

# QDA MINER 4

**QUALITATIVE AND MIXED-METHOD SOFTWARE**  
WITH UNPARALLELED COMPUTER-ASSISTED CODING,  
ANALYSIS, AND REPORT WRITING CAPABILITIES



# QDA MINER 4

## User's Guide

## **DISCLAIMER**

This software and the disk on which it is contained are licensed to you, for your own use. This is copyrighted software owned by Provalis Research Corp.. By purchasing this software, you are not obtaining title to the software or any copyright rights. You may not sublicense, rent, lease, convey, modify, translate, convert to another programming language, decompile, or disassemble the software for any purpose. You may make as many copies of this software as you need for backup purposes. You may use this software on up to two computers, provided there is no chance it will be used simultaneously on more than one computer.

## **WARRANTY**

The QDA Miner product is licensed "as is" without any warranty of merchantability or fitness for a particular purpose, performance, or otherwise. All warranties are expressly disclaimed. By using the QDA Miner product, you agree that neither Provalis Research nor anyone else who has been involved in the creation, production, or delivery of this software shall be liable to you or any third party for any use of (or inability to use) or performance of this product or for any indirect, consequential, or incidental damages whatsoever, whether based on contract, tort, or otherwise even if we are notified of such possibility in advance. (Some states do not allow the exclusion or limitation of incidental or consequential damages, so the foregoing limitation may not apply to you). In no event shall Provalis Research's liability for any damages ever exceed the price paid for the license to use the software, regardless of the form of claim. This agreement shall be governed by the laws of the province of Quebec (Canada) and shall inure to the benefit of Provalis Research and any successors, administrators, heirs, and assigns. Any action or proceeding brought by either party against the other arising out of or related to this agreement shall be brought only in a PROVINCIAL or FEDERAL COURT of competent jurisdiction located in Montréal, Québec. The parties hereby consent to in personam jurisdiction of said courts.

## **COPYRIGHT**

Copyright © 2004-2011 Provalis Research. All rights reserved. No part of this publication may be reproduced or distributed without the prior written permission of Provalis Research, 2414 Bennett Avenue, Montreal, QC, CANADA, H1V 3S4.

## **TRADEMARK**

Microsoft Windows is a registered trademark of Microsoft Corporation.

Excel, MS Access and FoxPro are products of Microsoft Corporation

SPSS/PC+ and SPSS for Windows are a registered trademark of SPSS Inc.

Other product names mentioned in this manual may be trademarks or registered trademarks of their respective companies and are hereby acknowledged.

# Table of Content

|                                                     |          |
|-----------------------------------------------------|----------|
| <b>Introduction to QDA Miner .....</b>              | <b>6</b> |
| Installation Instructions.....                      | 6        |
| Updating the Software .....                         | 6        |
| Project Structure .....                             | 7        |
| The Working Environment .....                       | 8        |
| <br><b>Data and Text Management</b>                 |          |
| Opening an Existing Project .....                   | 10       |
| Creating a Project from a List of Documents.....    | 12       |
| Creating a Blank Project.....                       | 15       |
| Importing an Existing Data File.....                | 19       |
| Importing Existing Documents as a New Project ..... | 23       |
| Project Description and Notes .....                 | 24       |
| Working with Menu Levels .....                      | 25       |
| Security & Multi-users Settings.....                | 27       |
| <br><b>Managing Variables</b>                       |          |
| Adding New Variables.....                           | 30       |
| Appending Variables from Another Data File .....    | 31       |
| Deleting Existing Variables.....                    | 32       |
| Changing Variable Types .....                       | 33       |
| Recoding Values.....                                | 36       |
| Editing Values Using the Spreadsheet Editor .....   | 37       |
| Editing Variable Properties.....                    | 38       |
| Computing Variables Statistics.....                 | 41       |
| <br><b>Managing Cases</b>                           |          |
| Adding and Deleting Cases.....                      | 43       |
| Appending New Documents and Images.....             | 46       |
| Adding New Cases from a Data File .....             | 48       |
| Filtering Cases .....                               | 49       |
| <br><b>Project Management Tasks</b>                 |          |
| Creating a Duplicate copy of a project .....        | 51       |
| Sending a Project by E-mail .....                   | 52       |
| Exporting Selected Cases to a New Project.....      | 53       |
| Exporting Coded Segments.....                       | 54       |
| Exporting Documents .....                           | 56       |
| Setting the Cases Descriptor and Grouping .....     | 57       |
| Importing a Document into a Case .....              | 60       |
| Importing an Image in a Case .....                  | 61       |
| Archiving a Project.....                            | 62       |

|                                                    |     |
|----------------------------------------------------|-----|
| Scheduled Backups.....                             | 62  |
| Using the Temporary Session Backup.....            | 64  |
| Merging Projects .....                             | 65  |
| Clearing all code assignments .....                | 70  |
| Retrieving Misplaced Codes.....                    | 71  |
| Using the Command Log.....                         | 72  |
| <b>Managing the Codebook</b>                       |     |
| Adding a Code.....                                 | 75  |
| Modifying Codes .....                              | 77  |
| Deleting Existing Codes .....                      | 77  |
| Moving Codes .....                                 | 78  |
| Splitting a Code .....                             | 78  |
| Merging Codes .....                                | 80  |
| Virtually grouping codes .....                     | 81  |
| Combining Codes .....                              | 81  |
| Importing a Codebook.....                          | 82  |
| Printing the Codebook.....                         | 84  |
| <b>Coding Documents and Images</b>                 |     |
| Assigning Codes to Text Segments .....             | 86  |
| Assigning Codes to Images .....                    | 87  |
| Working with Code Marks .....                      | 88  |
| Attaching a Comment to a Coded Segment.....        | 89  |
| Removing Coded Segments.....                       | 89  |
| Changing the Code Associated with a Segment .....  | 90  |
| Resizing Coded Segments .....                      | 91  |
| Hiding Code Marks .....                            | 92  |
| Highlighting and Hiding Coded Text Segments.....   | 92  |
| Highlighting and Hiding Areas of Coded Images..... | 93  |
| Modifying Code Mark Color Scheme .....             | 94  |
| Searching and Replacing Codes .....                | 95  |
| Consolidating Coded Segments.....                  | 95  |
| <b>Hyperlinking and Tagging Features</b> .....     | 97  |
| Adding an Hyperlink .....                          | 97  |
| Editing an Hyperlink .....                         | 100 |
| Deleting an Hyperlink .....                        | 100 |
| Following an Hyperlink.....                        | 100 |
| Geotagging of Documents and Images.....            | 101 |
| Importing Geographic Coordinates .....             | 103 |
| Attaching Icons to Geotags and Time-tags.....      | 105 |

|                                                |     |
|------------------------------------------------|-----|
| <b>Retrieval Features</b>                      | 107 |
| Saving and Retrieving Queries                  | 108 |
| Text Retrieval                                 | 111 |
| Thesaurus Editing                              | 117 |
| Section Retrieval                              | 119 |
| Keyword Retrieval                              | 123 |
| Cluster Extraction and Coding                  | 128 |
| Code Similarity Retrieval                      | 134 |
| Queby by Example                               | 137 |
| Coding Retrieval                               | 143 |
| Retrieving a List of Comments                  | 149 |
| Link Retrieval                                 | 151 |
| Creating and Editing Timelines                 | 154 |
| Exporting and Viewing Geotags                  | 157 |
| <b>Coding Analysis Features</b>                | 159 |
| Coding Frequencies                             | 160 |
| Coding Co-occurrences                          | 166 |
| Clustering Cases                               | 167 |
| Clustering Codes                               | 167 |
| Dendrogram                                     | 170 |
| 2d and 3d Concept Maps                         | 171 |
| Proximity Plot                                 | 175 |
| Code Sequence Analysis                         | 177 |
| Coding by Variables                            | 183 |
| Barchart or Line Chart                         | 187 |
| Heatmap Plot                                   | 191 |
| Correspondence Analysis                        | 196 |
| Bubble Charts                                  | 200 |
| Assessing Inter-coders Agreement               | 203 |
| Content Analysis                               | 210 |
| Statistical Analysis                           | 212 |
| <b>Miscellaneous Features</b>                  |     |
| Automatic Text Classification                  | 213 |
| Storing Code Statistics into Numeric Variables | 217 |
| Exporting Code Statistics                      | 219 |
| Exporting Co-occurrence or Similarity Matrix   | 221 |
| Coding Text Segments Using the Coding Table    | 222 |
| Using the Retrieved Segments Dialog Box        | 226 |
| Identifying and Substituting Unknown Words     | 228 |
| Managing Outputs with the Report Manager       | 230 |
| References                                     | 237 |
| Technical Support                              | 240 |

# Introduction to QDA Miner 4.0

QDA Miner is an easy-to-use qualitative data analysis software package for coding textual data and graphical, annotating, retrieving and reviewing coded data and documents. The program can manage complex projects involving large numbers of documents combined with numerical and categorical information. QDA Miner also provides a wide range of exploratory tools to identify patterns in codings and relationships between assigned codes and other numerical or categorical properties. Documents are stored in Rich-Text Format and support font and paragraph formatting, graphics and tables. Documents may be edited at any time without affecting the existing coding.

QDA Miner can import and export documents, data and results in numerous file formats (MS Word, WordPerfect, RTF, PDF, HTML, XML, MS Access, Excel, SPSS, Paradox, dBase, QSR N6, Nvivo, Atlas.ti, HyperResearch, Ethnograph, Transana, Transcriber, etc.). It also provides unique integration with advanced quantitative content analysis, text mining (WordStat) and statistical analysis (Simstat) tools, providing easy combination and integration of qualitative and quantitative methods.

## Installation Instructions

To install QDA Miner on your computer from the installation CD, place the CD in your CD drive. The installation program should start automatically. If it does not start, double-click the "My Computer" desktop shortcut and then on the CD drive icon to display the content of the CD. Double-click the INSTALL.EXE icon on the CD. Just follow the setup direction on screen.

If you downloaded the software, simply double-click the SetupQM.EXE icon and follow instruction displayed on your screen.

## Updating the Software

Minor revisions of QDA Miner are likely to be released on a regular basis and made available on the Provalis Research Web site. These releases may include bug fixes as well as minor improvements and sometimes even new features. Unless otherwise indicated, if your version of QDA Miner has the same starting number as the version available on the Web, you are entitled to a free update. For example, if you purchased the version 3.0 of QDA Miner, and the current version available on the Web site is 3.2.1, you should be able to upgrade to this version for free.

To update your software to the latest version, select the WEB UPDATE command from the HELP menu or go to the QDA Miner Web site and select the QDA Miner WEB PAGE from the HELP menu, or go to the following web page:

[www.provalisresearch.com/QDAMiner/QDAMiner.html](http://www.provalisresearch.com/QDAMiner/QDAMiner.html).

From there, go to the download section of the site, download the trial version of QDA Miner and install this version over your existing one. QDA Miner will automatically register the trial version and transform it into a fully functional version.

# Project Structure

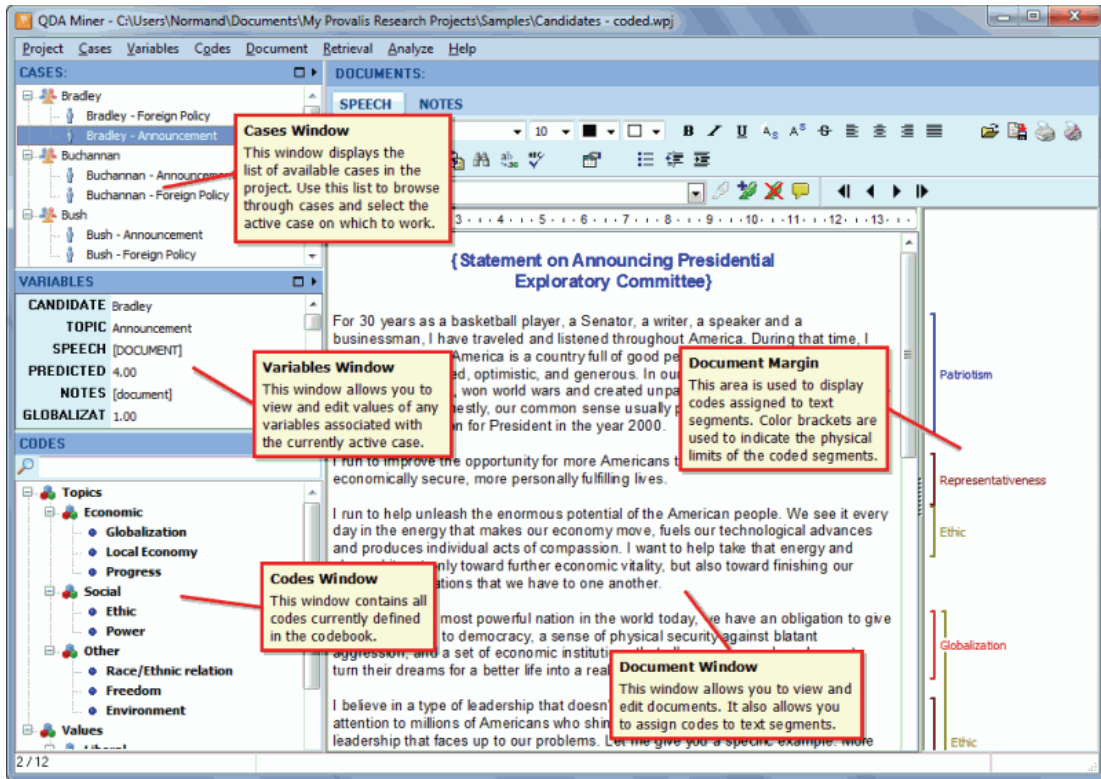
QDA Miner keeps all documents, coding schemes, codes, and notes in a set of files called a “project”. QDA Miner projects consist of multiple cases. A case is the basic unit of analysis of a project. It typically represents an individual, an organization, or a group. A case can contain several documents as well as numerous alphanumeric, numeric, categorical, date or Boolean variables. These variables are used to specify the properties associated with a case.

For example, if you want to analyze transcripts of in-depth interviews of numerous individuals, you may end up creating a project file where each case contains information associated with a single interviewee. One or several document variables may be created to contain transcripts of these interviews. You may also add other variables to specify sociodemographic information for the interviewee, group membership, etc, as well as the interview date and individual responses to closed-ended questions. Up to 2035 variables can be associated with each case. The number of cases in a single project is limited by the disk space, up to a maximum of 4 to 8 gigabytes.

One of the unique features of QDA Miner is its ability to explore relationships between any one of these properties and codes manually assigned to documents. For example, you can assess how the content of an interview is related to the interviewee’s gender or age, or how it relates to specific answers to a closed-ended question.

# The Working Environment

The QDA Miner working environment provides numerous tools to manage, view, edit, and code documents. It also gives access to a variety of retrieval and analysis tools. The picture below provides an example of the QDA Miner workspace.



The QDA Miner desktop consists of a menu bar at the top and four windows. The first three windows are located on the left side of the screen. These screens are, from top to bottom:

- The CASES window, which contains the list of available cases in the project. These cases are displayed either as a list or grouped as a tree. This window is used mainly to browse through cases and select the active case on which you want to work.
- The VARIABLES window, which displays and allows editing of values for any variable associated with the currently active case.
- The CODES window, located in the lower left corner of the screen, which contains all codes currently defined in the codebook.

The large DOCUMENT window on the right side of the screen is the main working space. This window is used to view and edit documents and images and assign codes to text segments or image areas. When a case contains more than one document or image, you can switch between them by selecting the document/image name from the list box located in the top left corner of this window. To the right of this window is a gutter area. This area is used to display codes assigned to text segments and image areas. A second gutter area is displayed at the bottom of images. Color brackets are used to indicate the physical limits of the coded segment. This area is used to review coding, remove or change assigned codes, and attach notes to any coded segments (see Working with Code Marks).

Each window may be resized either horizontally or vertically by dragging its border. The three windows in the right panel may be maximized to take up most of the available vertical space by clicking on the  button located in the top right corner of the window. Clicking on the  button restores the maximized window to its original size. You can also minimize a single window by clicking on the  button. The “folded” window may then be restored by clicking on the  button.

## The Main Menu

The PROJECT drop-down menu gives you access to various operations related to entire projects, such as project creation, opening and configuration. This menu also includes commands related to the import, export and backup of project files. It also gives access to various maintenance and teamwork operations (including merging, copying and emailing of projects) and to auditing tools such as the Command Log and the Report Manager features.

Operations associated with the main four windows of the working environment are grouped under menus with a corresponding name. For example, all operations relevant to the CASES windows are grouped under the CASES drop-down menu, while all commands relevant to the codebook currently displayed are located under the CODES drop-down menu.

The text and code retrieval features as well as the various exploratory tools available in QDA Miner (coding sequences or co-occurrences analysis, heatmaps and correspondence analysis plot, etc.) are all grouped under the ANALYSIS drop-down menu. This menu also provides access to external modules such as WordStat (quantitative content analysis and text mining) and Simstat (statistical analysis).

## Navigating through cases

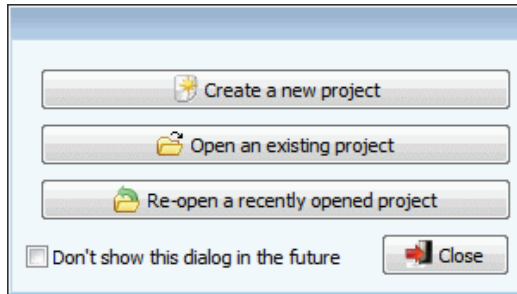
There are now three ways to navigate through cases. You can move to a specific case by clicking its entry in the CASES window. You may also browse through cases, by pressing the Ctrl-PgDn button to move to the next case or Ctrl-PgUp to move to the previous case. You may also use the navigation toolbar located above the document window. This toolbar displays the following four buttons

-  Move to the first case of the project.
-  Move back to the previous case
-  Move forward to the next case
-  Move to the last case of the project.

# Opening an Existing Project

## Opening a Project from the Introductory Screen

By default, QDA Miner stores data files in a folder named Data. The Data folder is located in the QDA Miner program folder. When you start QDA Miner, an introductory dialog box appears:



QDA Miner keeps track of the ten most recently use project files.

### To reopen a recently opened project

- Click the **Reopen a recently open project** button and select its name from the displayed drop-down list.

### To open another project

- Click the **Open an existing project** button. The **File** dialog box will appear.
- Locate and select your project (a QDA Miner project is stored with a .WPJ file extension).
- Click the **Open** button.

## Opening a Project When the QDA Miner Program Is Already Running

### To open an existing project:

- Select the OPEN command from the PROJECT menu. The **File** dialog box will appear.
- Locate and select your project (a QDA Miner project is stored with a .WPJ file extension).
- Click the **Open** button.

### **To open a recently opened project:**

- Select the REOPEN command from the PROJECT menu.
- Click the project file that you want to reopen.

## **Opening a Project from Windows Explorer**

You can also open a project by double-clicking its icon in Windows Explorer:



# Creating a New Project

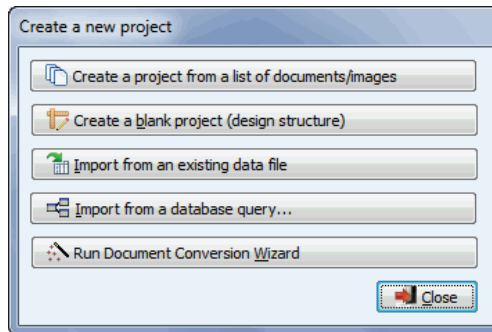
There are four ways of creating a QDA Miner project: (1) You can create a simple project by importing one or several documents; (2) you can create a new project from scratch by creating a project structure and then manually entering data and documents; (3) you can import existing data and documents stored in another file format such as Excel, MS Access, dBase, Paradox, QSR N6, HyperResearch 2.x, Transana, Transcriber, etc.; or (4) you can also use the Document Conversion Wizard utility program to import several documents at once, perform some transformation and data extraction, and store them in a new project.

## Creating an Project from a List of Documents

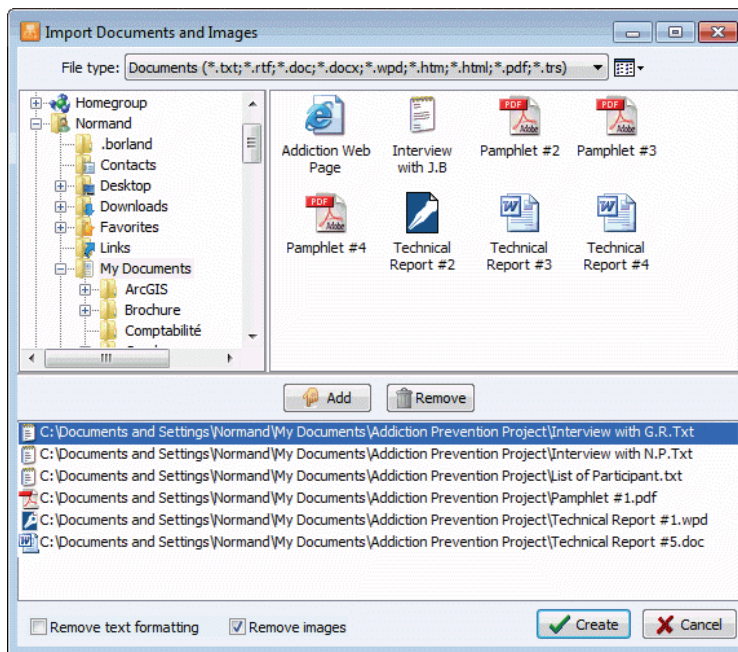
The easiest method to create a new project and start doing analysis in QDA Miner is by specifying a list of existing documents or images and importing them into a new project. Using this method creates a simple project with two or three variables: A categorical variable containing the original name of the files from which the data originated, a DOCUMENT variable containing imported documents and/or an IMAGE variable containing imported graphics. All text and graphic files are stored in different cases so, if 10 files have been imported, the project will have 10 cases with two or three variables each. To split long documents into several ones or extract numerical, categorical, or textual information from those documents and store them into additional variables, use the Document Conversion Wizard.

### To create a new project using this method

- Select the NEW command from the PROJECT menu. This command calls up a dialog box similar to the one below.



- Click the **Create a Project from a List of Documents/Images/** button. A dialog box similar to the one below will appear:



- Click a folder in the folders list on the upper left section of the dialog box to display its contents. If you want to see the contents of a drive, go to the folders list, click My Computer, and then double-click a drive.
- In the upper right section of the dialog box, QDA Miner displays all supported document file formats that may be imported, such as MS Word, WordPerfect, RTF, PDF documents, plain text files or HTML as well as all graphic file format supported (BMP, JPG, GIF, PNG and WMF). To display only files of a specific type, set the **File Type** list box to the desired file format.
- Click the file you would like to import. To select multiple files, hold down the **CTRL** key while clicking on the other files.
- Click the  **Add** button to add those files to the list of files to import, located at the bottom of this dialog box. You may also drag the files from the top right section to this list.
- To remove a file from the list of files to import, select that file name and click the  **Remove** button.

When importing formatted documents like Word, HTML or PDF files, images stored in the document may significantly increase the resulting document size and slow down the browsing and text-processing speed of QDA Miner and WordStat. For this reason, the Remove Images option has been set by default, resulting in smaller documents. To keep the images in the imported documents, simply disable this option.

If the text formatting of the existing documents such as the font styles and colors or the paragraph formatting are not relevant, one can further reduce the size of the imported documents by selecting

the Remove Text Formatting option. Enabling this option will convert all documents into plain-text documents without any formatting or images.

- Once all files have been selected, click the **Create** button. You will be asked to specify the name of the project that you want to create. If a project with an identical name already exists, you will be asked to confirm that you wish to overwrite the previous version of the project.

QDA Miner closes this dialog box, imports all the specified files into a new project and then brings you back to the main window.

## Some Notes about Importing PDF Files

The PDF file format is designed to display text in the same way on various platforms. Different strategies have been considered in QDA Miner to support PDF files. They could simply be imported as is and stored in the QDA Miner project, allowing one to view the documents exactly as they appear in Acrobat. One could then code the PDF documents directly. Such an approach has, however, several inconveniences. First, the document might not be edited, preventing one from removing unnecessary information, such as headers and footers, appearing on each page or unrelated information like advertising. Also, PDF files are often created by outputting each page and each line of text separately, often resulting in the loss of the original document structure. Carriage returns will often be inserted at the end of each line, paragraphs spread over two pages remain physically split up, as do hyphenated words. In multi-column documents, lines from different columns may even be mixed up. All these minor imperfections would prevent one from retrieving full sentences and paragraphs in QDA Miner and may reduce recall results when searching for specific words or phrases separated by hyphens, carriage returns or page limitations. Such flaws would also undermine the performance of numerous features of WordStat relying on phrase identification, word co-occurrence analysis or proximity rules.

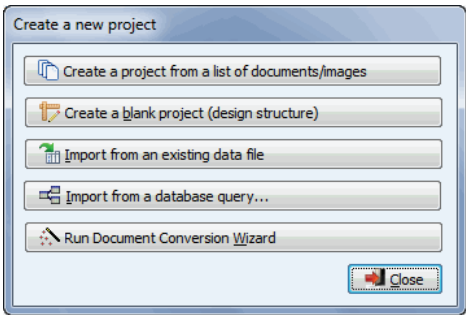
For all these reasons, we chose to import PDF documents by converting them into editable documents, storing them in rich-text format, just like any other document file type supported by QDA Miner. The conversion engine has been carefully designed to remove hyphens and unnecessary carriage returns, adjust the text flow in multi-column documents, and import tables and images correctly. While headers and footers will still be imported, which will break the flow of text across pages, one may now easily identify those and remove them from the text since the imported document will be fully editable.

Despite all those advanced importation features, some PDF documents may still not be imported properly. Several factors may explain such difficulties. Some PDF files consist of only scanned images of the original document and contain no text at all. These can be easily recognized by the fact that it is not possible to select any text segment or that text searches never return any hits. Other documents may have quite complex layout designs, making their proper importation quite difficult. For those documents, we recommend pre-processing them with full-fledged OCR (Optical Character Recognition) software like Abbyy's FineReader or Nuance's OmniPage or by some PDF to Word conversion utility tools like PDF Transformer by Abbyy. Although the latest version of Acrobat Professional does include some OCR features, its performance was deemed not good enough to preserve the document structure.

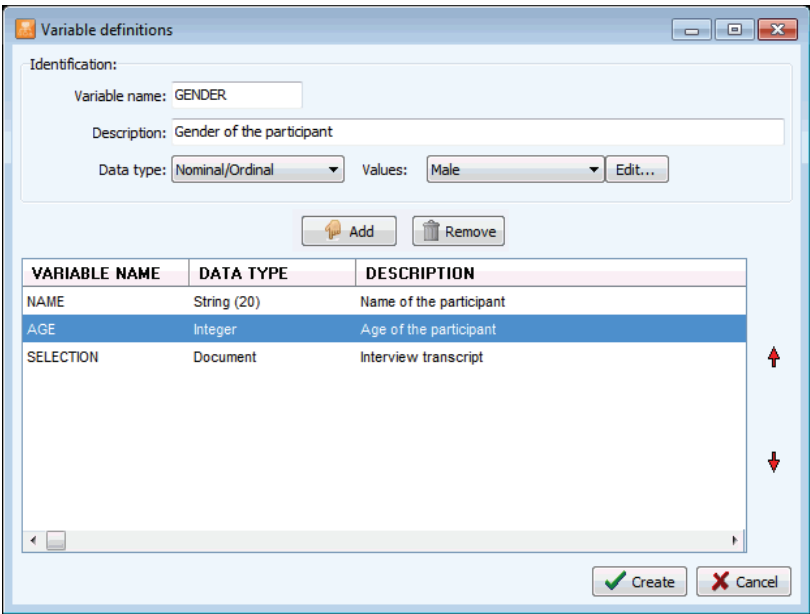
We also recommend removing images or scaling down the graphic resolution of images to the lowest setting because images will significantly increase the resulting document size and will slow down the browsing and text-processing speed of QDA Miner and WordStat.

# Creating a Blank Project

To create a new data file, select the NEW command from the PROJECT menu. This command calls up a dialog box similar to the one below:



Select the **Create a Blank Project** button. The **Variable Definitions** dialog box will appear.



The first step involved in creating a new project file is to define the initial structure of the project file. This structure is defined by the list of variables that each case will contain. A variable may contain a document or an image to be manually coded, but can also consist of a numeric value, a date, a Boolean value (true or false), etc. A single project can contain up to 2035 variables per case. It is possible to create several document and image variables for each case. The ability to store many documents per case is especially useful when the project involves a set of several document types. A document variable may also be created to store notes or comments that are specific to a case. QDA Miner can later be instructed to search for and analyze specific document variables or in all documents.

In the **Variable Definitions** dialog box (see above), you can define various attributes for these new variables, such as their name, and whether they will contain numeric, alphanumeric values, or dates.

**VARIABLE NAME** - The first edit box at the top of the dialog box allows you to enter a variable name. Each variable name must be unique (within that project file). Valid variable names begin with a letter and may contain letters, numbers or underscore characters. Punctuation marks, blank spaces, accentuated and other special characters are not permitted. The maximum variable name length is ten characters.

**DESCRIPTION** - The Description option is used to enter a variable label that describes the content of the variable in more detail. You may leave this column blank if you wish since it is always possible to add or edit a description later using the **VARIABLES | PROPERTIES** command.

**DATA TYPE** - Each variable in the data file must have a type. QDA Miner supports the following types:

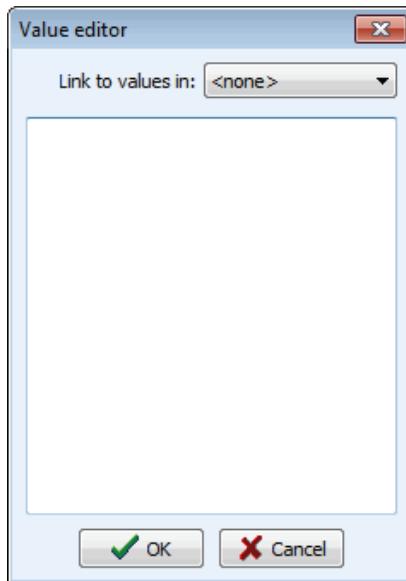
**Document** - This data type is used to store documents that will be manually coded. QDA Miner stores text in this data type using Rich Text Format (RTF). This format enabled the use of different fonts and styles and paragraph formatting. Graphics and tables may also be inserted in the document. Numerous file formats may be directly imported into document variables, such as plain ASCII files (\*.TXT), Rich Text files (\*.RTF), MS Word documents (\*.DOC; \*.DOCX), HTML (\*.HTM or \*.HTML), Adobe Acrobat (\*.PDF), and WordPerfect documents (\*.WPD) (see **Importing a Document or an Image in a Case**, on page 60).

**Image** - This data type is used to store graphics that will be manually coded. Numerous file formats may be directly imported into document variables, such as Windows bitmaps (\*.BMP), Windows Meta Files (\*.WMF), or Enhanced Meta Files (\*.EMF), CompuServe Graphic Interface (\*.GIF), Portable Network Graphics (\*.PNG) or JPEG files (\*.JPG or \*.JPEG) (see **Importing a Document or an Image in a Case**, on page 60).

**Numeric** - Numeric variables can contain either integer or floating-point numbers. When you choose floating point numbers, you will be asked to specify the number of decimal places to display. Floating-point numbers are stored in the data file using double precision values (at least 15 significant digits). The decimals option is used exclusively to control how numeric values are displayed in the **Variables** grid and in other locations and does not affect the internal precision of the variable.

**Nominal/Ordinal** - Nominal or ordinal variables are used to hold a limited number of short strings used to describe specific properties of a case. For example, you may choose to use this variable type to identify the gender of the interviewee ("male" or "female") or its group membership ("manager", "employee", "client", etc.). You may also use this data type to hold ordinal ranking ("novice", "intermediate", or "expert") or responses to close-ended questions such as Likert scale items (e.g.: from "strongly disagree" to "strongly agree"). You should use this data type instead of short strings if you plan to analyze this variable or examine the relationship between its values and any other variable or coding of documents.

When a nominal/ordinal data type is selected, you will be asked to first provide a list of values that this variable can take. To enter new values, click the **Edit** button. A dialog box similar to the one below will appear:



You can start typing values (one value per line) in the large edit box. If the current variable uses the same values as another existing variable, you may also establish a link to this other variable so that they will share the same list of values. Click **OK** to confirm the setting of these values. Note: This list may later be modified to add new values or edit existing ones.

**Date** - The date type holds a year, month and day. The display and data entry format used for dates is based on the Windows date setting.

**Boolean** - The Boolean type stores a value that can be either true or false (or Yes or No).

**Short String** - Short string variables can contain up to 254 alphanumeric characters. When creating a short string variable, you must specify first the maximum number of characters this variable will hold.

### To create a new variable:

- Enter a unique variable name.
- Enter a description (optional).
- Select the data type for this variable.
- Set the option associated with the chosen data type (see above).
- Click the **Add** button to add the defined variable to the list.

### To remove a variable from the list:

- Select the row containing the variable you want to delete.
- Click the **Remove** button.

### To change the position of a variable in the list

- Select the row containing the variable you want to move.
- Click the **up** or **down** arrows located to the right of the grid until the variable appears in the desired location.

### To create the data file:

When you have finished defining the structure of the new project, click the **OK** button.

You will be asked to specify the name of the project that you want to create. If a project with an identical name already exists, you will be asked to confirm the overwriting of the older project.

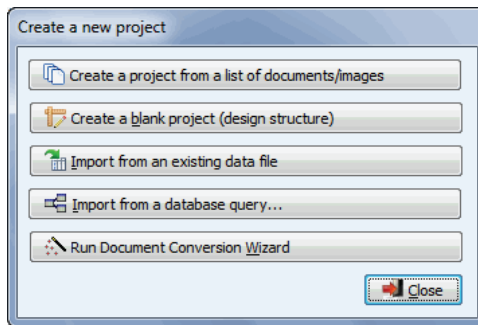
# Importing an Existing Data File

QDA Miner allows you to directly import data files from spreadsheet and database applications, as well as from plain ASCII data files (comma or tab delimited text). The program can read data stored in the following file formats:

- MS Access
- DBase
- Paradox (v3.0 - v5.0)
- Lotus 123 (v1.0 - v5.0)
- MS Excel
- Quattro Pro (v1.0 - v6.0)
- Comma Separated Values
- Tab Separated Values
- Triple-S XML files (interchange standard for survey data)
- QSR N6
- Atlas.ti 5.x and 6.x hermeneutic unit
- HyperResearch 2.x
- Enthograph 5.x project files (document, codebook and coding only)
- Transcriber
- Transana 2.x XML files.
- Reference Information System (RIS) data files.

## To import data from any of these applications:

- Select the **NEW** command from the **PROJECT** menu. The following dialog box will appear:



- Click the **Import from an existing database** button.
- Select the file format using the **List File of Type** drop-down list.
- Select the file you want to import and click **OK**.

We will now discuss the specific file formats in greater detail, including formatting requirements (if applicable), supported features and limitations.

## ASCII Data Files

QDA Miner will read up to 500 numeric and alphanumeric variables from a plain ASCII file (text file). The file must have the following format:

- Every line must end with a carriage return.
- The first line must include the variable names, separated by spaces, tabs and/or commas.
- Variable names may not be longer than ten characters. Longer strings are truncated at ten characters.
- The remaining lines must include numeric scores separated by spaces, tabs and/or commas.
- Each line must contain data for one case; variables must be in the same order for all cases.
- All invalid scores and all blanks encountered between commas or tabs are treated as missing values. A single dot can also be used to represent a missing value.
- Comments can be inserted anywhere in the file by putting an asterisk (\*) at the beginning of the line.
- Blank lines can also be inserted anywhere in the file.

## Spreadsheet Data Files

In spreadsheet programs, you can enter both numeric and alphanumeric data into the cells of a data grid. QDA Miner can import spreadsheet files produced by EXCEL, LOTUS 1-2-3 (v1.1 to v5.0) and QUATTRO PRO (v1.0 to v6.0).

When you select a spreadsheet file format for importing, the program displays a dialog box in which you can specify the spreadsheet page and the range of cells where the data are located. You must specify a valid range name or provide upper left and lower right cells, separated by two periods (such as A1..H20). If you set the **Range Name** list box to **ALL**, the program attempts to read the whole page.

### Formatting spreadsheet data

The selected range must be formatted so that the columns of the spreadsheet represent variables (or fields) while the rows represent cases or records. As well, the first row should preferably contain the variable names, while the remaining rows should hold the data, one case per row. QDA Miner will automatically determine the most appropriate format based on the data that it finds in the worksheet columns. Cells in the first row of the selected range are treated as field names. If no variable name is encountered, QDA Miner will automatically provide one for each column in the defined range.

## Database Files

### dBase and Paradox files

QDA Miner can import dBase and Paradox data files. However, the length of alphanumeric fields should not exceed 256 characters and memo fields are not supported. If you do have this type of data, you may use the exporting capabilities of your database program to create a data file that is more compatible with QDA Miner (such as a FoxPro 2.x data file or a tab delimited text file).

### MS Access data files

When you select a MS Access data file, the program displays a dialog box in which you can specify the table where the data are located. Once the table has been selected, click the **IMPORT** button to import the data file.

## Atlas.ti hermeneutic unit

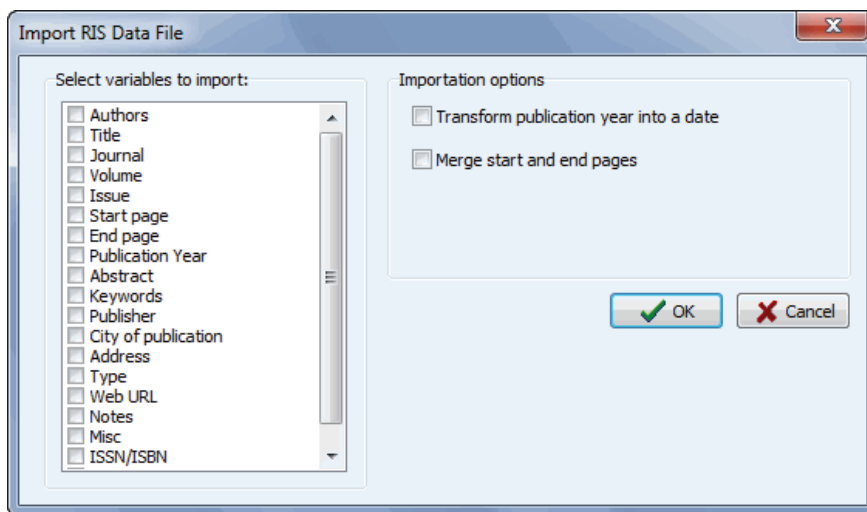
To import an Atlas.ti project, follow these instructions:

1. In Atlas-ti, open the project you want to import into QDA Miner.
2. Run **the XML | EXPORT HU to XML** command from the **TOOLS** menu.
3. When asked when you want to include documents and quotations, click **YES**.
4. Select **FILE** as the output destination and click **OK**. A Save File dialog box will appear.
5. Change the name of the XML file so that it has the same name as the project file but with an XML file extension (you may need to remove the "hu\_" prefix). Save this file in the same folder as the original .HPR5 file.
6. Close Atlas-ti.
7. Open QDA Miner, run the QDA Miner importation routine and select either the .HPR5 or .HPR6 file or the newly created .XML file.

## Reference Information System (RIS) data files

QDA Miner can import data created by reference management software (like EndNote, Reference Manager, ProCite, Zotero), as well as many digital libraries (such as IEEE Xplore, ScienceDirect, SpringerLink). To import data, you need to export from those programs or services the references into an RIS data file. RIS files are plain-text files and have by default a TXT file extension. To import such a file, set the file type list box to Reference Information Management (RIS) and then select the file you wish to import. Another approach is to change the file extension to RIS, and QDA Miner will automatically recognize this file extension and import the references in this file without the need to specify the proper file type.

Once a file has been selected for importation, QDA Miner will display a dialog box similar to this one:



The following importation options are available:

**Select variables to import** - This list box allows you to put check marks beside variables that you want to import. All unchecked items will be ignored.

**Transform publication year into a date** - The publication dates in RIS files can consist of a full date, with days and months, or only the year, or sometimes the publication month and year. By default, QDA Miner will import just the publication year and store it in an integer variable. Choosing this option will store all dates into a DATE variable. If only the year is specified, QDA Miner will set the day and month to January 1. If the publication date consists of a month and a year only, then QDA Miner will set the day to the first day of the month.

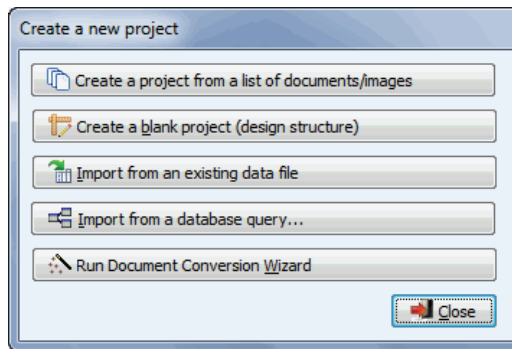
**Merge start and end pages** - By default, start-page and end-page numbers are stored in separate variables. Selecting this option will join both numbers with a hyphen character and store those in a single string variable.

# Importing Existing Documents as a New Project

The Document Conversion Wizard is a utility program used to import one or more documents into a new project file. This tool supports the importation of numerous file formats including plain ASCII, Rich Text Format, MS Word, HTML, Acrobat PDF files, and WordPerfect documents. It may be instructed to split large files into several cases and to extract numeric and alphanumeric data from these files. The Document Conversion Wizard may be run either as a standalone application or from within QDA Miner. When no splitting, transformation or extraction of information from documents are necessary, an easier method would be to use **Creating a Project from a List of Documents**.

## To Run the Document Conversion Wizard from Qda Miner:

- Select the NEW command from the PROJECT menu. The following dialog box will appear:



- Click the **Run Document Conversion Wizard** button.

The program then guides you through the necessary steps needed to import the documents and extract relevant data. For more detailed information on the document importing process using this utility program consult the **Document Conversion Wizard** help file.

## Project Description and Notes

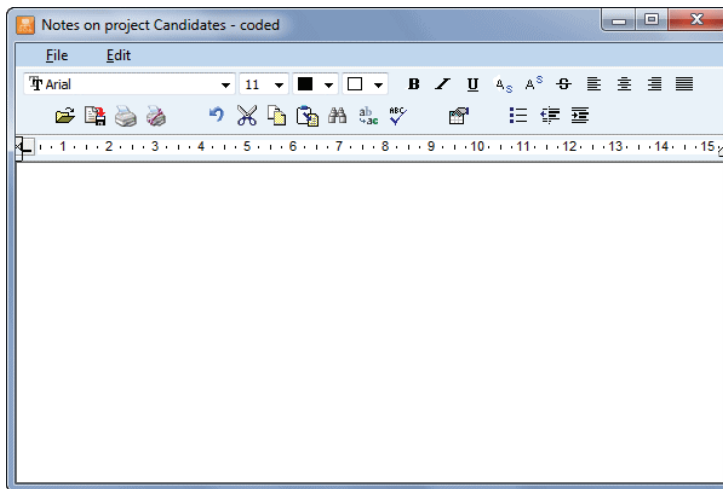
QDA Miner allows you to specify a description for the project as well as write general notes to yourself or to be shared with other people working on this project. The project description typically contains general information about the project and may be configured so that it will be displayed automatically when opening the project file. Notes may be used as reminders of decisions you made, things to be done, etc.

**To enter a project description:**

- Select the **PROPERTIES** command from the **PROJECT** menu.
- Enter a description for the project.
- To display this description automatically when the project file is opened, select the **Show Description upon Opening** checkbox.
- Click the **OK** button to save the changes you made.
- To quit this dialog box without saving the changes, click the **Cancel** button.

**To enter or edit general notes:**

- Select the **NOTES** command from the **PROJECT** menu, or press the **F3** key from anywhere in the program. This will open a Note editor like the one below:



- Enter your comments or observations. You can format text and paragraphs, change font attributes, insert graphics and tables. You may also save your notes to disk in RTF format.
- To close this dialog box, select the CLOSE button from the PROJECT menu.

# Working with menu levels

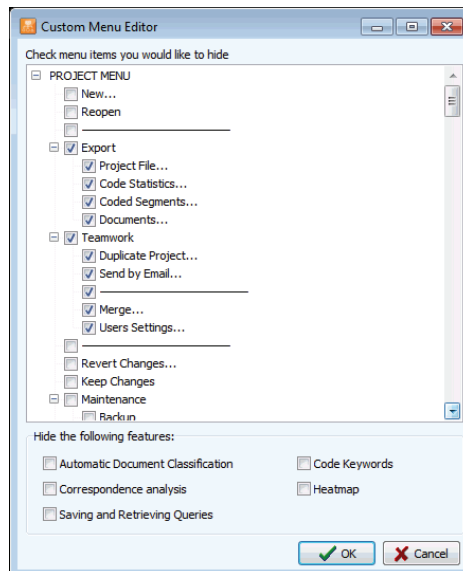
QDA Miner offers not only basic coding and retrieval functions but also advanced analytical tools that may be intimidating for novice users or unnecessary for users only involved in the coding of documents and not in the data management or data analysis of a research project. The MENU LEVEL command, available from the PROJECT menu, may be used to simplify the interface by reducing the number of visible menu items, making the software easier to learn and use. QDA Miner currently offers three modes, or levels, of menus: FULL showing the full set of commands and features including the analysis and data management features, BASIC that presents only the most commonly used commands for applying codes to documents, and CUSTOM that can be adjusted to display any chosen set of menu items suitable for a specific purpose.

By default, when a menu level is chosen it will be used across projects and across sessions until the user explicitly changes the menu level to a different setting. A project administrator can also associate different default menu levels to specific users, preventing them from accessing some commands and hiding unnecessary features. For more information on how to assign menu levels to users, see Setting Security & Multi-users Settings.

## Editing the Custom Menu

The definition of “custom menu” is specific to projects, so one could create different menus for different projects. To customize the menu:

- Select the PROPERTIES command from the PROJECT menu.
- Move to the Miscellaneous page and click the Edit button on the right of the Custom Menu option. A dialog box similar to the one below will appear:



- The top list box displays items from the pull-down menu that may be hidden from the user. Other advanced features are also listed in a box below this list. To hide an item, simply click beside it to insert a check mark.
- Once finished, click the OK button to confirm the changes. To close this dialog box without saving any changes, click the Cancel button.

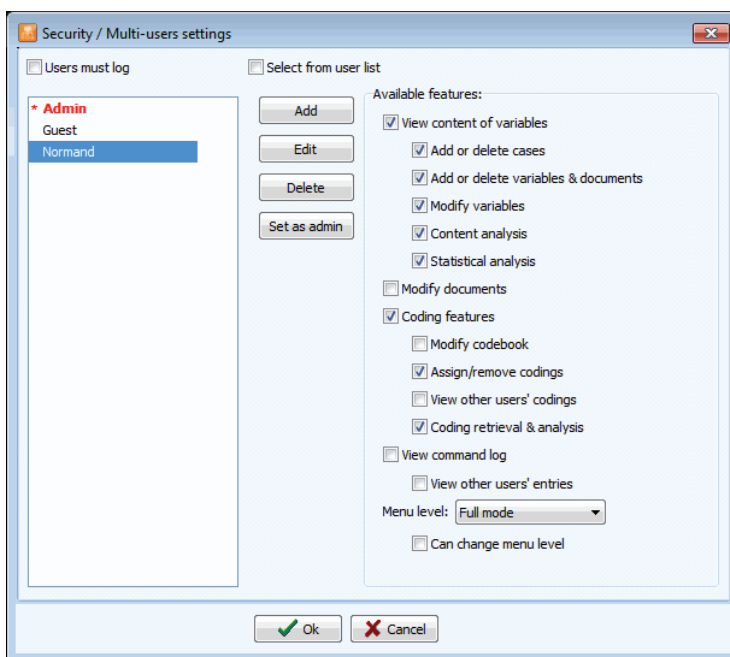
# Security & Multi-users Settings

It is possible to limit the number of people who can access and edit a project or limit the type of operations that can be performed by certain individuals by creating user accounts and requiring people to provide a user name and password when they access the project. This feature is useful for preventing deletion of existing cases or variables, or the editing of documents or values stored in other variables. The multi-user account feature is also useful for assessing the reliability of coding by allowing the same documents to be coded by different users (see **Assessing Inter-coders Agreement** on page 203).

When setting a project to support multiple users, one of these users should be able to control access to the project, create and delete user accounts and define passwords. This user is commonly known as the administrator. When creating a new project, an administrator account is automatically created. Both the default user name and password is ADMIN. If you choose to restrict access to some users, it is highly recommended that you change this user name and password to prevent unauthorized changes to user access rights.

## To change the multi-users settings of a project:

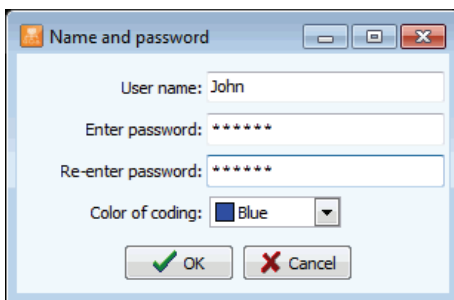
- Select the TEAMWORK | USERS SETTINGS command from the PROJECT menu. The following dialog box will appear:



By default, opening a project without using the user log screen gives the user all administrator access rights. Enabling the **Users must log** option will display a logon dialog box prompting the user to enter a valid user name and password in order to access the project file. The user name can be either entered in a text box or selected from a drop-down list box. To display existing user names in a list box, set the **Select from users list** option.

## To add a new user account:

- Click the **Add** button The following dialog box will appear:



- Enter in the user name.
- In the **Enter Password** edit box, enter this user's password.
- Enter the password a second time in the next edit box to make sure it was entered correctly.
- QDA Miner can emphasise coding assigned by different coders using a distinct color for each coder for displaying their code marks. The Color of Coding option can be used to select the color used to display codes assigned by this coder (see **Modifying Code Marks Color Scheme** on page 94).
- Click **OK** to save the new user account.

Once a user's account has been created, you have to specify which specific features this user will be able to access.

## To define the user access rights:

- Select the user for which you want to define or edit access rights.
- In the list of available features to the right of the dialog box, select the features you want this user to have access to and clear the features that you do not want him to access.

When creating several user accounts to establish inter-coders agreement, it is strongly recommended that you deactivate the **View other users' coding** option. This prevents a user from seeing code marks assigned to text segments by other users. Disabling this feature also restricts any coding retrieval procedures to the coding created by this user and prevents access to the inter-coders agreement dialog box. To prevent users from modifying existing documents or values stored in variables, clear the box beside the **Modify variables** and **Modify documents** options. It may also be a good idea to disable the **Add or delete cases** and **Add or delete variables/documents** options. Lastly, if you want to prevent users from adding new codes to the codebook, or deleting or modifying existing ones, make sure the **Modify Codebook** option is disabled.

- Adjust the **Menu Level** option to define a default menu level for this user. To prevent this user from changing to another menu level, remove the check mark beside the **Can Change Menu Level** option.

#### **To remove a user account:**

- In the list of existing accounts to the left of the dialog box, select the account you want to remove.
- Click the **Delete** button.
- Select **Yes** to confirm the deletion of this user's account.

To save the changes you made to the accounts and close this dialog box, click the **OK** button. To close this dialog box without saving any changes, click the **Cancel** button.

# Adding a New Variable

The **ADD** command in the **VARIABLES** menu is used to add new variables to the project file. This command displays a dialog box similar to the one used to create variables for a new data file (for more information on this dialog box and its options see **Creating a new empty project** on page 15).

Specify the name of the new variables and set the variable types, sizes and descriptions.

Click the **OK** button to create these new variables and add them to the end of the current data set.

# Appending variables from another data file

This command allows one to append new variables and values stored in an external data file to each case in the current project. The supported file formats for this external file are:

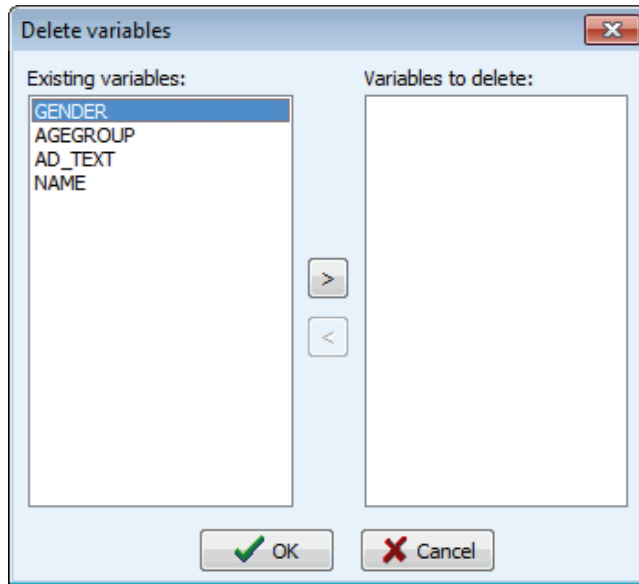
- QDA Miner projects (\*.wpj)
- Simstat data files (\*.dbf)
- MS Excel spreadsheets (\*.xls)
- MS Access data files (\*.mdb)
- Tab delimited files (\*.tab)
- Comma separated value files (\*.csv)

The importation of the variables is performed by matching cases using one or several key variables shared by the project file and the external data file. If no case of the external data file matches the key values of a case in the current project, missing values are assigned to the newly created variables. If several cases in the secondary file match the key values, values are extracted from the first matching case.

Variables with identical names may only be used to match cases, and their data cannot be imported into the current project. To replace the values of one of those variables with the ones in the external data file, simply delete this variable from the current project prior to this operation.

# Deleting Existing Variables

To delete one or more variables, select the DELETE command from the VARIABLES menu. The following dialog box will appear.



Highlight the names of the variables that you want to delete and click the  button to move them to the **Variables To Delete** list box.

To delete successive variables, click the first variable, drag the mouse cursor down the list to highlight multiple variables, and then click the  button.

To remove a variable from the list of variables to delete select the name of this variable from the **Variables to Delete** list box and click the  button.

Click the **OK** button to delete all selected variables.

# Changing Variable Types

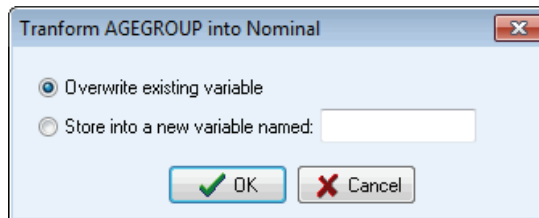
Some operations require specific variable types. For example, qualitative coding cannot be applied to nominal or string variables but only on documents or graphic variables. Also, a comparison of coding among subgroups of cases using the Coding by Variable command can only be made if the comparison variable is numeric or nominal. Another situation where a change in data type may be required is when one needs to add decimal values to a numeric variable created as an integer variable. One first needs to transform this integer variable into a floating-point numeric data type to do so.

QDA Miner offers the ability to change the type of an existing variable or to create a new variable containing values of the existing variable but storing those values using a different data type. The following transformations are currently supported:

Float -> Integer  
Float -> String  
Integer -> Float  
Integer -> String  
Integer -> Nominal  
Nominal -> String  
String -> Nominal  
String -> Document (codable)  
String -> Date  
Date -> String

To change the type of a specific variable:

- Select the variable you want to transform by clicking it in the Variable window.
- Select the TRANSFORM command from the VARIABLES menu. The allowed transformations will be listed in a submenu.
- Chose the desired data-type transformation. A dialog box similar to the one shown below will appear.

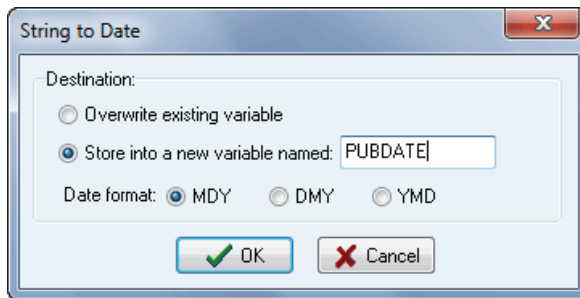


To change the type of the selected variable, choose **Overwrite Existing Variable** and click the **OK** button.

To copy the values of the selected variable into a variable of the new data type, choose **Store into a new variable named**, type the name of the new variable in the edit box and click the **OK** button. If the new variable name already exists, you will be prompted to confirm the overwriting of this variable.

## Transforming Strings Into Dates

QDA Miner can extract date information from a string and store the date into a new variable. It can recognize in a string various date formats such as 01/05/2011, January 5th, 2011, 5 JAN 2011, or 2011-01-05 and will ignore surrounding text. When this transformation is performed, a dialog box similar to the one shown below will appear:

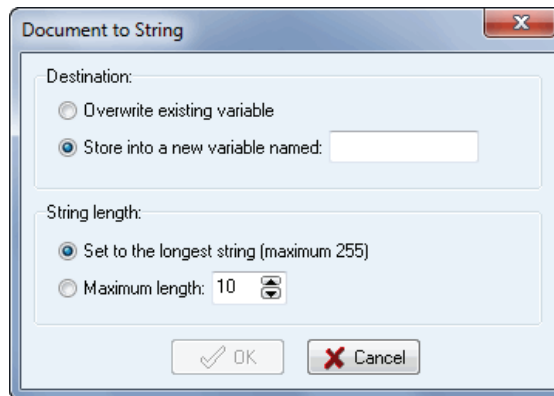


- To change the selected string variable into a date, choose **Overwrite existing variable**. To store the date values that are extracted from the selected variable into a new date variable, choose **Store into a new variable named** and type the name of the new variable in the edit box and then click the **OK** button. If the new variable name already exists, you will be prompted to confirm the overwriting of this variable.
- The Date Format option allows you to choose the specific sequence that is used for displaying dates. QDA Miner will recognize only one date sequence format at a time, either month/day/ year (MDY), day/month/year (DMY) or year/month/day (YMD). Select the sequence used in the most common date format. If no date with a specific format is found for a case, the date variable will remain empty.

## Transforming Documents Into Strings

QDA Miner can transform a document variable into a string with a maximum length of 256 characters. Several situations may justify such a transformation. While documents can be coded, one cannot filter cases based on the content of document variables. Also, case descriptors cannot include text stored in documents, while string variables may be used as part of the case descriptors. When importing several data files, long strings may be imported as document variables even if there was no intention of coding them.

When this transformation is requested, a dialog box similar to the one shown below will appear:



- To change the selected document variable into a string type, choose **Overwrite existing variable**. To store the date values that are extracted from the selected variable into a new date variable, choose Store into a new variable named and type the name of the new variable in the edit box. If the new variable name already exists, you will be prompted to confirm the overwriting of this variable.
- One can let QDA Miner set the size of the new string variable automatically by selecting the Set to the longest string option. QDA Miner will go through all cases and set the size to the longest string encountered. If a document is longer than 255 characters, the variable size will be set to 255 and the document will be truncated to this length. One can also set the size to a fixed length by setting the Maximum length option to the desired length. All documents longer than the set length will be truncated to this length.

# Recoding Values of a Variable

The RECODE command provides an easy way to apply multiple changes to the values of numeric, categorical or string variables or to create new variables based on a new grouping of the values of an existing variable.

To access this command, select the TRANSFORM | RECODE command from the VARIABLES menu. A recoding dialog box will appear with the following regions:

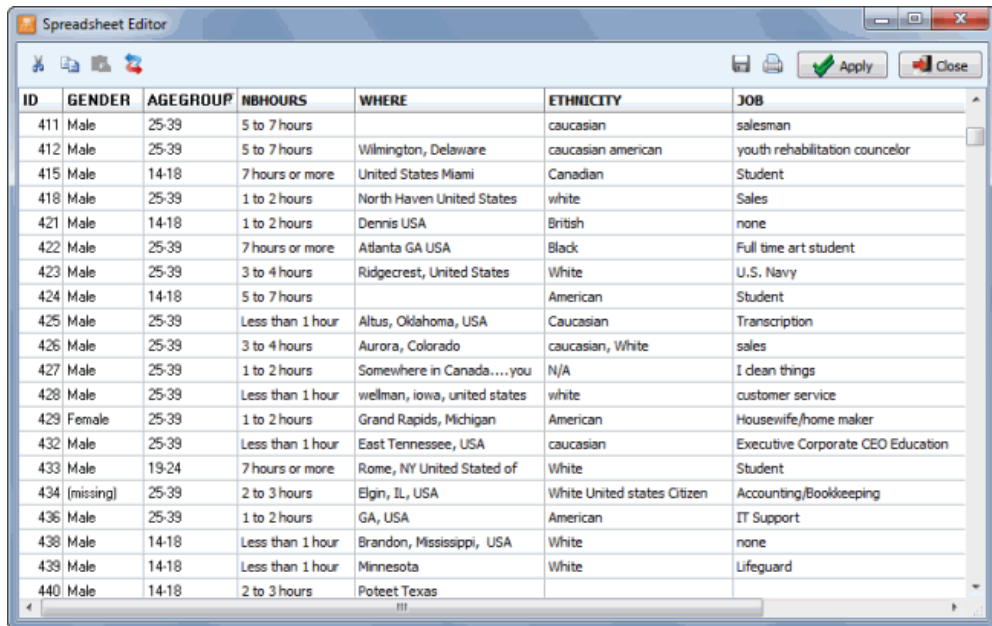
- At the top of the dialog box the **Destination** box allows you to specify where the computation results will be stored. To transform the values of the selected variable, choose the **Overwrite Existing Variable** option. To keep the selected variable intact and store the result in another variable, select the **Store in a New Variable Named** option and type in the name of the new target variable name in the edit box . If you enter the name of an existing variable, the program will ask you if you want to replace its values with those produced by the change. If the variable does not exist, the program will ask you to confirm the creation of this new variable.
- The **Existing Values** list box displays the list of all values found in the selected variable. Recoding is performed by selecting one or several items from this list and by either typing a new value in the **Recode As** edit box or selecting another existing value from its drop-down list. To confirm the recoding, click the **Add** button. The value transformations to be performed will be listed in the recoding list box. Repeat this operation until all desired transformations have been specified. Untransformed values remaining in the **Existing Values** list box will remain unchanged or will be copied to the destination variable if you selected to store values in another variable.

The special keyword **<missing>** is used to replace specific values with an empty cell and further treat those values as missing. Cases with missing values are ignored when performing some analyses involving those variables, such as the Coding by Variables command.

To remove a transformation from the list of recoding, select it and press the **<Delete>** keyboard key. Once a valid destination and a recoding list have been entered, you can leave this dialog box and perform the recoding by clicking the **OK** button. To leave the dialog box without performing any recoding, click the **Cancel** button.

# Editing Values Using the Spreadsheet Editor

Entering numerical, date or string values to variables in many cases can be time consuming when performed in the QDA Miner main screen, since the user only has access to values of one case at a time and thus needs to move from one case to another. A much more efficient data-entry or data-editing approach is to use the QDA Miner spreadsheet editor. To call the editor, run the SPREADSHEET EDITOR command from the VARIABLES menu. A dialog box similar to this one will appear:



| ID  | GENDER    | AGEGROUP | NBHOURS          | WHERE                        | ETHNICITY                   | JOB                               |
|-----|-----------|----------|------------------|------------------------------|-----------------------------|-----------------------------------|
| 411 | Male      | 25-39    | 5 to 7 hours     |                              | caucasian                   | salesman                          |
| 412 | Male      | 25-39    | 5 to 7 hours     | Wilmington, Delaware         | caucasian american          | youth rehabilitation counselor    |
| 415 | Male      | 14-18    | 7 hours or more  | United States Miami          | Canadian                    | Student                           |
| 418 | Male      | 25-39    | 1 to 2 hours     | North Haven United States    | white                       | Sales                             |
| 421 | Male      | 14-18    | 1 to 2 hours     | Dennis USA                   | British                     | none                              |
| 422 | Male      | 25-39    | 7 hours or more  | Atlanta GA USA               | Black                       | Full time art student             |
| 423 | Male      | 25-39    | 3 to 4 hours     | Ridgecrest, United States    | White                       | U.S. Navy                         |
| 424 | Male      | 14-18    | 5 to 7 hours     |                              | American                    | Student                           |
| 425 | Male      | 25-39    | Less than 1 hour | Altus, Oklahoma, USA         | Caucasian                   | Transcription                     |
| 426 | Male      | 25-39    | 3 to 4 hours     | Aurora, Colorado             | caucasian, White            | sales                             |
| 427 | Male      | 25-39    | 1 to 2 hours     | Somewhere in Canada....you   | N/A                         | I clean things                    |
| 428 | Male      | 25-39    | Less than 1 hour | wellman, iowa, united states | white                       | customer service                  |
| 429 | Female    | 25-39    | 1 to 2 hours     | Grand Rapids, Michigan       | American                    | Housewife/home maker              |
| 432 | Male      | 25-39    | Less than 1 hour | East Tennessee, USA          | caucasian                   | Executive Corporate CEO Education |
| 433 | Male      | 19-24    | 7 hours or more  | Rome, NY United States of    | White                       | Student                           |
| 434 | (missing) | 25-39    | 2 to 3 hours     | Elgin, IL, USA               | White United states Citizen | Accounting/Bookkeeping            |
| 436 | Male      | 25-39    | 1 to 2 hours     | GA, USA                      | American                    | IT Support                        |
| 438 | Male      | 14-18    | Less than 1 hour | Brandon, Mississippi, USA    | White                       | none                              |
| 439 | Male      | 14-18    | Less than 1 hour | Minnesota                    | White                       | Lifeguard                         |
| 440 | Male      | 14-18    | 2 to 3 hours     | Poteet Texas                 |                             |                                   |

This spreadsheet editor allows efficient data-editing operations. You can cut a cell or a range of cells and paste it into other rows, delete multiple cells, sort data, resize and rearrange columns, and print and save the spreadsheet to disk. Please note that pasting data from other cells can only be performed if the copied cells are from the same columns or if the data types are fully compatible.

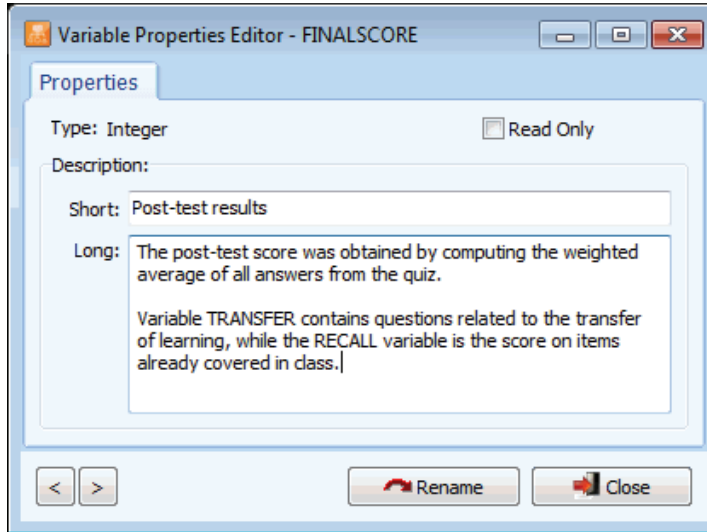
Clicking down the  button forces QDA Miner to synchronize the table with the main window so that moving from one row to another will cause the main window to display the corresponding case and document from which the selected text comes. This feature is useful if one needs to read the content of some documents in order to identify which value should be entered in a specific column.

Changes made to the spreadsheet data are not saved automatically. When a cell is modified, its background is set to a light-blue color, indicating that its value has been modified in memory but not yet applied. To apply changes made in the spreadsheet editor, click the  Apply button. To cancel all changes made to the values in the spreadsheet, click the  Close button. You will be asked to confirm the loss of all changes.

# Editing Variable Properties

QDA Miner allows you to edit various properties of existing variables in your project. For example, you can attach a short and a long description to a variable, set the number of decimal places for floating-point numerical values, edit values of categorical variables, etc. You can also set an individual variable as "read only" or change its name. To access the variable properties dialog box, use the following steps:

In the Variables windows, position the cursor on the variable that you want to edit. Choose the PROPERTIES command from the VARIABLES menu, or right-click anywhere in the Variables window and select the EDIT PROPERTIES command. This displays the **Variable Properties Editor** dialog box as shown below.



Once in the dialog box, you can navigate through variables by clicking either the  or the  button. When viewing or editing the properties of a categorical variable, a second page appears. You can add new values, as well as edit or delete existing ones on this page (see below). The first page of the dialog box offers the following options:

**READ ONLY** - When selected, this option prevents a variable from being modified. This option is useful to prevent accidental or unauthorized changes to the values of a variable. To prevent the modification of data for an entire project file, see **Security & Multi-Users Settings** on page 27. Setting a variable to **read only** only affects the editing of values in existing cases, but still allows users to create new cases and assign values to these new cases.

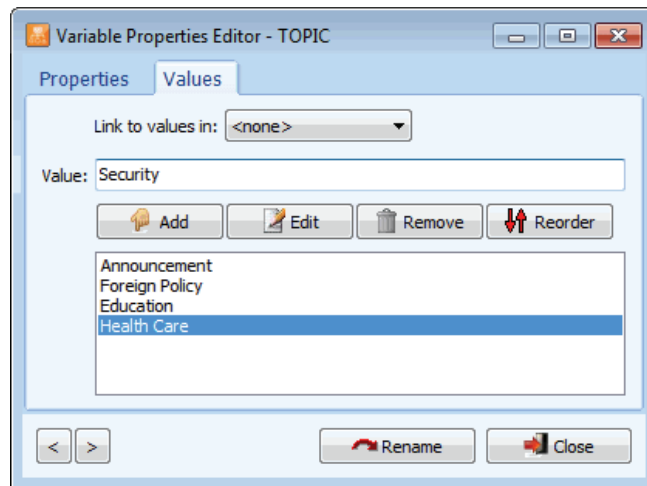
**DECIMALS** - When the variable is a floating-point number, the decimals option is used to specify how many decimal places to display in the data windows. Floating-point numbers are stored in the data file using double precision values (at least 15 significant digits). The decimals option is used exclusively to control how numeric values are displayed in the **Variables** grid and in other locations and does not affect the internal precision of the variable.

**DESCRIPTION** - This option lets you enter both a short single-line alphanumeric description as well as a detailed description of the variable. The short description is displayed in various locations in the program to help remind users of the exact content of this variable.

**RENAME BUTTON** - You can use this button to change the name of the current variable. When you click this button, you will be prompted for a new variable name. This new name must not exist in the current data file and should follow the basic rules for valid variable names.

## The Values Page

When viewing the properties of a categorical variable, a second page is shown. This **Values** page allows you to add new values, as well as edit or delete existing ones.



### To add a new value labels:

- In the Value edit box, enter the string that will be used to describe this new value.
- Click the **Add** button.

### To remove an existing value label:

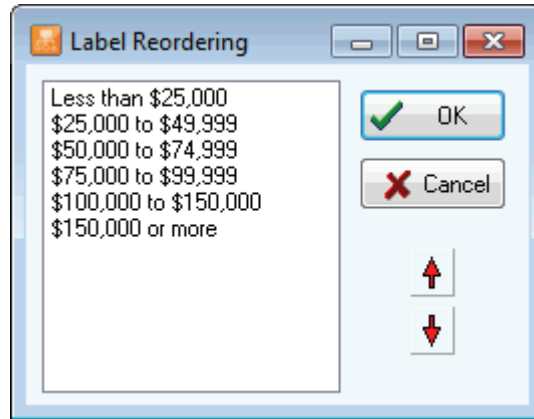
- In the list box located in the lower half of this page, select the label you want to delete.
- Click the **Remove** button

### To edit an existing value label:

- In the list box located in the lower half of the page, select the label you want to edit.
- Click the **Edit** button.
- Once you finished editing the label, click the **OK** button to apply the change.

## Reordering Value Labels

Ordinal variables assume a clear ordering of values. A good example of such a variable would be responses to a satisfaction questionnaire with values ranging from “Very Unsatisfied” to “Very Satisfied”, or an age group variable containing several age ranges going from “18 or less” to “60 or more.” To display those values in a proper order in various tables and to be able to apply some of the ordinal statistics available in QDA Miner, values of those variables should be properly ordered. The Reorder button allows one to change the natural order of values of ordinal variables. When this button is clicked, a dialog box like this one will appear:



To reorder values, simply select the value you would like to move and click the up or down arrow button until the label is in the proper order. To confirm the reordering, click **OK**. To return to the original ordering of values, click the **Cancel** button.

## Using an Existing Value Labels Definition

In some projects, several variables share the same value labels. For example, a questionnaire may use a common ordinal scale for several questions. Rather than re-entering the same value labels over and over again, QDA Miner can establish a link between a variable without value labels and an existing one that already contains labels.

To establish a link, click the down arrow button of the **Link to values in** list box, then select from the list of variables the one containing the value labels you want to use. These labels will appear in the **Value** list box.

Once a link has been established, every change made to the value labels list will affect the labels associated with the original variable as well as with all other variables currently linked to this variable.

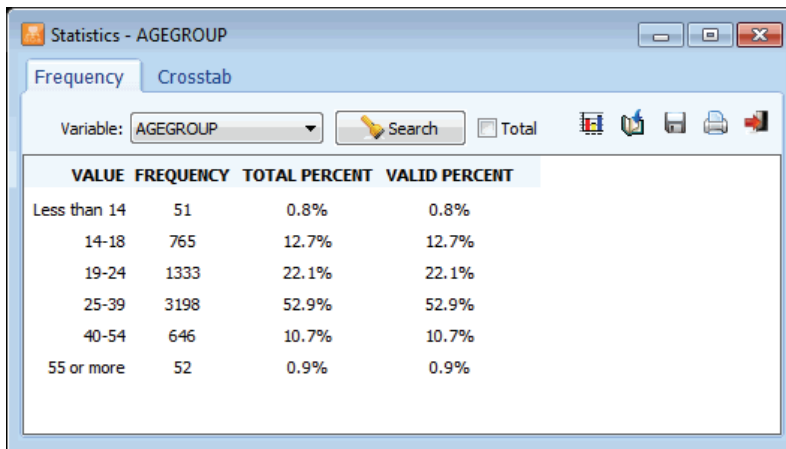
You may also use this feature to copy value labels from one variable to another. To do this, follow the previous instructions to link the current variable to the one from which you want to copy labels. The labels should appear in the value labels list. Then remove this link by setting the **Link to values in** option to none. You may now edit the newly copied labels without affecting the labels of other variables.

# Computing Variables Statistics

The VARIABLES | STATISTICS command allows one to quickly obtain the frequency and cross-frequency distribution of numerical, categorical, date and short-string variables. The univariate frequency table includes the frequency count for each value of the selected variable as well as the percentage of the count over all cases and over valid cases only. The contingency table describes the distribution of two variables simultaneously by displaying either the frequency or the row, the columns or the total percentages. Several types of charts may be created to illustrate the distribution or cross-frequency distribution of variables. You will find below a list of those charts.

## To obtain the frequency distribution of a variable

- From the main QDA Miner screen, in the Variables window, highlight the variable on which you would like to obtain distribution statistics.
- Select the STATISTICS command from the VARIABLES menu, or right-click and select the STATISTICS command.



- You may obtain a frequency table on any other numerical, categorical, alphanumerical, and date variable by selecting its name in the **Variable** list box and then clicking the **Search** button.

## To chart the distribution of a variable

Clicking the  button allows one to obtain up to five types of charts to visually display the distribution of specific codes. Some of those charts are available only for numerical and date variables (histograms and box-&-whiskers plots), while others like the bar charts and pie charts will be available in all situations when the total number of values is less than 100.



The vertical bar chart is the default chart used to display the frequencies of distinct values of a nominal or ordinal variable. It is especially useful to compare two or more values.



The horizontal bar chart displays the same information as the vertical bar chart. It is especially useful when the number of values is high and their labels cannot be displayed entirely on the bottom axis.



The pie chart is useful to display the relative frequency of each value and compare individual values to other values and to the whole. Numerical values displayed in pie charts are always expressed in percentages of either the total frequency or case occurrences.



The histogram graphically displays the distribution of a numeric variable. When selected, the program first separates the values into non-overlapping intervals of equal width, then plots bars that represent the frequencies of each interval.



The box-&-whiskers plot can be used to examine the distribution of numerical variables. It is especially useful to detect the presence of outliers and asymmetry in the data distribution. The box includes values that fall between the first and the third quartiles (about 50% of the values). The line in the middle of the box represents the median value while the whiskers extend to the farthest observations within 1.5 times the interquartile range measured from the nearest quartiles. Values that are situated farther than 1.5 times the interquartile range but within three times this distance are represented by a dot, while values farther than three times the interquartile range from the nearest quartile are represented by the letter X (for extreme).

## To obtain the joint distribution of two variables:

- Select the **STATISTICS** command, from the **VARIABLES** menu to display the **Statistics** dialog box (see above).
- Move to the **Crosstab** page. The dialog box will be similar to the one below:

|              | Female | Male  |
|--------------|--------|-------|
| Less than 14 | 0.1%   | 0.9%  |
| 14-18        | 4.9%   | 13.6% |
| 19-24        | 24.2%  | 21.6% |
| 25-39        | 55.6%  | 52.8% |
| 40-54        | 13.2%  | 10.3% |
| 55 or more   | 1.9%   | 0.7%  |

- Select from the **Tabulate** list box the variable you would like to be displayed on the rows of the table.

- Select the variable you would like to be displayed at the top of the table by selecting its name in the **With** list box.
- In the **Display** list box, select the statistics you would like to be displayed in the table.
- Click the **Search** button.

### To chart the joint distribution of two variables:

Bar charts or line charts are useful for visually comparing the joint distribution of two variables. To produce these types of charts:

- Set the **Tabulate**, **With**, and **Display** options so that the information to be viewed is displayed in the table.
- Click the  button.

For more information, see **Barchart and Line** on page 187.

### To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.

For more information on the **Report Manager**, page 230.

### To export the table to disk:

- Click the  button. A **Save File** dialog box will appear.
- In the Save As Type list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file (\*.CSV), HTML file (\*.HTM; \*.HTML), XML file (\*.XML), MS Word document (\*.DOC), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).
- Type a valid file name with the proper file extension.
- Click the **Save** button.

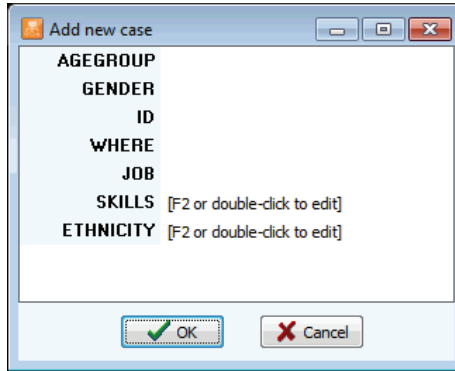
### To print the table:

- Click the  button.

# Adding and Deleting Cases

## Adding a New Case

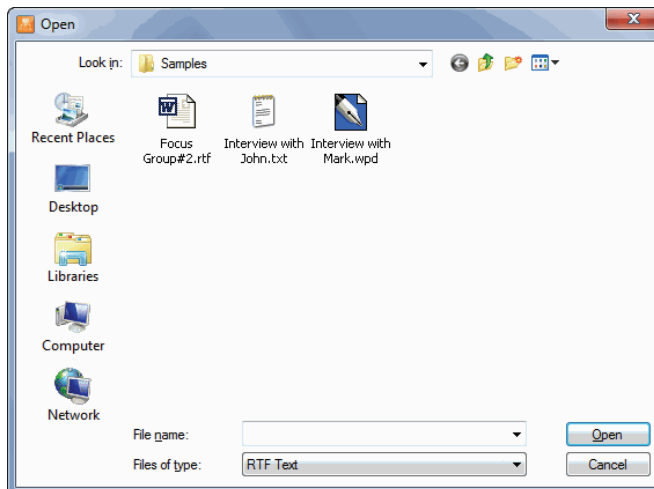
To create a new empty case, select the ADD command from the CASES menu. A data entry form will appear allowing you to enter values for each variable in the current project.

A dialog box titled "Add new case" with a list of variables: AGEGROUP, GENDER, ID, WHERE, JOB, SKILLS, and ETHNICITY. The SKILLS and ETHNICITY variables have a note "[F2 or double-click to edit]" next to them. At the bottom are "OK" and "Cancel" buttons.

To enter a new value, click the data entry cell located to the right of the variable name that you want to edit. If the variable is numeric or alphanumeric, you can start typing the data you want to store in this variable. For categorical variables, dates and Boolean values, press the **F2** key or double-click the cell to display the list of available values or open a date editor. The tab key will take you to the next variable.

To import a graphic file into an image variable, double-click the data entry cell or press **F2**. An open dialog box will appear, allowing you to select the graphic file to import.

To enter text or import an existing file into a document variable, double-click the data entry cell or press **F2**. A text editor window will appear. You can start entering the text you want to store in this document variable. You can also import an existing document by selecting the OPEN command from the FILE menu or by clicking the  button. When you select this command, an **Open** dialog box will appear.



Select the type of file you want to import. QDA Miner can read the following document types:

- RTF files (Rich Text Format)
- HTML documents
- ANSI files (or plain ASCII files)
- MS Word documents
- Windows Write documents
- WordPerfect documents
- Acrobat PDF files
- Transcriber XML file

Other file formats such as InWriter/Notetaker and PocketWord may also be supported, depending on your system installation.

Once you have finished editing the document, select the CLOSE command from the PROJECT menu or click the  button located on the upper right corner of the editor.

Click the **OK** button to confirm the creation of this new case. To exit this dialog box without creating this new case, simply click the **Cancel** button.

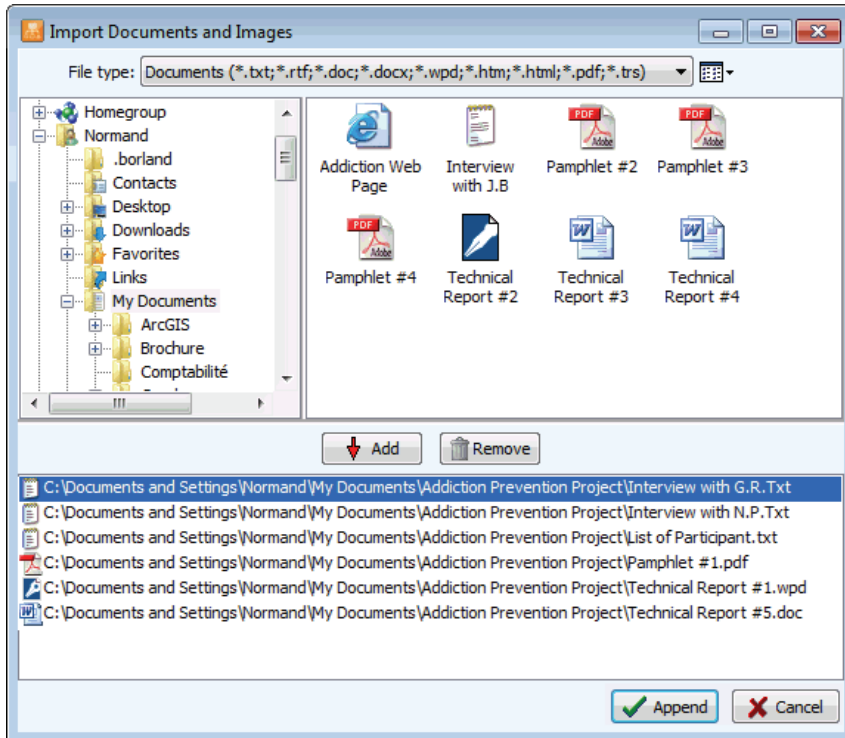
## Deleting Cases

To delete cases, select the DELETE CASES command from the CASES menu. A dialog box appears showing all cases in a list view or a tree view. Click in the appropriate box beside the cases you want to delete (check marks will appear). By default, the current case is automatically selected. Once you have selected all cases you want to delete, click the OK button.

To leave this dialog box without deleting any cases, click the CANCEL button.

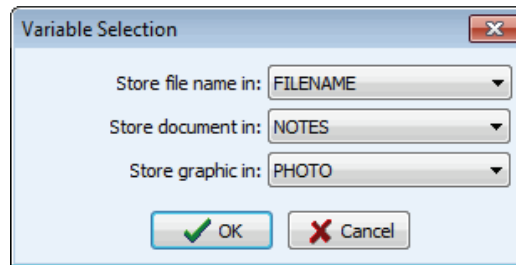
# Appending New Documents and Images

To append documents or image and store them in new cases, select the APPEND DOCUMENTS / IMAGES command from the CASES menu. A dialog box similar to the one below will appear:



- Click a folder in the folders list on the upper left section of the dialog box to display its contents. If you want to see the contents of a drive, go to the folders list, click **My Computer**, and then double-click a drive.
- In the upper right section of the dialog box, QDA Miner displays all supported document or graphic file formats that may be imported. To display only files of a specific type, set the **File Type** list box to the desired file format.
- Click the file you would like to import. To select multiple files, hold down the **CTRL** key while clicking the other files.
- Click the  **Add** button to add those files to the list of documents and images to import, located at the bottom of this dialog box. You may also drag the files from the top right section to this list.
- To remove a file from the list of file to import, select that file name and click the  **Remove** button.

- Once all files have been selected, click the **Append** button. If the project contains more than one categorical, document or image variable, a dialog box similar to this one will appear:



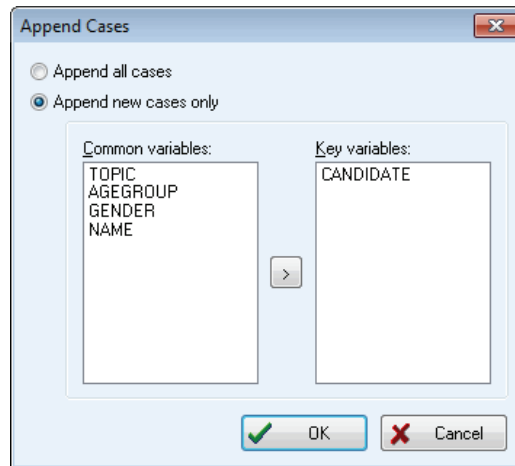
- Select the categorical variable in which the file name will be stored. Set this list box to **<none>** to prevent the program from storing this information in the project.
- Select the document variable where the imported documents should be stored.
- Select the image variable where the imported graphic should be stored.
- Click the **OK** button.

# Adding New Cases from a Data File

The **CASES | APPEND FROM A DATA FILE** command allows one to append cases stored in an external data file to the current project. In order for data to be properly imported, both data files need to share variables with identical names and compatible data types (numerous type conversions are supported). If a variable do not exists in the external file or if its type does not match or cannot be converted, the value of this variable will be set to missing. QDA Miner can import cases stored in the following file formats:

- QDA Miner projects (\*.wpj)
- Simstat data files (\*.dbf)
- MS Excel spreadsheets (\*.xls)
- MS Access data files (\*.mdb)
- Tab delimited files (\*.tab)
- Comma separated value files (\*.csv)

QDA Miner will first ask you to select the data file containing the cases to be imported. It will then display a dialog box similar to this one:



If the **Append all cases** option is chosen, QDA Miner will import all cases found in this other file and store them as new cases in the current project.

If the **Append new cases only** option is chosen, QDA Miner will require the identification of one or several key variables that will be used to differentiate already existing cases that should be ignored from new cases that will be appended to the current project.

To identify a key variable, simply select it in the **Common variables** list box and click the  button to move it to the **Key variables** list box. If a single key variable is identified, then all cases matching an existing values in the current project will be ignored while all the other ones containing new values will be imported as new cases. If several key variables are chosen, a case will be considered as existing and be ignored only if it matches values on ALL the key variables.

# Filtering Cases

The FILTER command in the CASES menu temporarily selects cases according to some logical condition. You can use this command to restrict analysis to a subsample of cases or to temporarily exclude some subjects. The filtering condition may consist of a simple expression, or include up to four expressions joined by logical operators (i.e., AND, OR).

The following table shows the various operators available for each data type:

| DATA TYPE         | AVAILABLE OPERATORS                                                                                                                                    |
|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| NOMINAL / ORDINAL | Equals<br>Does not equal<br>Is empty<br>Is not empty                                                                                                   |
| NUMERIC and DATE  | Equals<br>Does not equal<br>Is greater than<br>Is lesser than<br>Is greater than or equal to<br>Is lesser than or equal to<br>Is empty<br>Is not empty |
| BOOLEAN           | Is true<br>Is false                                                                                                                                    |
| STRING            | Contains<br>Does not contain<br>Is empty<br>Is not empty                                                                                               |
| DOCUMENT          | Is empty<br>Is not empty<br>Is coded<br>Is uncoded                                                                                                     |
| IMAGE             | Is empty<br>Is not empty                                                                                                                               |

Once a filtering expression has been entered, you can apply the filter and leave this dialog box by clicking the **Apply** button. If the filter expression is invalid, a message will appear exiting from the dialog box will not occur.

To temporarily deactivate the current filter expression, click the **Ignore** button. The filter expression will be kept in memory and may be reactivated by selecting the FILTER CASES command again and clicking **Apply**.

To store the filtering expression, click the  button and specify the name under which those filtering options will be saved.

To retrieve a previously saved filter, click the  button and select from the displayed list the name of the filter you would like to retrieve (for more information, see **Saving and Retrieving Queries** on page 108).

To exit from the dialog box and restore the previously active filtering expression, click the **Close** button.

# Creating a duplicate copy of a project

The TEAMWORK | DUPLICATE PROJECT command from the PROJECT menu provides an easy way to store an exact copy of the existing project under a different project name or into a different location.

To create a new project with an identical structure as the current project but with only a subset of cases, use the EXPORT | PROJECT FILE command instead. To create such a file, set the case-filtering conditions of the active project to display the cases as they should be stored in the new data file (see Filtering Cases) before running the exportation command (see Exporting Selected Cases).

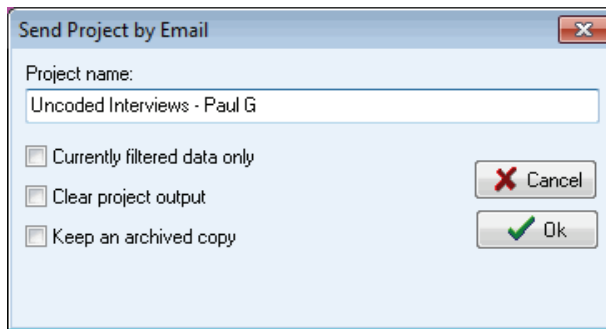
# Sending a Project by E-mail

The SEND BY EMAIL command provides an easy way to electronically send a project to another team member or to any other interested party. If a case filter has been applied, one can choose to send the whole project or only the currently filtered cases. The project may be renamed to prevent any confusion and outputs in the Report Manager may also be deleted. The custom project is then stored in a compact ZIP archive and sent to the user as an email attachment. When executed, your default e-mail program is called allowing you to enter recipient names, add any additional information, or select any other options.

## To send a project by email:

- If needed, filter the cases to display only the data that should be sent.
- Select the TEAMWORK | SEND BY EMAIL command from the PROJECT menu. A dialog box similar to the one below will appear:

- 



- If needed, edit the project name.
- If a case filter is active, enable the **Currently Filtered Data Only** option if you want to send only those selected cases, or disable it to ignore the filter and to send all cases.
- Set the **Clear Project Output** option if you want to clear the content of the Report Manager of this project.
- If you want to keep a zipped copy of the project sent to the user, select the **Keep Archived Copy** option.
- Click the **OK** button. If an archived copy was requested, you will be asked to provide a destination for the ZIP file. The default email application is called. Enter one or more recipient names and type any required information.

# Exporting Selected Cases

The PROJECT | EXPORT command saves a copy of the current project under a different name or exports the project to another file format. QDA Miner supports the following export formats:

- DBase
- Paradox
- Lotus 123
- Excel
- Quattro Pro
- Comma Separated Values
- Tab Separated Values
- Triple-S XML (interchange standard for survey data)
- XML

When exporting a data file, QDA Miner will use the current filtering condition to determine which cases will be exported. The active sorting order will also be used to control the case sequence in the new file.

## To export data to any of these applications:

- Set the filtering (see page 49) and sorting conditions (see page 57) of the active data file to display the cases as they should be exported.
- Select the EXPORT | PROJECT FILE command from the PROJECT menu.
- Select the file format you want to create using the **Save As Type** drop-down list.
- Enter a valid filename with the proper file extension.
- Click the **Save** button.

## Creating a File with Subsets of Cases

The export feature can also be used to create a new data file with a structure identical to the current data file, but with only a subset of cases.

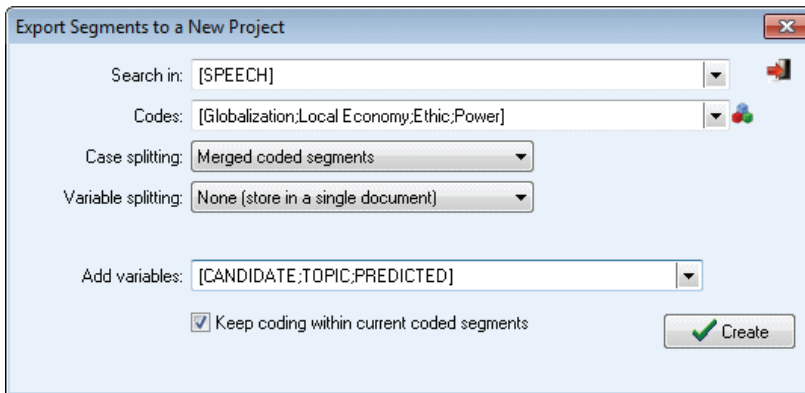
## To create a project with a subsets of cases:

- Set the filtering and sorting conditions of the active data file to display the cases as they should be saved in the new data file.
- Select the EXPORT | PROJECT FILE command from the PROJECT menu.
- Set the **Save As Type** drop-down list to QDA Miner Project
- Enter a valid filename.
- Click the **Save** button.

# Exporting Coded Segments

In some situations the format in which the data has been collected does not correspond to the design of a research project or is not appropriate for the kinds of analyses that need to be performed. For example, if one collects transcripts of group discussions and would like the unit of analysis to be the individuals, one will need to split the transcript so that interventions by each speaker will be stored in separate cases, creating for each participant a single new document of everything this person said. Another example would be when one has several structured documents but is only interested in analyzing specific sections of them and would like to eliminate unnecessary sections. One may also want to split the different sections of those documents and store each of those sections in separate document variables, allowing faster browsing through specific sections. Alternatively, one may want to regroup, under a single document, segments about a specific topic currently found in several document variables. The Export Coded Segments feature is a flexible data restructuring tool that may be used to perform all the above changes. It may also be used to transform a set of unstructured individual interviews into a structured database, storing answers to different questions into separate document variables or regrouping text about a specific topic discussed by an interviewer at different moments under a single document variable.

To perform this procedure, select the EXPORT | CODED SEGMENTS command from the PROJECT menu. A dialog box similar to this one appears:



**SEARCH IN** - This option allows you to specify which document variables the extraction should be performed in. If the current project contains more than one document variable, you will have a choice of selecting either one or several document variables. By default, all document variables are selected. To restrict the extraction to only a few of them, click the arrow key at the right of the list box. You will be presented with a drop-down list of all available document variables. Select the variables on which you want the extraction to be performed.

**CODES** - This option allows you to select the codes defining the segments that should be extracted. To select codes, first click the arrow key at the right of the list box and then select the desired codes. You may also click  button to select them from the codebook tree structure.

**CASE SPLITTING** - This option is used to specify whether the original case structure should be preserved or whether existing cases should be split into several ones based on the coding made of their documents. Three options are available. Choose **None (keep the same cases)** if you want to

keep the same number of cases. If the **Every coded segments** option is selected, the program will extract each coded segment associated with any one of the specified codes and will store the segment in a separate case. For example, if a specific document contains five coded segments associated with Code A and three coded segments associated with Code B, this single document will result in eight new cases, each one containing a single text segment. If **Merged coded segments** is chosen, a new case will be created for each selected code. This new case will contain all segments associated with this code. So, if a specific document contains five coded segments associated with Code A and three coded segments associated with Code B, this single document will result in the creating of two cases, one containing all five text segments associated with Code A and the second one containing the three text segments associated with Code B.

**VARIABLE SPLITTING** - This option can be used to specify whether extracted segments should be stored in a single document variable or in different ones. The available options depend on the setting of the **Case Splitting** option and will be any one of the following:

**None (store in a single document)** - Use this option to store all extracted segments in a single document variable named SEGMENT.

**Current document variables** - When segments are extracted from several document variables, this option will create a project with the same document variables as those specified in the **Search In** option and will store the extracted segments in the variable from which they come.

**Above codes** - This option will create for each code selected in the **Codes** option a separate document variable and will store text segments associated with specific codes in their corresponding document variable.

**Other codes** - This option is available only when the **Case Splitting** option is set to **Every Coded Segment** or **Merged Coded Segments**. It allows one to further split the text segments extracted from the first step into separate document variables based on a second list of codes. An application of this would be splitting a group discussion so that codes associated with individuals are used to create different cases and storing them in different document-variable text segments associated with different topics.

**ADD VARIABLES** - This drop-down checklist box may be used to append the values of existing variables to the exported data file along with the extracted text segments.

**KEEP CODING WITHIN CURRENT CODED SEGMENTS** - By default when extracting coded segments, codes used to extract the coded segments as well as those codes enclosed in the extracted segments are both stored along with the extracted text and are displayed in the new project as new coded segments. The current codebook is also copied to the new project file. Disabling this option removes all codes and codings from within the resulting project, and will create a project without this codebook.

# Exporting Documents

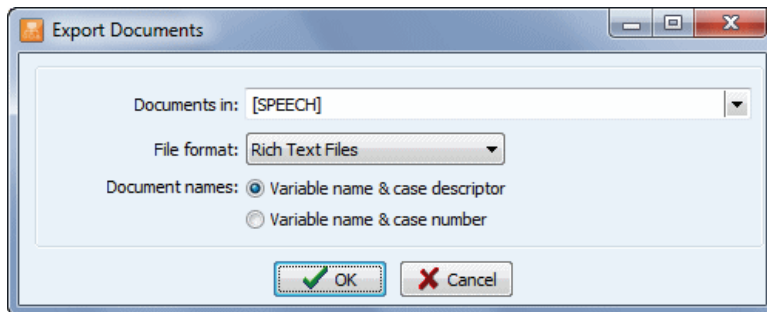
The EXPORT | DOCUMENTS command saves all documents in a project file associated with one or several document variables into separate files. QDA Miner supports the following document formats:

- Rich Text
- HTML
- Plain text file

When exporting documents, QDA Miner will use the current filtering condition to determine which documents will be exported.

## To export documents to disk:

- Set the filtering condition of the active data file to display the cases containing the documents you want to export. If you want to export all documents, remove any filtering condition.
- Select the EXPORT | DOCUMENTS command sequence from the PROJECT menu. The following dialog box will appear:



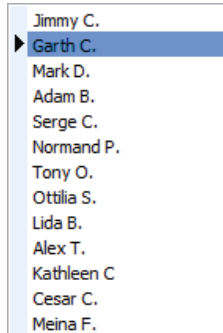
- Select the variables containing the documents you would like to export to disk.
- Set the **File Format** option to the desired format.
- In the **Document Names** option, select the method that should be used to automatically create file names.
- Click the **OK** button.

You will be asked to specify a folder under which document files should be stored.

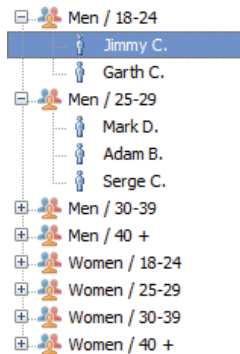
# Setting the Cases Descriptor and Grouping

The CASES window shows all currently active cases in the project. By default, cases are listed in their order of creation and are identified by their physical position in the data file (e.g. case #1, case #2, etc.). However, you can edit the list order, group cases in categories, and define a custom descriptor that will be used to identify a case based on the values contained in one or more variables.

There are two major ways to display the active cases. These cases may be displayed either in a single sorted list or grouped by category based on the values of one or two variables. The example below shows a list of cases sorted alphabetically on interviewee name.



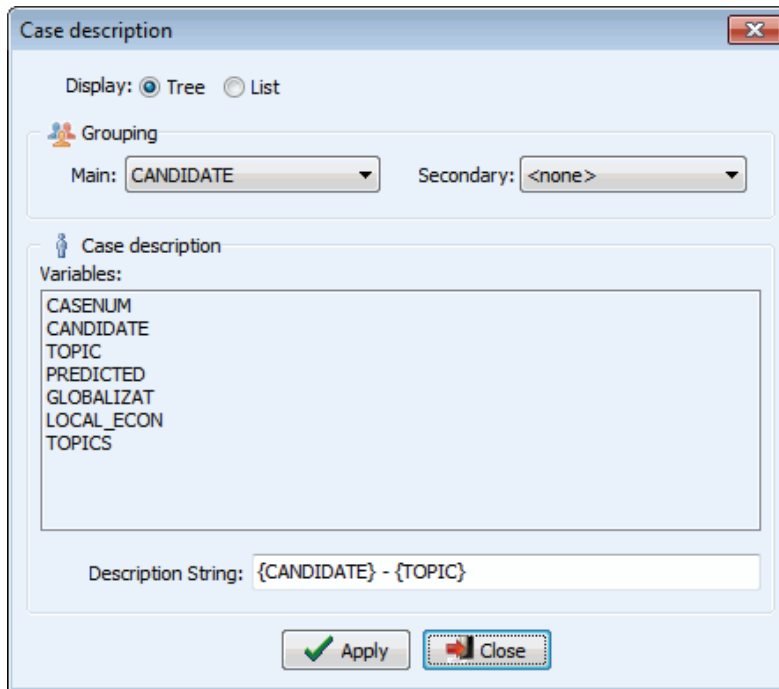
The second example below shows the same cases grouped by gender and age.



In this last example, all cases appear in a tree in which each node represents the combined value of the grouping variables. You can display or hide cases under this category by clicking the + or - sign located to the left of the node or by double-clicking the node itself. To display or edit a specific case, simply select its case descriptor.

**TIP:** Displaying cases grouped in a tree means that the program must read all cases in the project, which may cause some delays when working with large data files containing several thousand cases. For this reason, when working with such a large project, it may be preferable to display cases as a list.

To adjust the descriptors used to identify cases or to choose how these cases will be organized, select the **GROUPING/DESCRIPTOR** command from the **CASES** menu. The following dialog box will appear:



The dialog box is titled "Case description" and has a close button in the top right corner. It contains two radio buttons for "Display": "Tree" (selected) and "List". Below this is a "Grouping" section with a person icon, containing two dropdown menus: "Main:" with "CANDIDATE" selected and "Secondary:" with "<none>" selected. Below the grouping section is a "Case description" section with a person icon, containing a "Variables:" list box with the following items: CASENUM, CANDIDATE, TOPIC, PREDICTED, GLOBALIZAT, LOCAL\_ECON, and TOPICS. At the bottom of the dialog is a "Description String:" text field containing "{CANDIDATE} - {TOPIC}" and two buttons: "Apply" (with a green checkmark icon) and "Close" (with a red X icon).

**DISPLAY** - This option lets you select whether cases will be displayed in **TREE** mode or as a single **LIST**.

When you choose to display cases in a tree, you will be asked to provide up to two variables that will be used to group cases. The **MAIN** list box allows you to select the first grouping variable, while the **SECONDARY** list box allows you to further break down the cases into the subcategories defined by these two variables. For example, if you choose **GENDER** as the main grouping variable and **AGE** as the second grouping variable, and if we assume that the **AGE** variable contains only three values (16, 17 and 18), cases will be grouped under the following six categories:

MALE - 16  
MALE - 17  
MALE - 18  
FEMALE - 16  
FEMALE - 17  
FEMALE - 18

If you choose to display cases in a single list, cases will be displayed by default in order of creating. The **MAIN** and **SECONDARY** list boxes may however be used to re-order these cases and display them sorted in ascending order on one or two variables.

**CASE DESCRIPTION** - This section of the dialog box allows you to specify a label that will be used to describe each case. The label may be changed by editing the text in the **DESCRIPTION STRING** edit box. To insert the value stored in a specific variable into the description, simply enter the variable name in uppercase letters and enclose this name between braces. Alternatively, you can insert a variable name at the current caret location by clicking the corresponding item in the VARIABLES list located just above the edit box.

If you enter the following string:

`{GENDER} subject - {AGE} years old`

The {GENDER} and {AGE} strings will be replaced with their corresponding value for this specific case. If the current case contains information about a seventeen-year-old male, the above string will be displayed as:

Male subject - 17 years old

Beside the name of variables, it is also possible to insert the following string:

`{CASENUM}`

This string will display a unique case number, representing the physical order of this case in the project file.

# Importing a Document into a Case

The Document window allows users to enter text directly in the document editor or paste text from the clipboard. It is also possible to import an existing document stored on disk into the project.

## To import a document into an existing case:

- From the **Cases** window, select the case in which you would like to store the document.
- If your project contains more than one document variable per case, make sure the Document window points to the proper document by selecting its name from the list box on the title bar of this window.
- Select the **DOCUMENT FILE | IMPORT** command from the **DOCUMENT** menu or click the  button on the document editor toolbar. An **Open** dialog box will appear.
- Select the file format of the document that you want to import from the **Files of Type** list box located at the bottom of this dialog box.
- Highlight the file name that you want to import and click **Open**.

Please note that importing a document will overwrite existing text stored in the currently selected document. To append a file to an existing text, simply select this text, copy it to the clipboard, import the new document, and paste into the text back in the proper location.

## To import a document into a new case:

- Select the **ADD** command from the **CASES** drop-down menu. A data entry dialog box will appear.
- Double-click the document variable to access the text editor.
- Select the **OPEN** command from the **PROJECT** menu or click the  button on the editor toolbar. An **Open** dialog box will appear.
- Select the file format of the document you that want to import from the **Files of Type** list box located at the bottom of this dialog box.
- Highlight the file name that you want to import and click **OPEN**.
- Exit the editor by selecting the **CLOSE** command from the **FILE** menu or clicking the  button.
- After entering data for other variables, click the **OK** button.

# Importing an Image in a Case

## To import an image into an existing case:

- From the CASES window, select the case in which you would like to store the image.
- If your project contains more than one image variable per case, make sure the Document window points to the proper variable by selecting its name from the list box on the title bar of this window.
- Select the FILE | IMPORT command from the IMAGE menu or click the  button on the Image Editor toolbar. An **Open** dialog box will appear.
- Select the file format of the image that you want to import from the **Files of Type** list box located at the bottom of this dialog box.
- Highlight the file name that you want to import and click the OPEN button.

Please note that importing a graphic will overwrite the existing image stored in the currently selected variable.

## To import an image into a new case:

- Select the ADD command from the CASES drop-down menu. A data entry dialog box will appear.
- Double-click the image variable to access the Image Editor. An **Open** dialog box will appear.
- Select the file format of the image that you want to import from the **Files of Type** list box located at the bottom of this dialog box.
- Highlight the file name that you want to import and click the OPEN button.
- After entering data for other variables, click the **OK** button.

# Archiving a Project

During a research project, data files are frequently modified (cleaning, coding, etc.). At some point you may need to return to a previous version of an existing data file to recover lost variables or cases that has been transformed or deleted, or just to make routine verifications. In order to do this, several successive backup copies of this data file should be created and kept. Another reason to make backup copies of data files is to prevent the accidental loss of an entire data file caused by a hardware failure or a software malfunction.

QDA Miner provides a simple archiving procedure that allows users to quickly create backup copies. This procedure stores a copy of the currently active project and all its related files (structure definition, documents, codebook, value labels, etc.) in a single compressed file. QDA Miner uses the industry standard ZIP format as its own archive file format so that you can easily manage these archives files outside QDA Miner using any application that can manipulate these files. As well, the high level of compression achieved on typical QDA Miner data files (usually more than 90-95%) allows you to store backup copies of large projects in much smaller files and thus keep several copies of your data file on your hard drive without sacrificing precious disk space.

Another benefit of this archiving feature is to facilitate the transfer of your data file and all its related files to another computer by creating a single file that includes all associated files.

## To make an archived copy of the current data file:

- Select the MAINTENANCE | BACKUP | CREATE command sequence from the PROJECT menu. An **Export Data** dialog box will appear.
- Set the drive and directory setting to the location in which you want to store the compressed file.
- Enter a valid file name and click **OK**.

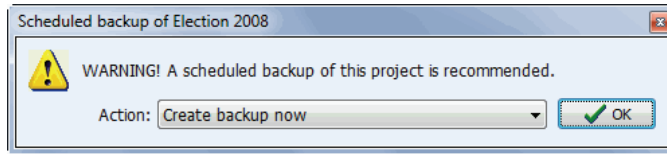
## To restore a file from an archived copy:

- Select the MAINTENANCE | BACKUP | RESTORE command sequence from the PROJECT menu. An **Open** dialog box will appear, displaying all ZIP files.
- Select the proper ZIP file and click the **OK** button. A second dialog box will appear, prompting you to specify the location in which data should be restored. The default location is the same directory as the archive file.
- If needed, change the drive and directory setting and click the **OK** button to proceed with the extraction.

## Scheduled Backups

Even though creating backups can be easy in QDA Miner, sometimes one will fail to perform those backups, forget about them or simply lose track of when a backup on a specific project has last been done. For this reason, we implemented in QDA Miner an automatic backup feature for project files. This

backup is set by default to occur one month after the last time a backup was performed, and will be prompted only if some changes have been made to the project. One can set a different time cycle for scheduled backups or disable this feature by accessing the Project Properties dialog box. When QDA Miner detects that a scheduled backup is due to run, a dialog box like the one below will appear:



This dialog box offers a choice among four types of actions:

- Selecting **Create backup now** will institute the backup creation procedure described previously and will require a backup name and location.
- The **Skip backup** option can be used to ignore the warning and delay the backup until the next time the project is accessed.
- Selecting **Postpone the backup (another cycle)** will also ignore the warning but it will reset the time counter to the current date so that the next request for backup will occur only after the defined time period.
- Finally, **Adjust the scheduled backup** will institute the Project Property dialog box, allowing one to either disable the scheduled backup feature or adjust how often the backup will be prompted.

# Using the Temporary Session Backups

QDA Miner automatically stores all changes made to the project. This ensures that no data will be lost if a power failure or a computer crash occurs. The drawback of this is that all changes made to a project are permanent, and cancelling them may sometimes require some work or may even be virtually impossible (such as when you delete a case or a variable). The temporary session backup provides a way to avoid such situations. When the temporary session backup option is enabled (see the **PROGRAM SETUP** command), QDA Miner automatically creates a temporary copy of a data file upon its opening. One may then cancel all coding, data transformation or editing performed during a session and restore the file to its original state.

Use the **REVERT CHANGES** command in the **PROJECT** menu to cancel editing operations and to restore the original project file or move back to the last time you updated this backup.

The **KEEP CHANGES** command in the **PROJECT** menu (or pressing the **CTRL-K** key combination) refreshes this temporary backup and ensures that all modifications to the project made so far during the session will not be lost if you later decide to revert to a previous version of the project.

# Merging Projects

The QDA Miner MERGE command allows one to combine two or more project files into one. This operation is achieved by opening a "master" project that will act as the recipient and then selecting external project files from which information will be retrieved. The MERGE command will perform up to 4 types of data import in a single operation.

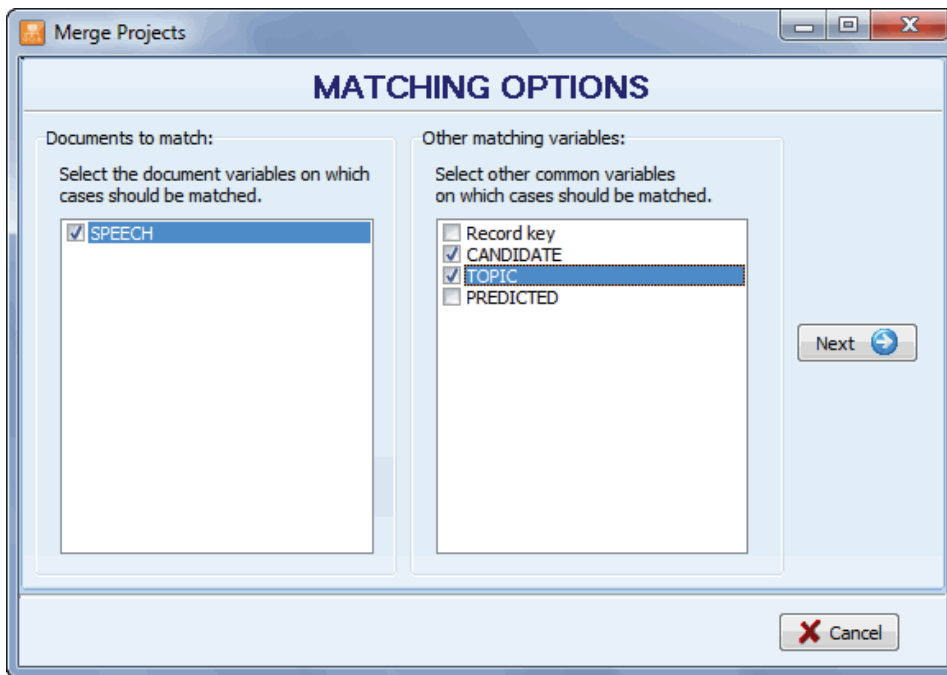
- It will attempt to match cases, documents and images and to import for common documents any code assignment that is not already in the master project.
- If necessary, it will also import the codebook from the external file and merge its codes and categories into the existing codebook (see also **Importing a Codebook** on page 82 for the single importation of codebooks from other projects).
- If the program cannot match cases in the imported project to existing ones in the main project, it can optionally append those cases along with their associated documents and codings.
- If the imported project contains unique variables, the program will allow selecting from those variables the ones that should be appended to the existing master project.

This merging feature is useful to allow one to synchronize one's own work performed on different computers. As well, both an individual or various team members may work on distinct sets of documents and then merge them along with their associated codings into a single master project.

The Merge feature may also be used by a team of researchers working on the same set of data. They can perform their own coding and editing on their own copy of a master project file, working on different computers at different times. Once individual coding and data editing has been done, they can then combine all codes assigned by each researcher into a single project file. Such a scenario may be used to separate the amount of work among several people. It is also useful to assess the reliability of coding made by different coders, by allowing different researchers to code independently the same sets of documents (for more information how to assess the reliability of coding, see **Assessing Inter-Raters Agreement** on page 203).

## To merge two project files:

- Open the project in which the information currently contained in other project files will be stored.
- Select the TEAMWORK | MERGE command from the PROJECT menu. An **Open** dialog box will appear.
- Select the second project from which to import codes, coding, cases, or variables. The following dialog box will appear:

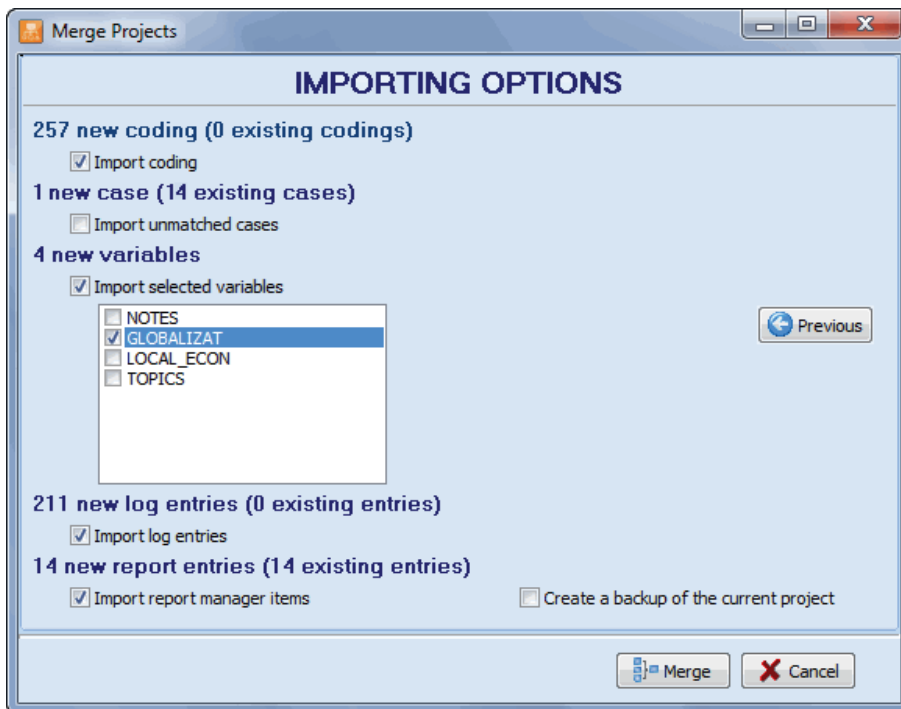


In order to merge coding made to duplicate documents stored in different projects files, QDA Miner needs to decide whether two documents are the same. This is performed by comparing the first 200 alphanumeric characters in each document (if the document is shorter, its entire text is compared). Spacing, parentheses and punctuation marks, as well as any other formatting characters including tabs and end-of-lines, are ignored. While it is unlikely that two documents will start exactly the same way, such a situation may still occur, especially when dealing with very short responses. To prevent such a situation from occurring, it is possible to instruct the program to consider a match when two cases share identical values on any other variables found in both project files.

The first list box on the left side of the dialog box is used to specify on which documents and image this pairing should be made. Since coding may only be imported for documents that have been previously matched, this list box also specifies where the codings to be imported are located.

After selecting at least one document/image variable, select from the list of all variables that are common to both files any other variable on which the pairing of cases should be made. Two cases will be considered identical only if all the selected documents/images and all the values of those selected variables are the same.

Once the matching options are set, click the  button. This will instruct QDA Miner to apply the matching criteria, check for possible conflicts, and then move to the second page of the dialog box. If two records in the master project share the same values on all selected variables, QDA Miner will display an error message stating that the master project contains duplicate entries. If the selected key variables are unique to each case, then the program will move on to the next page.



The second page of the dialog box previews the predicted outcome of the merging operation and specifies what should be imported. Depending on the nature of the information stored in the second project file, up to file import options will be available:

**Import coding** - While attempting to match cases on common documents, QDA Miner also compares all the codings applied to those documents. If it finds codings that do not exist in the master project, it will offer the option to import those codings. The number of new codings that would result from a merge is reported at the top of the second page of the dialog box, as is the number of common codings found in both project files. Select the IMPORT CODING box to instruct QDA Miner to import those additional codings. Note that this option only applies to documents contained in both project files. Any new document imported by the addition of new cases or new variables will always be merged into the current project along with all its associated codings.

**Import unmatched cases** - After attempting to pair all cases in both project files, QDA Miner will report the number of cases in the second project for which no match has been found. Select this option to append those cases at the end of the current project.

**Import selected variables** - If some variables are unique to the imported project file, it will be possible to integrate all or some of those unique variables into the main project by enabling this option and selecting from the list of variables located below those that should be imported.

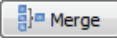
**Import log entries** - Log entries stored in the command log may be appended to the current log file.

**Import Report Manager items** - If some quotes, tables and charts have been added to the Report Manager from within the secondary project, the number of new items will be listed here. Importing those entries will append them to the current project and will be accessible from the Report Manager. However, please note that modified versions of items already in the project will not be imported.

Merging several projects into a master file may result in several important changes to the original project file. Some of those changes may not correspond to what was expected. Since it could be very difficult to reverse all the changes resulting from a merge and recover the original project file as it was before this operation, it is highly recommended either to perform this operation on a copy of the original project file or to create a backup of your original project before merging any data into it. While the creation of backups may be performed at any time in QDA Miner by using the MAINTENANCE | BACKUP command, an option has been added to this dialog box to facilitate the access to the BACKUP command. Enabling this option will instruct the program to first perform a backup of the current project file before importing begins.

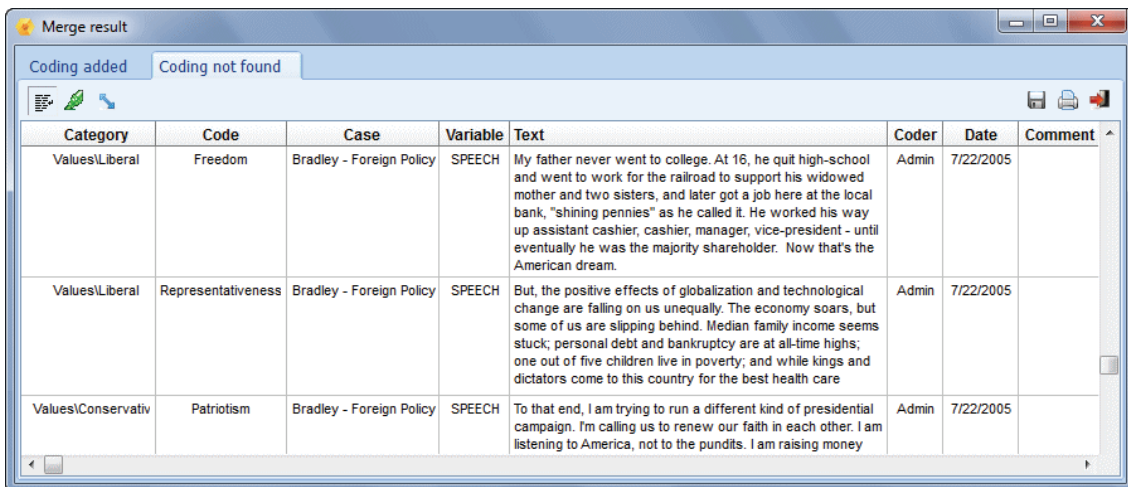
To cancel the operation and return to the QDA Miner main screen, click the  button.

To go back to the first page of the dialog box to adjust matching settings, click the  button.

Once ready, click the  button to begin merging the projects. If the backup option has been set, you will be presented with a dialog box that will ask you to specify the name of the backup file. The program will then proceed to the import of all the information and will display a message summarizing the changes made to the master project.

## Reviewing imported coding

When QDA Miner imports codings, it attempts to match the original text and location of a coding in the imported project to a segment of the master document that points to the exact same text and the same or nearby position. However, it may sometimes fail to find the proper location, usually because one of the two documents has been modified. After importing codings, QDA Miner displays a dialog box with a list of all codings added to existing documents as well as any coding that could not be assigned to the documents in the master project (see: "Coding not found" tab below).

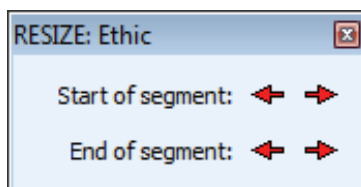


The first tab of this dialog box displays all the codings that have been added to the master project. The second tab displays any coding which QDA Miner was unable to locate. This second table of unmatched coding includes the text to which this code had been originally assigned in order to make the match manually.

Selecting an item in this table, either by clicking it or by using the keyboard cursor keys (such as the UP or DOWN arrow keys), will automatically display a portion of the document text located near the code. If slight changes have been made to either version of the document, the highlighted segment in the main window should be relatively close to the correct location.

### To manually adjust the location of an unmatched code:

- Click the  button. The selection resizing tool will appear:



- From this dialog box, click the corresponding arrow button to move the starting or ending location of the selected text in the main window. Clicking a < button moves the chosen limit of the selection to the left and up while clicking a > button moves this limit to the right and down. Repeat the same move several times by clicking and holding the mouse button down. Release the mouse button when done.
- Once the selection has been adjusted to closely match the text originally associated with this code, click the  button to assign this code to the new selection.

# Clearing all code assignments

A researcher may want to start over with a fresh uncoded project or create a copy of the original project without any coding in order to allow someone else to code the same set of documents. To clear all codings made to the current project:

- Select the MAINTENANCE | CLEAR ALL CODINGS command from the PROJECT menu, and click **Yes** to confirm the deletion of all code assignments.
- As an extra precaution, a dialog box will be displayed requiring the typing of the word **YES** in an edit box. Typing **YES** and clicking **OK** will result in the removal of all codings and their associated memos.

To remove specific codings, see **Removing Coded Segments** on page 143.

# Retrieving Misplaced Codes

When a code is assigned to a selected segment of a document, the code itself is stored within this document at the exact location where it was assigned. The associated coding will always be preserved, even if a portion of text or an entire document is copied to the clipboard and pasted somewhere else (whether in the original project or another one) or if it is saved to a disk as an RTF document. This method of storing codings instead of keeping an index of their location has the advantage of allowing one to modify the original document without the risk of corrupting the coding index. However, one inconvenience is the possible insertion within a project of documents already containing codings for which there is no corresponding code in the current project codebook. A software or hardware malfunction causing an abrupt termination of the program may also result in the loss of some codes in the codebook or in some codings being left without a corresponding code.

The RETRIEVE MISPLACED CODES command instructs QDA Miner to browse through all the documents in the current project and search for any coding without an associated code in the codebook and to append any new code to the existing codebook. All new codes will be stored under a special category named **Recovered**.

To run this command, select the MAINTENANCE | RETRIEVE MISPLACED CODES command from the PROJECT menu. The program immediately browses through all documents and reports the number of codes that have been retrieved.

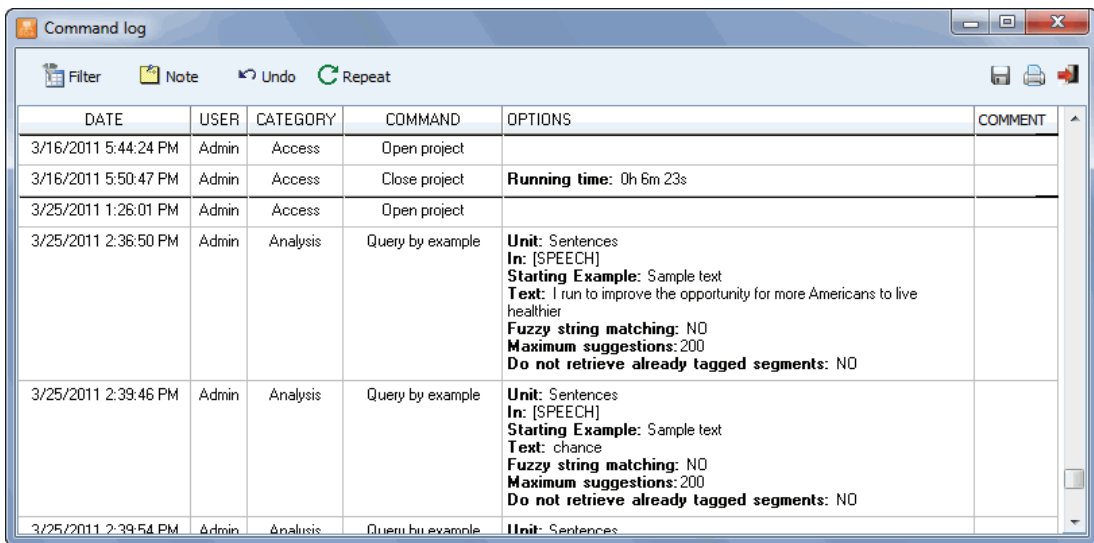
If the retrieved codes are obsolete or no longer valid, simply erase the code in the codebook, which will result in the removal of all codings associated with it (see **Deleting Existing Codes** on page 77).

# Using the Command Log

The COMMAND LOG option allows one to keep traces of almost every project access and operations performed (codebook editing, case and variable transformation, analysis, etc.). Comments may also be added to specific log entries. Such a log may serve as a part an audit trail, and used to repeat previously performed queries or analysis, or to reverse some operations (code moving, splitting of codes, autocoding, etc.).

The command entries stored in the log can be set by the administrator by adjusting the Command Log options in the **Project Properties** dialog box. Each entry includes the date and time the operation was performed, the name of the logged user, the broad category of the command, and the name and options of this command.

To access the log, simply run the COMMAND LOG command from the PROJECT menu. A dialog box similar to this one will appear:



| DATE                 | USER  | CATEGORY | COMMAND          | OPTIONS                                                                                                                                                                                                                                                 | COMMENT |
|----------------------|-------|----------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| 3/16/2011 5:44:24 PM | Admin | Access   | Open project     |                                                                                                                                                                                                                                                         |         |
| 3/16/2011 5:50:47 PM | Admin | Access   | Close project    | Running time: 0h 6m 23s                                                                                                                                                                                                                                 |         |
| 3/25/2011 1:26:01 PM | Admin | Access   | Open project     |                                                                                                                                                                                                                                                         |         |
| 3/25/2011 2:36:50 PM | Admin | Analysis | Query by example | Unit: Sentences<br>In: [SPEECH]<br>Starting Example: Sample text<br>Text: I run to improve the opportunity for more Americans to live healthier<br>Fuzzy string matching: NO<br>Maximum suggestions: 200<br>Do not retrieve already tagged segments: NO |         |
| 3/25/2011 2:39:46 PM | Admin | Analysis | Query by example | Unit: Sentences<br>In: [SPEECH]<br>Starting Example: Sample text<br>Text: chance<br>Fuzzy string matching: NO<br>Maximum suggestions: 200<br>Do not retrieve already tagged segments: NO                                                                |         |
| 3/25/2011 2:39:54 PM | Admin | Analysis | Query by example | Unit: Sentences                                                                                                                                                                                                                                         |         |

Click the  button to select the information displayed in the table. When depressed a filtering panel will appear at the bottom of the toolbar with three list box that will allow you to select log entries associated with a specific **User**, a single **Category** of commands, or a specific **Command**. To disable any filtering condition, set it to <all>.

Clicking the  button toggles the display of an edit box at the bottom of the dialog box. This edit box may be used to enter a comment or edit an existing one. To attach a comment to a specific log entry, simply move to this log entry and type the comment in the edit box at the bottom of the dialog box.

Some operations may also be reversed by selecting them and clicking the  button. The following operations may be reversed (under some conditions).

- Manual assignment of a code (from the main screen or from any hit lists)
- Autocoding performed from any retrieval tools
- Manual recoding of a coded segment
- Search and replace codes
- Add variables
- Rename a variable
- Rename a code or category
- Move a code or category

Note: for canceling other operations, see the Using **Temporary Session Backups**, on page 64.

Some logged commands may be executed again by selecting them and clicking the  button. When a command is executed this way, the original options are restored and the dialog box is displayed, allowing you to edit the original option before executing the command. The following operations may be recalled from the log entries:

- Analysis / Text Retrieval
- Analysis / Section Retrieval
- Analysis / Query by Example
- Analysis / Keyword Retrieval
- Analysis / Coding Frequency
- Analysis / Coding Retrieval
- Analysis / Coding Co-occurrences
- Analysis / Coding Sequences
- Analysis / Coding by Variables
- Analysis / Coding Agreement
- Analysis / Content Analysis
- Project / Export / Code Statistics
- Project / Export / Documents
- Cases / Filter
- Variables / Compute
- Codes / Consolidate Codes

### To export the log table to disk:

- Click the  button. A Save File dialog box will appear.
- In the list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), tab delimited file (\*.TAB), comma delimited file (\*.CSV), MS

Word (\*.DOC), HTML (\*.HTM; \*.HTML), XML (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLX), and SPSS data file (.SAV).

- Type a valid file name with the proper file extension.
- Click the **Save** button.

### To print the log table:

- Click the  button.

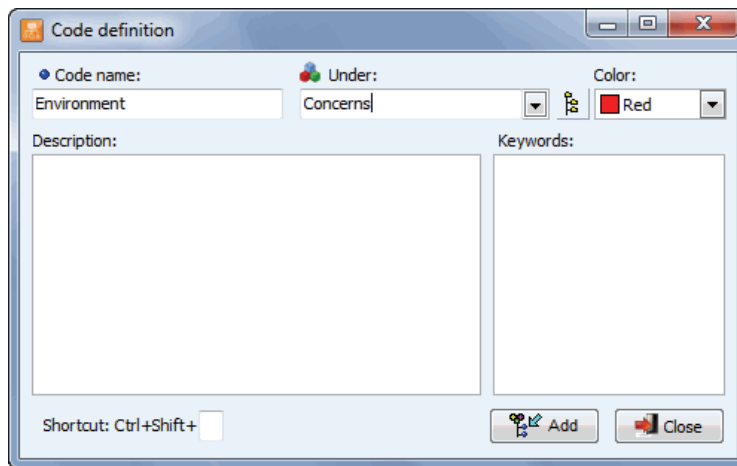
# Managing the Codebook

The main function of QDA Miner is to assign codes to selected text segments and then analyze these codes. Available codes associated with a project are located in the CODES window in the lower left corner of the application workspace. Codes are grouped into categories. The list of available codes, known as the codebook, is displayed as a tree structure, in which categories are the nodes under which associated codes can be found. You can expand a category to view all associated codes or close it by clicking the plus or minus sign located to the left of the category name. When a new project is created, this window is empty.

This section provides basic instructions on how to add new codes, organize them into categories and edit and delete existing ones. It also provides instructions on how to change the location of a code in the codebook, how to merge a code into another and how to import an entire codebook from another project

## Adding a Code

To add a new code to the existing codebook, select the ADD command from the CODES menu. The following dialog box will appear:

The image shows a 'Code definition' dialog box with a light blue border and standard window controls (minimize, maximize, close) in the top right. Inside the dialog, there are three input fields at the top: 'Code name:' with the text 'Environment', 'Under:' with a dropdown menu showing 'Concerns', and 'Color:' with a red square and the text 'Red'. Below these are two large text areas: 'Description:' on the left and 'Keywords:' on the right. At the bottom left, there is a 'Shortcut:' label followed by 'Ctrl+Shift+' and an empty text box. At the bottom right, there are two buttons: 'Add' with a green plus icon and 'Close' with a red minus icon.

When you create a new code, you must (i) specify a unique name and (ii) select a category.

The **UNDER** list box allows you to select the category under which this code will be stored. This control can be used both as an edit box to create a new category and as a list box from which you can select an existing category.

If you want to add the new code to an existing category, select the category name from the list of available categories by clicking the down arrow key to the right of the list box and by selecting the category name.

To add the new code under a new category, simply enter the name of the new category. By default, the new category is listed at the top level of the codebook. To create a subcategory underneath another one,

type its full path separating each level by a backslash. For example, typing `Topics\Economy` will create a subcategory Economy under the Topics main category. If this top-level category does not exist, the program will create both this category and its subcategory and then add the specified code into this subcategory. To store a new subcategory into an existing path, one may also type this new category name and then click the  button, which will display the tree view of the codebook categories. From that, select the parent category under which this new category will be stored. Selecting any existing category from this tree view automatically inserts the full path before the new category name in the edit box.

The maximum number of levels a codebook can contain is set to 4 by default for every new project. Since the last level can only contain codes, the actual maximum level of a subcategory is equal to this maximum minus 1. To increase or decrease the maximum number of levels allowed in a project codebook, run the **PROPERTIES** command from the **PROJECT** menu, and set the **Maximum Levels in the Codebook** option to a number between 2 and 8.

When a code is assigned to a text segment, a bracket appears in the **MARGIN** of the document to indicate the beginning and end of the code along with the code name. The **COLOR** option can be used to select the color of this bracket and of its associated label.

The **DESCRIPTION** option allows you to enter a definition or a detailed description of the code. You can use this section to specify coding instructions for the coders, along with examples or non-examples, and related codes that may be used in place of or in conjunction with this code.

The **KEYWORDS** section allows you to enter words, word patterns, or phrases that are characteristic of the code. QDA Miner may then search text for any of those items using the **KEYWORD RETRIEVAL** command and retrieve sentences, paragraphs or documents containing any of those items (see Keyword Retrieval on page 123). The following rules should be followed when entering keywords:

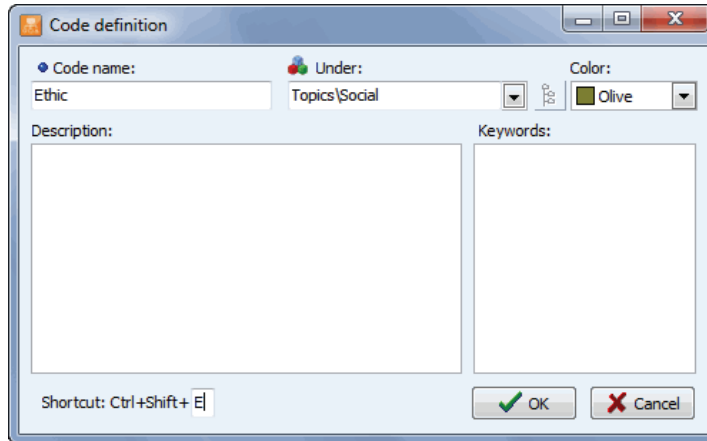
- Each item must be entered on a separate line.
- The underline character is used to join words of a phrase. Since space characters are not allowed in keyword entries, they are automatically transformed into underline characters.
- Wildcards such as `*` and `?` are supported. For example, the following entry `REPORT*` will match words like `REPORT`, `REPORTS`, `REPORTERS`, while the word pattern `EXP?RT` will match both `EXPERT` and `EXPORT`.

The **SHORTCUT** option at the bottom of the dialog box allows you to assign a keyboard shortcut to a code. This shortcut can be used to quickly assign code to whole paragraphs in the main editor without the need to highlight the paragraph and select the desired code in the codebook. To assign a shortcut, move to the shortcut edit box and type either a letter or a numeric character. If the typed character has already been assigned to another code, an error message will indicate which code has already received this shortcut assignment.

To use the shortcut, move the editing cursor in the main editor until it is located in the paragraph you want to code. Then press and hold down the **CTRL** and the **SHIFT** keys along with the character that has been assigned to the code. The same key combination can also be used in hit tables produced by the various retrieval functions of QDA Miner to code text associated with a specific hit.

## Modifying Codes

To modify an existing code or category, simply select the code or the category that you want to edit and then select the EDIT command from the CODES menu. If the currently selected item is a category, you will be prompted to provide a new name for this category. If a code is selected, the CODE DEFINITION window will appear, allowing you to change the code name, its category, its description, its keywords, its keyboard shortcut, and the color used to identify segments that has been assigned to this code:



Once finished, you can click **OK** to apply the changes, or **Cancel** if you do not want to save any changes.

Changing the name of a code will trigger QDA Miner into browsing all documents in the active project in order to apply this change.

## Deleting Existing Codes

To delete an existing code, simply select the codes you want to delete and then select the DELETE command from the CODES menu. You can also delete all codes associated with a category by selecting this category instead of a code. If you attempt to delete a category, you will be prompted to confirm the deletion of all codes within this category.

Deleting one or more codes from the codebook results in the removal of the corresponding codes from all documents associated with the current project. If you want to delete an existing code but assign another code to all segments currently associated with this code, use the code merging feature (see below).

## Moving Codes

Codes and categories in the codebook may be freely moved in the codebook using one of the following methods:

### To move a code using drag & drop:

- Click the code or category that you want to move, holding down your mouse button.
- Drag the item to its new location and release the mouse button. How the dragged item is inserted in the target position depends on the exact position where you drop it.
  - If the item is dropped on a category name, it will be stored as the last item under this category.
  - Dropping a code over another code name inserts the dropped item before the target code.

### To move a category and its items using drag & drop:

- Click the category that you want to move, holding down your mouse button.
- Drag the category to its new location and release the mouse button. If the item is dropped on another category name, it will be inserted before the target category. If it is dropped over a code, the item will become a subcategory of this code category.

QDA Miner will not allow a category to be dropped in a location underneath it. It will also forbid dropping a category at a location if the number of levels resulting from such a move exceeds the highest level allowed for this project codebook.

To increase or decrease the maximum number of levels allowed in a project codebook, run the **PROPERTIES** command from the **PROJECT** menu, and set the **Maximum Levels in the Codebook** option to a number between 2 and 8.

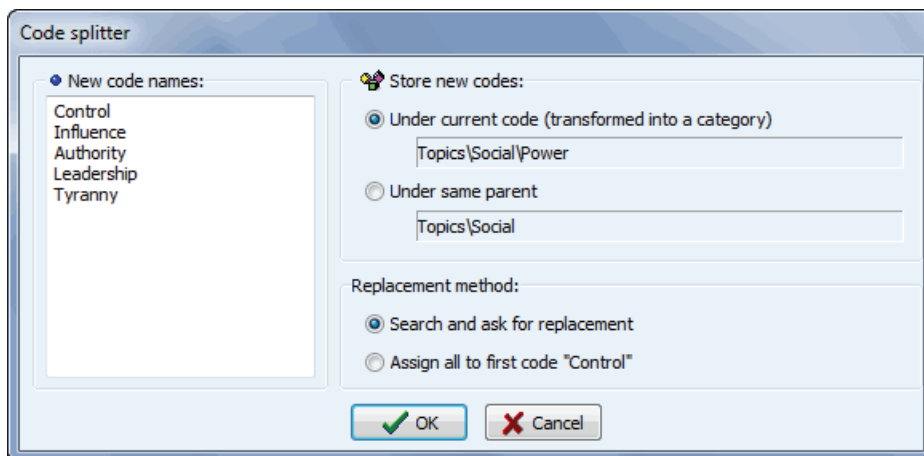
## Splitting a Code

Sometimes, one may feel the need to refine a code that has been created at an early stage of analysis in order to make an important distinction one didn't initially consider or to introduce some nuance that would allow a finer-grained analysis. Frequency analysis of codings may also reveal that a code definition is too broad and occurs too often to be of any value for the interpretation of the data. The splitting code function of QDA Miner can be used to solve this. The procedure allows one to subcategorize a code into several other codes in a single step that involves:

1. The transformation of the initial code into a category.
2. The creation of the codes underneath this category
3. The assignment of all existing text segments associated with the initial code to either one of these new codes.

## To split a code into several ones:

- In the CODES window, select the code you want to split.
- Select the SPLIT CODE command from the CODES menu. A dialog box similar to the one below will appear:

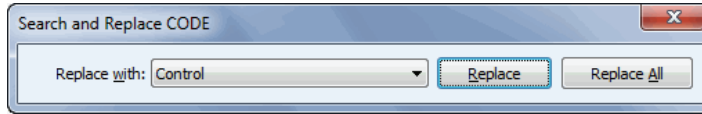


- In the **New Code Names** edit box, enter the name of all the new codes into which the initial code could be recoded, one code name per line.
- In the **Store New Codes** section of the dialog box, specify where those new codes will be stored. By default, the initial code is transformed into a category, and the new codes are stored under this newly created category. This can be achieved by selecting the **Under Current Code** radio button. One may instead wish to store all new codes in the same location as the initial code by selecting the **Under Same Parent** radio button. Please note that when the code to be split is at the deepest level allowed by the project setting, only this second option will be allowed. In that case, to make sure all new items will be grouped under a new category, one can either move the code to a higher level of the hierarchy prior to the execution of this split command (see **Moving Codes**, page 78), or increase the maximum number of levels allowed in this project codebook.

Next, select which replacement method will be used to replace all coded segments currently tagged with the initial code. Two options are available: (1) The program can search among all documents for any coded segment with the split code, display it and, for each coded segment that is found, ask the user to choose among all the new codes the one that should be attached to this segment; or (2) Choose to immediately replace all coded segments with the first code specified in the list of new codes. The user may return later to the coded segments associated with this first code in order to review those and, if needed, modify the code associated with those segments.

- Click **OK** to proceed to the splitting of the code.

If the **Search and Ask for Replacement** option was selected, the program will search among all documents and all cases for text segments that have been tagged with the initial code that has just been split. It will stop at the first occurrence, highlight the text segment, and display a dialog box similar to the one below.



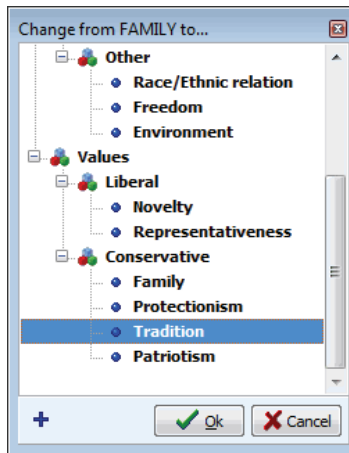
In the Replace With list box, select the code that is appropriate to this text segment and click the **Replace** button. The program will then search for the next occurrence and stop again, allowing one to select another code and assign it to this other text segment. The program proceeds this way until all text segments associated with the initial code have been assigned a new code. However, one may choose to stop this interactive search-and-replace procedure by choosing a code and clicking the **Replace All** button, instructing the program to replace the initial code for all remaining text segments with a new code. The user may then return later to the coded segments associated with this code to review those and, if needed, modify the code associated with those segments, either individually by selecting their code marks and using the RECODE INTO command, or globally using the SEARCH & REPLACE command from the CODE menu (see **Searching and Replacing Codes**, page 95).

## Merging Codes

During the analysis process, you may choose to merge several codes together. QDA Miner offers you the possibility of merging a single code into another or of merging all codes in a single category into one code. When merging a code into another one, QDA Miner removes the first code from the codebook and automatically recodes all text segments tagged with the first code by giving them a new code. For example, if you decide to merge the code FAMILY into TRADITION, QDA Miner will delete FAMILY from the codebook. It will then search for all documents in the project for text segments that have already been tagged with the FAMILY code and change the assigned code to TRADITION. When merging all codes of one category, the program removes all those codes, transforms the category into a new code, and then automatically recodes all text segments tagged with any of the deleted ones by this new code. If a segment has been tagged using two or more of the merged codes or when two of those coded segments overlap each other, the coded segments are automatically consolidated so that only one tag is attached to the text segments.

### To merge a code into another one:

- In the CODES window, select the code you want to eliminate.
- Select the MERGE INTO command from the CODES menu. The following dialog box will appear:



- Select which code you want to attach to text segments already assigned to the code you have decided to delete.
- Click the **OK** button to proceed.

## Virtually grouping codes

Codebook categories are used mainly to organize codes in a hierarchical way. By default, categories cannot be assigned to text segments and frequency statistics can only be computed on codes, not on their categories. This restriction has been implemented to maintain the independence of observations required by some routines available in QDA Miner, like clustering or multidimensional scaling. However, one may sometimes wish to obtain descriptive statistics at a category level. While one could transform codes within a category into a single code by merging all codes underneath this category (using the CODES | MERGE command), such a solution results in a permanent loss of information since all coding made using those codes are replaced with a single code. The VIRTUAL GROUPING function allows one to temporarily group all codes in a category and instruct QDA Miner to compute frequencies and obtain statistics at the category level without reporting statistics on codes underneath it. For example, if a category named POSITIVE contains five codes, by default, the CODING FREQUENCIES command will report statistics on each of those. If you virtually group codes in this category, the same command will no longer display statistics to each code, but only a single line of statistics for the POSITIVE category. Export coding statistics will also result in a single variable containing frequency data for the whole category.

The VIRTUAL MERGE function will affect the following features of QDA Miner:

- ANALYSIS | CODING FREQUENCIES
- ANALYSIS | CODING CO-OCCURRENCES
- ANALYSIS | CODING SEQUENCES
- ANALYSIS | CODING BY VARIABLES
- ANALYSIS | CODING AGREEMENT
- VARIABLES | COMPUTE COMMAND
- FILE | EXPORT | STATISTICS

The presence of virtual groupings will not affect the text or coding retrieval features, as well as any coding or other codebook management procedures.

### To virtually group codes into their parent category:

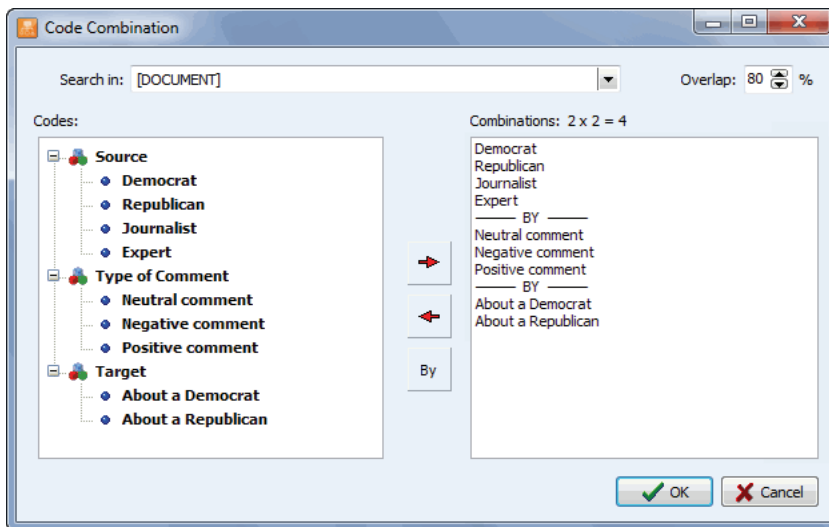
- In the codebook, highlight the category on which you would like to obtain statistics.
- Run the CODES | VIRTUAL GROUPING command. The category icon will be transformed into a  icon while icons of all codes underneath it will be replaced with a  icon.

### To ungroup codes in a virtually grouped category:

- In the codebook, highlight the category you would like to ungroup.
- Run the CODES | VIRTUAL GROUPING command. The original icons of the highlight category and codes will be restored, indicating that they are no longer grouped.

## Combining Codes

Quite often, a codebook will be multidimensional with several aspects of a phenomenon being coded independently. For example, if you are analyzing communication processes, you may code the source of a comment (the speaker), its target or the specific topic discussed as well as the tone of the message (positive, negative or neutral). In such situations, the main interest lies not in the appearance of a specific code but in the combination of codes from those different dimensions. QDA Miner offers various tools to analyze code co-occurrences and code sequences or to retrieve codes overlapping or close to each other. The CODES | COMBINE command allows one to join codes from multiple dimensions and create new compound codes that combine all those dimensions. The coding is performed using a dialog box similar to the one below:



The left side of the dialog box first shows all current codes in your codebook. To combine those, you select codes associated with the first dimension and then click the  button to move them to the right panel.

You then click the  button to separate codes that belong to the second dimension and continue until all codes from all the various dimensions have been added. In the example above, codes from three different dimensions were used. The number of compound codes that will be created will be equal to the product of the number of codes in each dimension. In the above example, we combine four source codes by three tones by two target codes to produce a total of 24 new compound codes ( $4 \times 3 \times 2$ ). Those new codes will be created and stored in a separate category and will be named by linking the name of their component codes.

In order to be combined in such a way, codes should overlap. The amount of overlap that is used as the criteria for deciding whether they co-occur can be adjusted by setting the **Overlap** option. The **Search In** option can also be used to restrict the combination to specific document variables.

Clicking the **OK** button will add all the compound codes to the codebook and then apply those codes every time every one of their component codes overlap each other.

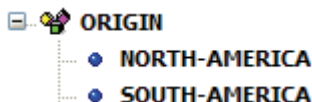
## Importing a Codebook

Quite often codebooks developed for a specific project may be used again for other coding purposes. QDA Miner can import a codebook and merge its codes and categories into the existing codebook. When merging codebooks, codes or categories whose names already exist in the current codebook are simply ignored.

WordStat categorization dictionaries may also be imported and transformed into the codebook. When importing a content analysis dictionary, low-level categories are automatically transformed into codes and their items are imported as keywords. Higher-level categories containing other content categories are transformed into new categories. For example, the importation of the following WordStat dictionary:



will create two codes, NORTH-AMERICA and SOUTH-AMERICA, under the ORIGIN category:



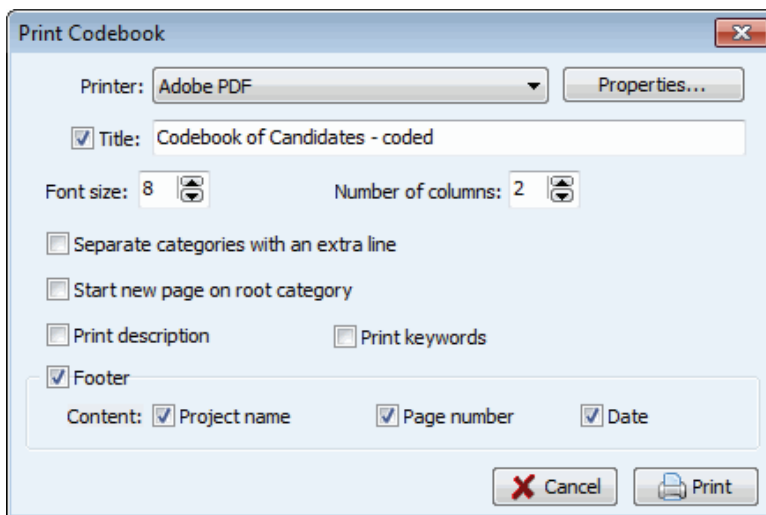
Items stored under the two codes will automatically be imported as keywords.

### To import an existing codebook or a WordStat dictionary:

- Select the IMPORT command from the CODES menu. An **Open** dialog box will appear.
- Select the .WPJ project file containing the codebook or the .CAT file that you want to import and click the **Open** button.
- You will be asked to confirm the merging of imported codes into the current codebook. Click **Yes** to confirm the merging of codes.

## Printing the Codebook

To create a printed version of the codebook, select the PRINT CODEBOOK command from the CODES menu. A Print dialog box like the one below will appear, allowing you to set various options of what and how the codebook should be printed.



**PRINTER** - Select the desired printer. To adjust the printer settings, such as the page size and orientation or printer resolution, click the **PROPERTIES** button.

**TITLE** - Use this option to specify a particular line of text that will appear at the top of each page.

**FONT SIZE** - This option is to adjust the font size used to print the dictionary items.

**NUMBER OF COLUMNS** - The codebook may be printed with up to seven columns per page, allowing one to print large codebooks on fewer pages. Please note that when increasing the number of columns per page, it may be necessary to decrease the font size to prevent the overlapping of items in adjacent columns.

**SEPARATE CATEGORIES WITH AN EXTRA LINE** - Setting this option skips a line every time a category name is encountered, allowing it to clearly delineate groups of codes belonging to the same category.

**START NEW PAGE ON ROOT CATEGORY** - Root categories are the codebook categories still visible when the codebook is fully collapsed. Selecting this option instructs the program to start the printing of all items starting with each root category at the top of a new page.

**FOOTER** - Enable this option to print a footer at the bottom of each page. A footer can consist of up to three items: The **Project Name** (printed on the left margin of the footer), the **Page Number** (located at the bottom center of the page) and the (printed on the right margin of the footer).

**PRINT DESCRIPTION** - Select this option to print not only the code names, but also their descriptions. The description is printed in italics just below the code name.

**PRINT KEYWORDS** - This option prints the keywords and key phrases associated with each code.

# Coding Documents and Images

The simplest and most obvious method of qualitatively code textual information is to read each document in a project and manually assign existing codes in a codebook to a selected segment of text. QDA Miner allows you to easily attach a code to a block of text that may be as small as a single character but that will more often consist of a few words, or one or more sentences or paragraphs. A code may even be assigned to an entire document.

The following section provides instructions on how to assign codes to text segments and image areas, resize the selected segment, change the associated code, attach a comment to a coded segments, as well as how to delete one or more codings.

## Assigning Codes to Text Segments

### To assign a single code to a segment of text:

You can use several methods to assign codes to text segments.

#### Method 1 - Quick assignments of codes to whole paragraphs

- Select the code from the code list (bottom left of the screen) by clicking it and holding the mouse button down.
- While holding the mouse button down, drag the mouse cursor over the paragraph you want to code.
- Release the mouse button to assign the dragged code to the paragraph found under the mouse cursor.

#### Method 2 - Highlight text and double-click the code

- Highlight the text segment that you want to code.
- Double-click the code in the code list.

#### Method 3 - Highlight text and drop the code

- Highlight the text segment you want to code.
- Drag an item from the code list to the selected segment.

#### Method 4 - Using the code list tool button.

- Select the proper code from the list box located in the toolbar above the document.

- Highlight the text that you want to code.
- Click the  button.

The last used code can be quickly assigned to a newly selected text segment by right-clicking in the document and selecting the CODE AS command from the shortcut menu.

### **To Assign Several Codes to the Same Segment:**

If you hold the **CTRL** key down while assigning a code to a text segment, the text segment will remain highlighted. You can then assign other codes to the same segment by double-clicking them in the code list or by dragging them from the code list to the highlighted segment.

## **Assigning Codes to Graphic Regions**

### **To assign a single code to a graphic area:**

You can use several methods to assign codes to a graphic area.

#### **Method 1 - Select the area and double-click the code**

- Select the area of the graphic you want to code.
- Double-click the code in the code list.

#### **Method 2 - Select the area and drop the code**

- Select the graphic area that you want to code.
- Drag an item from the code list into the selected area.

#### **Method 3 - Using the toolbar.**

- Select the proper code from the list box located in the toolbar above the image.
- Select the area that you want to code.
- Click the  toolbar button.

### **To assign several codes to the same area:**

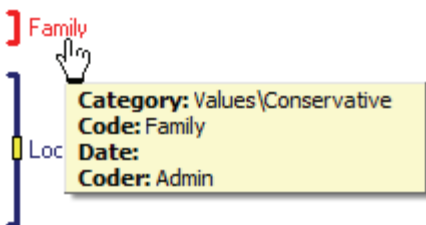
If you hold the **Ctrl** key down while assigning a code to a graphic area, the selection will remain active. You can then assign other codes to the same area by double-clicking them in the code list or by dragging them from the code list to the selected region.

## Working with Code Marks

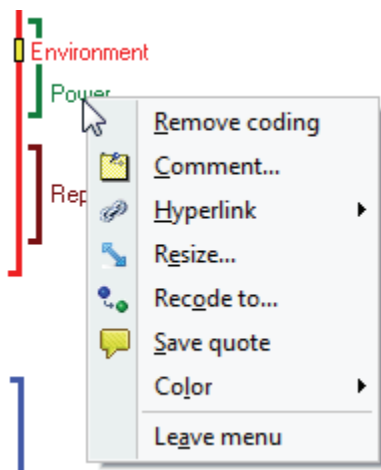
When a text segment has been successfully coded, a code mark will appear next to it in the margin area on the right side of the document. Code marks look like the following:



Color brackets are used to indicate the physical limits of coded segments. Moving the mouse cursor over a code mark (either the bracket or the code label) will display more detailed information about this specific coded segment, such as the category to which it belongs, its creation date, the name of the coder who created it, and any comment assigned to this coded segment.



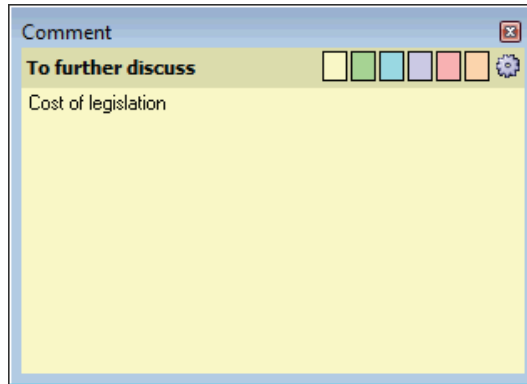
Clicking a code mark selects the text segment to which this code has been assigned. Clicking the same code mark a second time opens a shortcut menu. This menu can also be accessed by moving over a code mark and right-clicking it.



## Attaching a Comment to a Coded Segment

### To attach a comment to a coded segment:

- Select the coded segment to which you want to attach a comment by clicking its code mark.
- Click a second time to display the shortcut menu.
- Select the COMMENT command. A small window like the one below will appear.



- Enter the text that you want to associate with the selected coded segment.
- Click the **X** button in the upper right corner to save the comment and close the window.

You can choose among six types of comments by clicking one of the color rectangles in the upper-right corner of the dialog box (each color corresponds to a different comment type). The labels associated with those comment types appear in the upper-left corner of this window and may be customized by clicking the gear button.

When a note has been assigned to a coded segment, a small yellow square is displayed in the middle of the code bracket. To edit an existing code, follow the same steps. To remove a comment, simply open the note editor and delete all text in the note editing window.

To retrieve a list of all comments in a project file see page 149.

## Removing Coded Segments

Two methods may be used to remove codes assigned to text segments.

### Method 1 - Using the coded segment panel

- Select the coded segment that you want to delete by clicking its code mark.
- Click this mark a second time or right-click to display a shortcut menu.
- Select the REMOVE CODING command.

## Method 2 - Using the code deletion dialog box

The **Code deletion** dialog box allows you to delete one or more codes in the currently selected document or image.

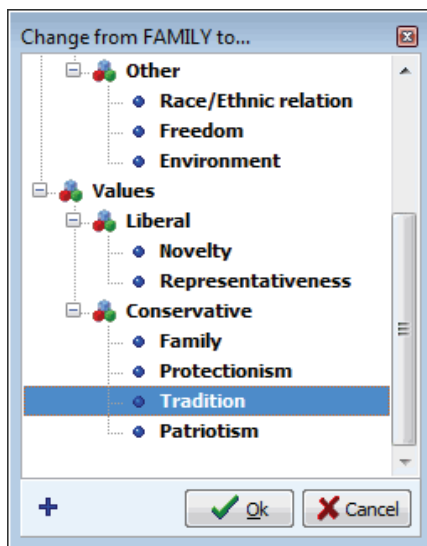
- Click the  button located on the toolbar of the document window. A list of all codes assigned to the current document or image will appear.
- Select the names of all the codes that you want to remove.
- Click the **OK** button.

Note: The above operations are only used to remove code assignments to text segments or image areas. They do not result in the removal of the code from the codebook. To remove a code along with all associated assignments, see **Deleting Existing Codes** (on page 77). To remove all code assignments in a project see **Clearing All Code Assignments** (page 70).

## Changing the Code Associated with a Segment

To change the code associated with a text segment:

- Select the coded segment that you want to remove by clicking its code mark.
- Click a second time to display the shortcut menu.
- Select the RECODE TO command. A window containing all available codes will be displayed.

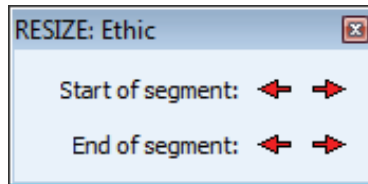


To change the code currently assigned to a new one, select the new code by clicking it, followed by the **OK** button. You can also double-click the desired code.

## Resizing Coded Segments

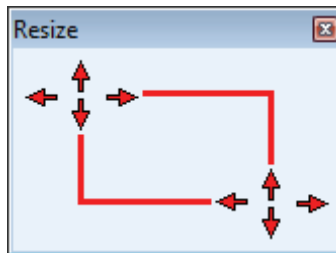
### To resize a coded segment:

- Select the coded segment that you want to resize by clicking its code mark.
- Click a second time to display the shortcut menu.
- Select the RESIZE command. For document variables, a small window like the one below will appear.



From this dialog box, you can move the starting or ending location of the selected segment by clicking the corresponding arrow button. Clicking a  button moves the chosen limit of the segment back to the left and up while clicking a  button moves this limit forward to the right and down.

For image variables, the resizing dialog box will look like this:



From this dialog box, you can move the upper-left and lower-right coordinates of the selected area by clicking the corresponding arrow button.

You can repeat the same move several times by clicking and holding the mouse button down. Release the mouse button when you are done.

## Hiding Code Marks

Sometimes, one would rather focus on, or examine, specific coding or assigned codes. In such a situation, it may be preferable to hide unrelated code marks in the document margin.

### To hide code marks associated with a specific code:

- In the codebook list, select the code for which the associated code marks will be hidden.
- Select the HIDE CODING command from the CODES menu.

Codes for which associated code marks have been hidden are displayed in the codebook using a lighter font and a grey bullet.

### To hide code marks associated with a category of codes:

- In the codebook list, select the category containing all the codes for which the associated code marks will be hidden.
- Select the HIDE CODING command from the CODES menu.

### To reveal hidden code marks:

- In the codebook list, select the code or the category for which the previously hidden code marks will be displayed.
- Select the HIDE CODING command from the CODES menu which will remove the check mark and will reveal all associated code marks.

Please note that hiding code marks only affects the display of those marks in the document margin. It has no effect whatsoever on the analysis or retrieval of coded segments.

## Highlighting or Hiding Coded Segments

Sometimes one may find it useful to hide or dim coded text in order to clearly focus on the portions of text that have not yet been coded. Alternatively, one may also wish to highlight text segments associated with specific codes. The CODED TEXT item from the DOCUMENT menu allows you to either hide, mask, dim, or highlight coded text segments. The following display properties are available:

- NORMAL (default, black text on white background)
- HIGHLIGHT (yellow background)
- DIM TEXT (gray text on white background)
- HIDE TEXT (white text on white background)
- MASK TEXT (black text on black background)
- CODE COLORS (the code color is used as the background color)

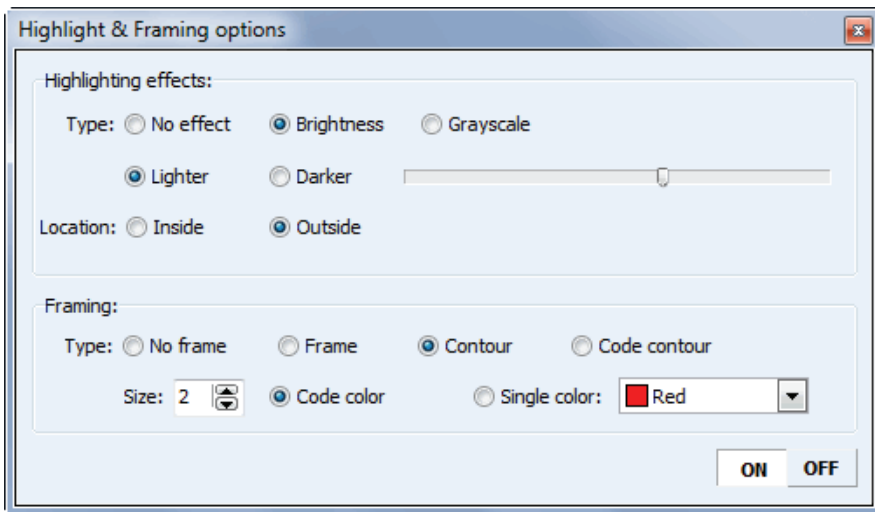
Please note that when choosing CODE COLORS as the highlighting mode and when more than one code has been assigned to the same text segment, only one background color is used.

The highlighting or hiding of coded segments is applied to all coded segments in which code marks are displayed. One may thus change display properties of text segments associated with specific codes by using the HIDE CODE MARKS command to hide code marks for all other codes (see section Hiding Code Marks above).

## Highlighting or Hiding Areas in Coded Images

Just like the highlighting feature for coded text, it is also possible to dim, mark or highlight specific regions of a coded image associated with specific codes.

To customize the highlighting function, click the Down arrow on the  button and select SETTINGS. The following dialog box will appear:



The first group of options at the top allows one to select what highlighting effects to apply, while the second group of options below it let you adjust how coded areas will be framed. The following options are available:

**TYPE** - This set of options allows one to choose whether the selected regions will be made Lighter or Darker as well as the intensity of this change using the sliding cursor. It also allows you to choose whether only the Brightness will be affected or whether the highlighted region should also be transformed into Grayscale. Selecting No effect will disable those settings. However, such a setting will not affect how coded areas will be framed (see below).

**LOCATION** - This option lets you choose whether the highlighting effect should be applied inside the coded areas or outside it.

The **FRAMING** option box offers various options to control how frames around coded regions will be drawn.

**TYPE** - This option offers four different choices:

|                     |                                                                                                                                                                                                              |
|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>No Frame</b>     | Prevents frames from being displayed around coded regions.                                                                                                                                                   |
| <b>Frame</b>        | Frames will be drawn around each coded region.                                                                                                                                                               |
| <b>Contour</b>      | Frames will be drawn only at the outer edges of all coded areas. If two or more areas are overlapping, lines crossing other coded areas will not be drawn even if their associated codes differ.             |
| <b>Code Contour</b> | Frames will be drawn at the edges of overlapping areas as long as they are associated with the same code. If two or more of these areas are overlapping, lines crossing those coded areas will not be drawn. |

**SIZE** - This option allows one to change the frame lines width.

**COLOR** - The color option allows one to choose the color for frame lines. If Code Color is selected, the color of frames around areas will be set to the color associated with the code. One may also choose a Single Color. A list box allows one to select any available color.

## Modifying Code Mark Color Scheme

The color scheme used to display code marks is set by default to highlight different codes. It may, however, be changed to emphasize the different coders who have assigned codes to the text segments.

### To change the coding scheme to highlight different coders:

- Select the CODE MARKS COLOR command from the CODES menu.
- Choose the BY CODER menu item.

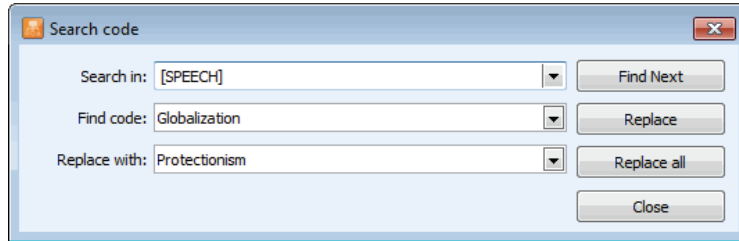
### To change the coding scheme to highlight codes:

- Select the CODE MARKS COLOR command from the CODES menu.
- Choose the BY CODE menu item.

To modify colors associated with different codes, see **Modifying Codes** on page 77. For instructions on how to assign different colors to different coders, see **Security and Multi-Users Settings** on page 27.

## Searching and Replacing Codes

The SEARCH & REPLACE command from the CODES menu may be used to interactively review codes assigned to text segments and, if needed, replace them with another code. When this command is executed a dialog box similar to the one below will appear:



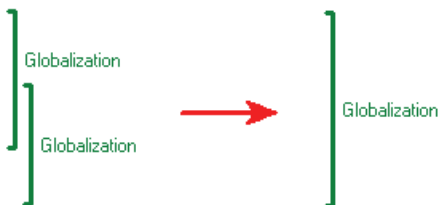
The **Search In** option allows you to specify which document or image variables to search. If the current project contains more than one document or image variable, you will have a choice of selecting either one or a combination of them. By default, all document variables are selected. To restrict the analysis to only a few of them, click the down arrow key at the right of the list box. You will be presented with a list of all codable variables. Select the variables on which you want the search to be performed.

The **Find Code** option allows you to choose which code you want to find. In the **Replace With** list box, select the replacement code.

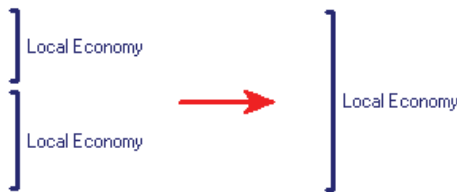
To replace all instances of the code at once, click **Replace All**. Or, to replace one instance at a time, click **Find Next**, and then click **Replace**.

## Consolidating Coded Segments

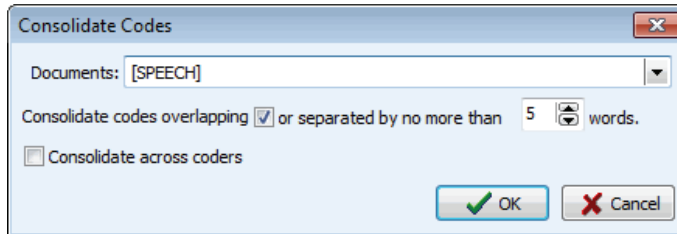
When performing extensive manual coding, autocoding and merging of codes, one may end up with some identical codes assigned more than once to the exact same text segment or to partially overlapping text segments. The CONSOLIDATE command allows one to replace codes assigned by the same coder to overlapping text segments by a single coded segment (see below).



This feature may also be used to join successive text segments sharing the same code and separated by less than a specified distance into a single, larger coded segment, like in the example below.



To consolidate overlapping or adjacent codes, run the **CONSOLIDATE** command from the **CODES** menu. The following dialog box will appear:



The **DOCUMENTS** option allows you to specify on which document variables the consolidation will be performed. If the current project contains more than one document variable, you will have a choice of selecting either one or a combination of them. By default, the currently displayed document variable is selected. To add other document variables, click the down arrow key at the right of the list box. You will be presented with a list of all available document variables. Select the variables on which you want the consolidation to be performed.

You may choose to consolidate overlapping coded segments only or also merge coded segments that are near each other without overlapping. To consolidate such adjacent coded segments, select this option by putting a check mark in the check box and specify the maximum number of words that can appear between the two segments. Setting this value to zero will only merge codes that appear next to each other, without any text between them. Setting this option to 5 will merge codes immediately next to each other or those that are separated by no more than five words.

By default, the consolidation of codes is performed only for codings made by the same user. The project administrator has the additional capability of consolidating codes across coders. Such a feature is especially useful in a situation where some documents have been coded by more than one user to establish inter-rater reliability and when one needs to calculate code statistics without counting the same code twice.

To proceed with the code consolidation, click the **OK** button. To leave the dialog box without modification to existing coded segments, click the **Cancel** button.

Please note that code consolidation cannot be undone. If you are not entirely sure as to the overall effect of code consolidation, you should make a backup copy of your project before attempting this operation.

# Hyperlinking

Hyperlinking allows one to attach to a specific text segment or coded segment a link to another part of the project or to some external resources. For example, one may use such a feature to link an example illustrating a fact or an idea to other examples corroborating or contradicting this first example, or to recreate a sequence of events over time by connecting successive events. The QDA Miner hyperlink feature may also be used to associate a coded segment to a specific geographic location or a specific time, allowing one to locate an event both in space and time.

Six types of hyperlinks are available in QDA Miner, allowing one to associate a piece of information with:

- 1) a web page;
- 2) a file on your computer system or available on your network;
- 3) a geographic location (with optional time information);
- 4) a time tag or time range;
- 5) another case;
- 6) a text or a coded segment.

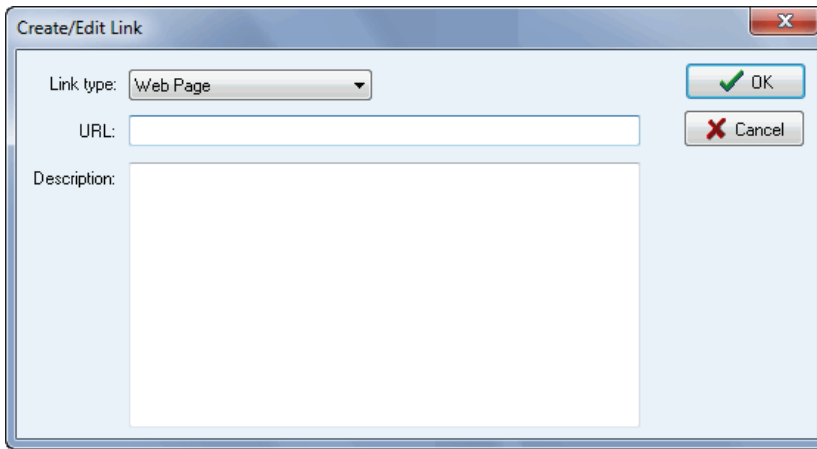
## **To add a hyperlink to a text segment (document variables only)**

- Select the text to which you want to attach a hyperlink.
- Right-click to display the shortcut menu and select the ADD command from the HYPERLINK menu item.

NOTE: Because of the way hyperlinks are stored in the document, they cannot include the start or the end of a coded segment. For this reason, when such a text segment is selected, the command to create a hyperlink is disabled. An alternative solution is to attach the hyperlink to a code rather than to the text itself.

## **To add a hyperlink to a coded segment (document and image variables)**

- Select the coded segment to which you want to attach a link by clicking its code mark.
- Click a second time to display the shortcut menu.
- Select the ADD command from the HYPERLINK menu item. A dialog box like this one will appear:



The **Link Type** list box allows you to specify which type of link to create. While the available options in this dialog box may differ depending on the selected link type, all link types share a common option that allows one to enter a Description. This edit box can retain the reasons why such a link was created, specify the nature of the relationship between the source and the target items, or add any comment. One may also type in some keywords that may later be searched for using the Link Retrieval feature.

The following section describes the specific options and data requirements associated with the different types of links. Geographic links will be discussed in a separate section.

### Web Page

**URL** - This option lets you specify the URL of the web page you would like to associate with the selected item. In order to be recognized as a web link, the `http://` or `https://` prefixes should be included. The most convenient way to write a web URL is by opening the desired web page, copying the address from the URL box that appears at the top of the browser, and pasting it into the URL edit box.

### External File

**File** - This option lets you type in a file name you would like to have associated with the selected item. This file may be a document, an image, a video or any other type of file, including executable applications. Clicking the  button allows you to browse through your computer to locate the desired file. Once the file is selected, its full path appears in the edit box. The OK button used to confirm the creation of the hyperlink will only be enabled if the typed file name points to an existing file.

### Geographic Location

See Geotagging Documents and Images (page 101)

## Time Stamp or Range

This type of link may be used to attach a specific time tag or time interval to the selected text or coded segment, allowing one to later perform coding retrieval based on dates or create timelines representing a sequence of events over time. The **Label** option may be used to store a single-line description of this event. The **Date** option is used to specify the basic date and time information. To specify a time range rather than a single time point, check the **To** box and then set the ending date and time..

An Icon may also be associated with a time stamp or time range. This icon may later be displayed in timelines created by QDA Miner (see Creating and Editing Timelines). You can change the icon of a specific time link by clicking the Icon button to the right of the dialog box and by choosing a new icon from a palette. QDA Miner allows one to choose among available Google Earth icons or offers the option to create your own icon (see Attaching an Icon to an Hyperlink).

NOTE: While clicking other types of links causes QDA Miner to move to another location or another application, clicking a time stamp or range link has no effect. Links with dates may, however, be searched based on date intervals (see Link Retrieval) and plotted on a timeline (see Creating and Using Timelines).

## Another Case

To create a link to another case, you first need to select it by clicking the [PICTURE LinkCase.bmp] button. You will be presented with a dialog box showing a list of all cases. Select the case you want to point to and click OK. The case description will automatically be stored in the Case edit box. This description can be edited to provide additional information.

## Text or Coded Segments

Creating a link to another text or coded segment is done in a two-step process:

### Step 1. Set the source segment

- Choose the text or coded segment to which the link should be attached and run the HYPERLINK | ADD command.
- Set the link type to Text or Coded Segment.
- Enter a description (optional) and then click OK. You will be returned to the main QDA Miner screen.

### Step 2. Set the destination segment

- Move to the segment where the source segment should point.
- Select it and run the HYPERLINK | SET AS DESTINATION command from the contextual menu.

You will be asked whether you want to create a link back to the source. Selecting Yes will attach a second hyperlink to the destination segment that will point to the original segment, allowing you to jump back and forth between the two locations. Selecting No will create a single link from the source segment to the destination segment.

### **To edit a hyperlink**

- For hyperlinks associated with a coded segment, select its code mark from the right margin of the document (for images, you can also select from the bottom margin). For hyperlinks associated with a text segment, position the editing cursor anywhere within the text.
- Right-click to display the shortcut menu and select the EDIT command from the HYPERLINK menu item. Close the dialog box when finished editing. Note that it is not possible to change the hyperlink type with the EDIT command. To perform such a change, the hyperlink must be deleted and recreated.

### **To delete a hyperlink**

- For hyperlinks associated with a coded segment, select its code mark in the right margin of the document. For hyperlinks associated with a text segment, position the editing cursor anywhere within the text.
- Right-click to display the shortcut menu and select the DELETE command from the HYPERLINK menu item.

### **To follow a hyperlink**

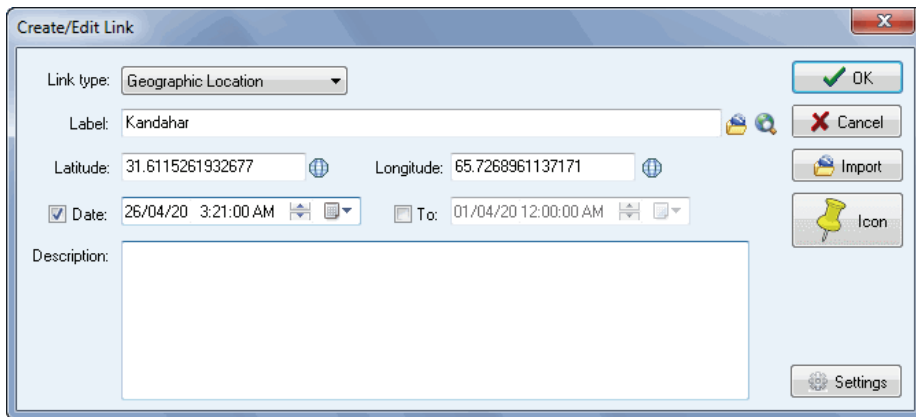
- Move the mouse cursor over the code mark or the coded segment containing the hyperlink.
- Hold the Control key down and click the link. One may also right-click to display the shortcut menu and select the FOLLOW LINK command from the HYPERLINK menu item.

# Geotagging of Documents and Images

Geographic tagging allows one to associate a specific text or coded segment to a geographic location and, optionally, to a point in time or a time period, allowing one to situate an event both in space and in time. One can enter the coordinates of the geographic location manually in decimal values or in degrees, minutes and seconds, import the coordinates from Google KMZ or KML files or ESRI Shapefiles, or simply copy and paste placemarks from Google Earth.

## To create a geographic tag

- Select the text or coded segment to which you want to attach such a tag.
- Right-click to display the shortcut menu and select the ADD command from the HYPERLINK menu item.
- Set the Link Type option to Geographic Location. A dialog box similar to this one will appear:



The screenshot shows a 'Create/Edit Link' dialog box with the following fields and controls:

- Link type:** A dropdown menu set to 'Geographic Location'.
- Label:** A text input field containing 'Kandahar'.
- Latitude:** A text input field containing '31.6115261932677' with a globe icon to its right.
- Longitude:** A text input field containing '65.7268961137171' with a globe icon to its right.
- Date:** A checked checkbox followed by a text input field containing '26/04/20 3:21:00 AM' and a calendar icon.
- To:** A text input field containing '01/04/20 12:00:00 AM' with a calendar icon.
- Description:** A large empty text area.
- Buttons:** 'OK' (green checkmark), 'Cancel' (red X), 'Import' (folder icon), 'Icon' (pushpin icon), and 'Settings' (gear icon).

The **Label** edit box allows one to add a short description for this geographic tag. Such a label will be used as the placemark name in Google Earth and in other locations to identify these specific coordinates so that you may later retrieve them by name. We recommend entering a concise but unique label. The **Description** edit box could be used to enter a longer description. It may also be used to enter some keywords or key phrases that may later be searched for.

A geographic coordinate can be defined by specifying a **Latitude** and a **Longitude**. There are four ways to specify a geographic coordinate:

### 1. By entering data manually

You can specify the latitude and longitude manually by either entering the floating-point values associated with the geographic coordinates or by clicking the  button and entering the geographic coordinates in degrees, minutes, seconds and cardinal compass point.

## 2. By importing coordinates from a Google KMZ or KML file

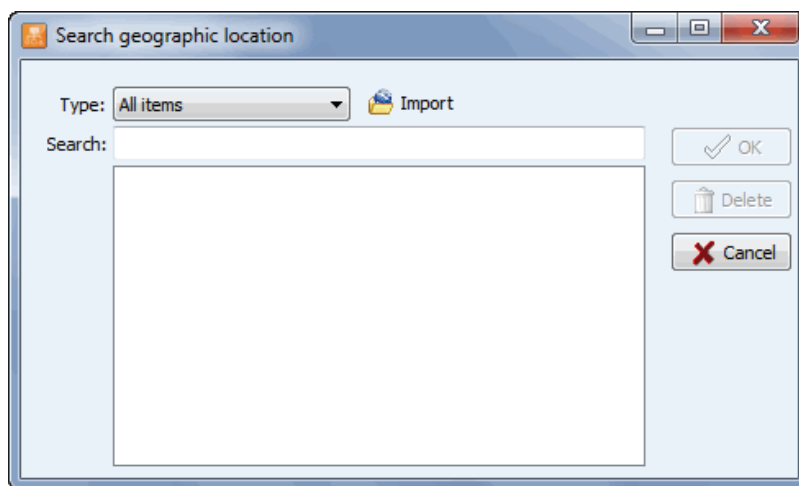
Click the  button. An **Open File** dialog box will appear, allowing you to select either a Google KMZ or KML file. If several geographic locations are stored in the selected file, you will be asked to choose which geographic location to import. The placemark label and description (if any) will also be imported.

## 3. By cutting and pasting Google Earth placemarks

One of the easiest methods to identify the geographic coordinates of a specific location and to import these coordinates into QDA Miner is to use Google Earth to find the proper location and create a placemark in Google Earth itself. Once the placemark has been created, select it and press **Ctrl-C** (to copy it to the clipboard) then move back to QDA Miner and paste it (using the **Ctrl-V** key combination).

## 4. By searching previously entered or imported locations

All geographic coordinates entered by the user in QDA Miner are automatically stored in QDA Miner along with their associated labels allowing one to search previously typed or imported coordinates. To access the search dialog box, click the  button located to the right of the **Label** edit box. A dialog box similar to this one will appear:



This dialog box allows one not only to search previously entered geographic locations but also to import geographic coordinates from ESRI Shapefiles or Google Earth KMZ or KML files (see Importing Geographic Coordinates). To select a geographic location, simply locate its label in the main list box, select it, and then click the **OK** button. Listed locations may be limited to those obtained from specific sources by setting the **Type** list box to a desired value. Setting this option to All will display coordinates from all sources. One may also filter coordinates by entering a

specific keyword or key phrase in the **Search** edit box. The list of locations is updated while typing so as to display only those entries with labels containing the typed string (Note that the search is not case sensitive).

A date stamp or range may also be associated with the geographic location by selecting the Date check box and by specifying the date and time information. To specify a time range, set the To check box and then type the ending date and time.

Google Earth displays placemarks using a set of icons that may be customized by the user. These icons may also be displayed in timelines created by QDA Miner (see **Creating and Editing Timelines**, page 154). You can change the icon of a specific geographic location by clicking the Icon button to the right of the dialog box and by choosing a new icon from a palette. QDA Miner allows one to choose among available Google Earth icons or offers the possibility of creating your own icon (see **Attaching an Icon to an Hyperlink**, page 105)

Once a geographic tag has been added to a text or coded segment, clicking this link while holding the CTRL key creates a temporary KMZ file and opens it using Google Earth or any other application associated with this file type (for example, ArcGIS Explorer).

For more information on how to edit, delete or search hyperlinks, see pages 97-100).

## Importing Geographic Coordinates

Geographic coordinates may be entered one at a time either by typing those coordinates, by cutting and pasting Google Earth placemarks, or by importing a single geographic coordinate stored in a KMZ or KML file. It is also possible to import multiple coordinates from Google KMZ or KML files or ESRI Shapefiles so that they become readily accessible for searching in QDA Miner. This importing feature of QDA Miner is available through the **Search Geographic Locations** dialog box. This dialog box can be accessed by clicking the  button found in the **Geographic Hyperlink Editor** or in the **Link Retrieval** dialog box.

### To import new geographic coordinates:

- Click the  button. An **Open File** dialog box will appear, allowing you to view all files with a .kml, kmz or .shp file extension.
- Select the file containing the coordinates to be imported and click **Open**. QDA Miner will read the selected file and display extracted coordinates and other information about each location in a dialog box similar to this one:

citiesx020.shp - 28706 coordinates

Save under:

Label: [Click here to create a custom label](#)

| NAME         | STATE | COUNTY           | LATITUDE         | LONGITUDE         |
|--------------|-------|------------------|------------------|-------------------|
| Autaugaville | AL    | Henry County     | 31.5721492767334 | -85.2505264282227 |
| Booth        | GA    | Wilcox County    | 31.9926471710205 | -83.3064193725586 |
| Jones        | LA    | Vermilion Parish | 29.9751052856445 | -92.1347503662109 |
| Prattville   | MS    | Lafayette County | 34.5032501220703 | -89.5034255981445 |
| Bay Minette  | SC    | Abbeville County | 34.1785202026367 | -82.3788909912109 |
| Daphne       | WI    | Clark County     | 44.9461212158203 | -90.3164901733398 |
| Fairhope     | AR    | Scott County     | 35.0737190246582 | -94.1944198608398 |
| Foley        | NM    | Colfax County    | 36.3048858642578 | -104.26139831543  |
| Fort Morgan  | TX    | Hill County      | 31.8852500915527 | -97.0741577148438 |
| Gasque       | VA    | Craig County     | 37.4294128417969 | -80.1391220092773 |
| Gulf Shores  | WV    | Upshur County    | 38.9220352172852 | -80.3229293823242 |
| Lillian      | NC    | Bladen County    | 34.5174560546875 | -78.7242736816406 |
| Loxley       | TX    | Carson County    | 35.3934478759766 | -101.459587097168 |
| Davidson     | ND    | Richland County  | 46.4477081298828 | -106.73005703125  |

LABEL: <none>

- In the **Save Under** edit box, enter the name of the list in which those locations will be saved. If an existing list name is typed, the new locations and their coordinates will be added to that existing list.
- A Label for each location also needs to be specified. This label consists of data from a single column or a combination of several columns. When importing KMZ or KML files, this label is automatically set to the NAME of the placemark. For ESRI Shapefiles containing several text data, one needs to select columns that will be used to construct the label. To build such a label, click the **Click here to create a custom label** link. A dialog box similar to this one will appear:

Location Label

Variables:

FEATURE  
NAME  
STATE  
COUNTY

Label String: {NAME}, {STATE}

OK Close

The label may be changed by editing text in the **Label String** edit box. To insert the value stored in a specific field into the description, simply enter the column name (in uppercase letters) enclosed between braces. Alternatively, you can insert a column name at the current location by clicking the corresponding item in the **Column Names** list located just above the edit box.

If you enter the following string:

{NAME}, {STATE}

The{NAME}and{STATE}strings will be replaced with their corresponding column value for each location.

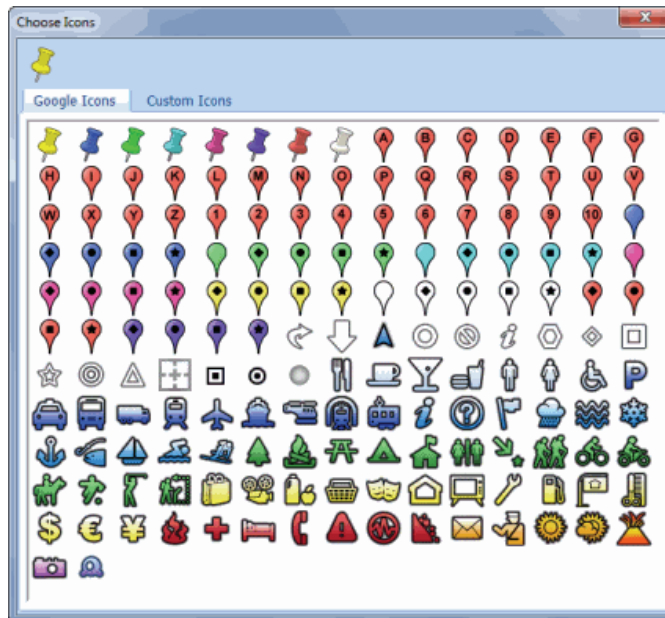
- To prevent importing of specific locations, simply select those locations and press the Del key.
- Once all options have been set, click the **OK** button to proceed with importing.

## Attaching an Icon to a geo-tag or time tag

QDA Miner allows one to customize the icon that will be used to display placemarks in Google Earth as well as icons in timelines.

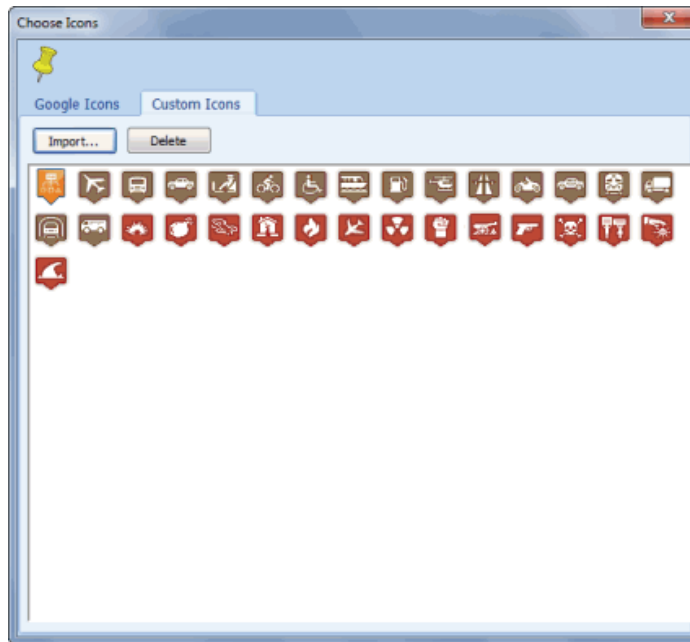
### To change the icon associated with a geographic or time tag:

- Select the hyperlink which icon has to be changed, right-click and run the HYPERLINK | EDIT command.
- Click the  button. And Icon palette dialog box similar to this one will appear:



- To select an icon from the Google icon collection, simply click the desired icon and close the dialog box.

- To select a custom icon, click the **Custom Icons** tab.



- Click the desired icon and close the dialog box.
- To create a custom icon from an existing image file, click the **Import** button. An **Open File** dialog box will appear, allowing you to select this image file. Once selected, the image will be transformed into an icon, inserted into your list of custom icons and selected.
- To delete a custom icon, click it and then click the **Delete** button.

# Retrieval Features

QDA Miner provides several search and retrieval tools to assist in the coding task. These tools may be used to speed up or systematize the coding of documents, ensure coding consistency, or retrieve relevant segments of documents, images or comments.

The **Text Retrieval** tool searches for specific text patterns in documents. Complex search patterns may be performed using Boolean operators. This search can be performed on all documents in a project or restricted to specific ones or to some coded segments.

The **Section Retrieval** tool searches structured documents for sections bounded by fixed delimiters. This feature is especially useful to automatically assign codes to recurring sections within a document or in multiple documents. You can search in all documents in a project or restrict the search to specific document variables.

The **Keyword Retrieval** feature can retrieve any document, paragraph, sentence, or coded segment containing a specific WordStat keyword or a combination of keywords.

The **Cluster Extraction** tool allows one to quickly identify all similar paragraphs or sentences in a large collection of documents and code them quickly.

The **Code Similarity** retrieval tool quickly identifies sentences or paragraphs that share some similarities with existing coded segments in the current project or those stored in another coded project.

The **Query by Example** retrieval tool may be used to search for text segments that are, in some ways, similar to a given example. It may also be used to find text similar to all segments associated with a specific code, and thus retrieve sentences or paragraphs that may potentially be coded the same way.

The **List Comments** feature allows one to retrieve comments attached to coded segments.

The **Link Retrieval** tool allows one to obtain a list of all hyperlinks, geographic tags or time tags in a project or to restrict this list to some link types or to those meeting specific search criteria. This dialog box also allows one to create timelines out of time tags as well as geographic maps for all links containing geographic coordinates.

## Saving and Retrieving Queries

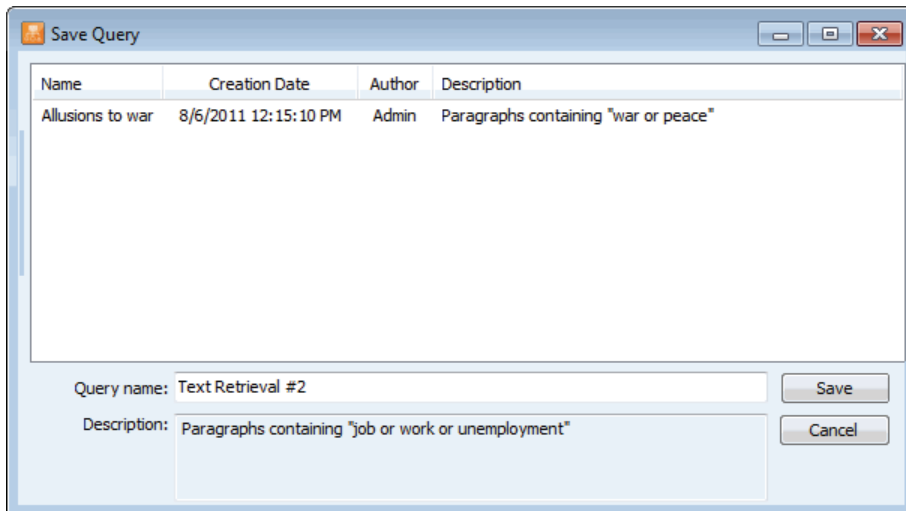
QDA Miner provides different types of queries that enable you to find specific text patterns, to select subsets of cases, and to explore the relationships among codes and between those codes and case properties. The program allows you to save with the project different kinds of queries and analysis options and retrieve them later. QDA Miner provides savings and retrievals for the following functions:

- Text Retrieval
- Section Retrieval
- Keyword Retrieval
- Coding Retrieval
- Coding Co-occurrences
- Coding Sequences
- Inter-coders Agreement
- Case Filtering

Dialog boxes associated with each of those query types have two buttons: The  button is used to save the current question options with the project, while the  button is used to retrieve previously saved queries.

### To Save a Query

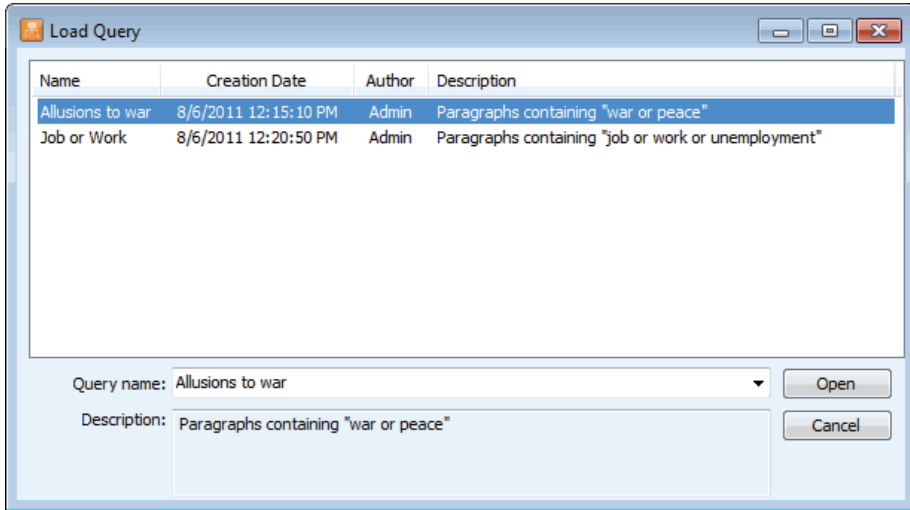
- Set the various options on the dialog box.
- Click the  button. A dialog box similar to the one below will appear:



- The program automatically provides a generic query name and a description. Edit this name and description and click the **Save** button to store this query with the project. If a query with an identical name already exists, you will be asked to confirm that you wish to overwrite it.

## To Retrieve a Previously Saved Query

- Open the dialog box associated with the task you want to perform.
- Click the  button. A dialog box similar to the one below will appear:



Please note that if no query of this type has been previously saved, this option will be disabled.

- From the list of previously saved queries, you can select the query you want to retrieve, either by double-clicking its line, or click it once and then clicking the Open button. You may also type its name in the Query Name edit box or select it from a drop-down list and then click the **Open** button to open it.

## To delete queries

- From the **Save Query** or **Load Query** dialog box, select the query you would like to delete.
- Click the **Del** keyboard key or right-click to display a pop-up menu and choose the **Delete** command.

## To rename a query

- From the **Save Query** or **Load Query** dialog box, select the query you would like to rename.
- Right-click to display a pop-up menu and choose the **Rename** command.
- Enter the new name and click **OK**.

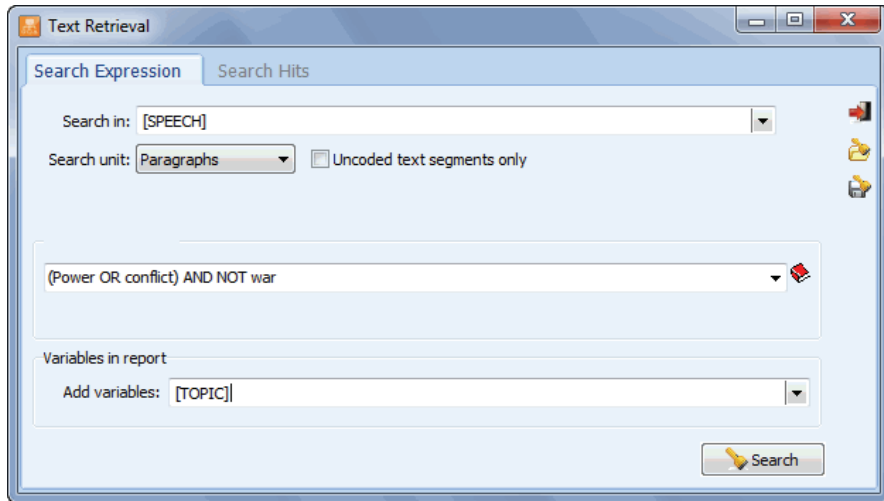
## To change the description of a query

- From the **Save Query** or **Load Query** dialog box, select the query with the description you would like to modify.
- Right-click to display a pop-up menu and choose the **Edit Description** command.
- Change the description text and click **OK**.

# Text Retrieval

The TEXT RETRIEVAL function searches for specific text or combination of text in documents. You can search in all documents in a project or restrict the search to specific document variables. Searches can also be restricted to specific coded segments.

To start the text search feature, select the TEXT RETRIEVAL command from the ANALYSIS menu. The following dialog box will appear:



On the first page you can set the search conditions while results are displayed on the **Search Hits** page. When entering this dialog box, the second page is normally disabled. As soon as a search returns at least one hit, QDA Miner will enable this second page, allowing you to review the search hits and optionally assign codes to them.

The **SEARCH IN** option allows you to specify on which document variables the search will be performed. If the current project contains more than one document variable, you will have a choice of selecting either one or a combination of them. By default, all document variables are selected. To restrict the analysis to only a few of them, click the down arrow key at the right of the list box. You will be presented with a list of all available document variables. Select the variables you wish to search.

The **SEARCH UNIT** option determines the search unit on which the search will be performed as well as what will be retrieved. You can select three different search units:

If you select **Documents** as the search unit, QDA Miner will apply the search expression on each document associated with a specific case. If a specific document meets the search condition, its location will be returned.

When selecting **Paragraphs** as the search unit, QDA Miner will return any paragraph meeting the search condition.

Select **Sentences** to instruct QDA Miner to return sentences meeting the search condition.

Selecting **Coded Segments** as the search unit restricts the search to the text segments that have already been coded. When a coded segment meets the search condition, its entire text is returned, no matter whether it is a single word or several paragraphs. To restrict the search to specific codes, select the **Selected** radio button on the right side of this option and select the codes from the drop-down checklist by clicking the arrow button at the right end of this list. These codes can also be selected from a tree representation of the codebook by clicking the  button. To perform the search on all codes, select the **All** radio button.

When searching sentences or paragraphs, one may restrict the retrieved text segments to those that have not been previously coded by enabling the **Uncoded Text Segment Only** check box.

**SEARCH EXPRESSION** - This edit box is used to specify words or phrases being searched for. If disabled, the search will retrieve all text segments, no matter what their content. To search for a single word, simply enter the word in the edit box. In order to search for a phrase, you need to put the phrase between quotation marks or replace the space between words with underscore characters. For example, entering either:

"quality of life"  
or  
quality\_of\_life

with retrieve text segments containing this phrase.

Boolean operators **AND**, **OR** and **NOT** may also be used to build complex search expressions. For example, the following search string:

angry OR mad

will retrieve any search unit containing either the word "angry" or the word "mad". Parentheses may also be used to specify the evaluation order in complex search expressions. For example, if you enter the following search expression:

(angry OR mad) AND (son OR daughter OR child)

and set the Search Unit option to Paragraphs, QDA Miner will return all paragraphs containing either "angry" or "mad", but only if this paragraph also contains either one of the three words found in the second set of parentheses (e.g. "son", "daughter" or "child").

If several words are entered without a Boolean operator, then QDA Miner will assume that all those words are required, and thus behave as if you had entered an "AND" operator between those words.

## Using wildcards in search expressions

A wildcard is a character or a set of characters that may be used in a search expression to represent one or more other characters. QDA Miner support the following wildcard:

- |                   |                                                                                                                                                                                    |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>?</b>          | Matches any single character. For example <b>F?T</b> will find <b>FAT, FIT</b> and any other combination of three characters beginning with 'F' and ending with 'T'.               |
| <b>*</b>          | Matches any number of contiguous characters. For example, <b>T*ENT</b> will return every word starting with 'T' and ending with 'ENT' such as <b>TENT, TORRENT, TANGENT</b> , etc. |
| <b>#</b>          | Matches any numeric character.                                                                                                                                                     |
| <b>[ ]</b>        | Matches any of the specified characters at that position. For example, <b>[abc]</b> will match if a or b or c at that position.                                                    |
| <b>[^] or [!]</b> | Matches anything but the specified characters at that position. For example, <b>[^abc]</b> will match any single character, but a, b, or c.                                        |
| <b>[a-e]</b>      | Matches a through e at that position.                                                                                                                                              |
| <b>_</b>          | The underscore character may be used to replace space within words. For example searching for <b>QUALITY_OF_LIFE</b> will match any instance of this phrase.                       |

## Performing thesaurus-based searches

A thesaurus-based search allows one to search for several words or phrases associated with a single thesaurus entry previously defined. For example, by entering a single category **@GOOD** (with an ampersand character "@" as a prefix), the program can automatically search for items that have been associated with this category, like "good", "fine", "excellent", "all right", "topnotch", etc. Names of categories may be typed directly in the search expression by preceeding the category with a # character. For example, the following text search expression:

**@GOOD and @SERVICE**

will retrieve all text units containing either one of the words or phrases associated with the thesaurus entry **GOOD** along with any word or phrase included in the **SERVICE** thesaurus entry.

One may also insert a category, by clicking the  button to display the thesaurus editing dialog box, selecting the thesaurus entry and then clicking the Insert button. The Thesaurus Editing dialog box allows one to create new entries or edit existing ones. For more information on thesaurus searches, see **Thesaurus Editing** on page 117.

**WHOLE WORDS** - This option defines whether the words in the search expressions may be part of a word or if only whole words are to be found. For example, if this option is disabled, searching for a word like "bank" will return any search unit containing not only "bank", but also "banking", "bankruptcy", etc.

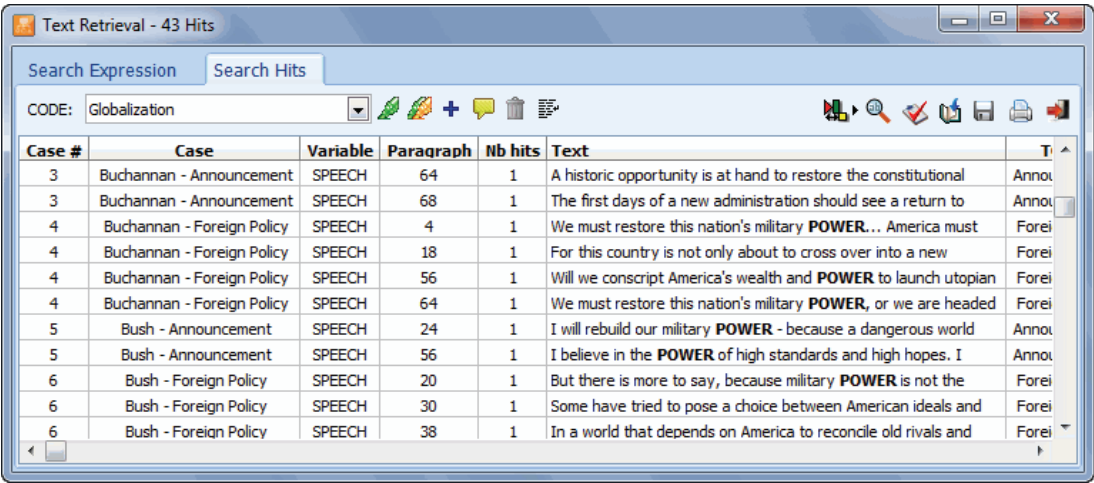
**CASE SENSITIVE** - This option defines whether the search is case-sensitive or not. If enabled, only words with the exact same cases will be returned.

**ADD VARIABLES** - This drop-down checklist box may optionally be used to add the values stored in one or more variables to the table of retrieved segments for the specific case from which a text segment originate.

To store the search expression and options, click the  button and specify the name under which those query options will be saved.

To retrieve a previously saved query, click the  button and select from the displayed list the name of the query you would like to retrieve (for more information, see **Saving and Retrieving Queries** on page 108).

To perform the search, click the  button. The results of a search are displayed in a table located on the **Search Hit** page.



The screenshot shows a window titled "Text Retrieval - 43 Hits" with two tabs: "Search Expression" and "Search Hits". The "Search Hits" tab is active, displaying a table of search results. The table has columns: Case #, Case, Variable, Paragraph, Nb hits, Text, and a partially visible column on the right. The search code is "Globalization". The table contains 12 rows of results, with the word "POWER" highlighted in several text segments.

| Case # | Case                      | Variable | Paragraph | Nb hits | Text                                                                   |       |
|--------|---------------------------|----------|-----------|---------|------------------------------------------------------------------------|-------|
| 3      | Buchanan - Announcement   | SPEECH   | 64        | 1       | A historic opportunity is at hand to restore the constitutional        | Anno  |
| 3      | Buchanan - Announcement   | SPEECH   | 68        | 1       | The first days of a new administration should see a return to          | Anno  |
| 4      | Buchanan - Foreign Policy | SPEECH   | 4         | 1       | We must restore this nation's military <b>POWER</b> ... America must   | Forei |
| 4      | Buchanan - Foreign Policy | SPEECH   | 18        | 1       | For this country is not only about to cross over into a new            | Forei |
| 4      | Buchanan - Foreign Policy | SPEECH   | 56        | 1       | Will we conscript America's wealth and <b>POWER</b> to launch utopian  | Forei |
| 4      | Buchanan - Foreign Policy | SPEECH   | 64        | 1       | We must restore this nation's military <b>POWER</b> , or we are headed | Forei |
| 5      | Bush - Announcement       | SPEECH   | 24        | 1       | I will rebuild our military <b>POWER</b> - because a dangerous world   | Anno  |
| 5      | Bush - Announcement       | SPEECH   | 56        | 1       | I believe in the <b>POWER</b> of high standards and high hopes. I      | Anno  |
| 6      | Bush - Foreign Policy     | SPEECH   | 20        | 1       | But there is more to say, because military <b>POWER</b> is not the     | Forei |
| 6      | Bush - Foreign Policy     | SPEECH   | 30        | 1       | Some have tried to pose a choice between American ideals and           | Forei |
| 6      | Bush - Foreign Policy     | SPEECH   | 38        | 1       | In a world that depends on America to reconcile old rivals and         | Forei |

The table provides basic information on each hit, such as the case number and label, as well as the document variable in which it was found, the hit location, etc. To sort this list of hits in ascending order on any column values, simply click this column header. Clicking a second time on the same column header sorts the rows again in descending order.

Selecting an item in this table, either by clicking it or by using the keyboard cursor keys, such as the **UP** or **DOWN** arrow keys, automatically displays the corresponding case and document in the main window. If the search unit was set to **Paragraphs** or **Coded Segments**, selecting a specific hit will also highlight the corresponding text segment. This feature provides a way to examine the retrieved segment in its surrounding context. You can also assign an existing code to the highlighted segment.

### To assign a code to a specific search hit:

- In the table of search hits, select the row corresponding to the text segment you want to code.
- Use the **CODE** drop-down list located above this table to select the code you want to assign.
- Click the  button to assign the selected code to the highlighted text segment.

### To assign a code to all search hits

- Use the **CODE** drop-down list located above this table to select the code that you want to assign
- Click the  button to assign the selected code to all text segments meeting the search expression.

Prior to the assignment of a code to all search hits, you may want to remove selected hits that do not correspond to what you were looking for. To remove a search hit from the list, select its row and then click the  button.

### To display the automatic text classification panel:

- Click the  button. A panel will appear at the bottom of this hit list, allowing you to apply an automatic text classification model on the retrieved text segments. Clicking this button again, will hide the panel.

For more information on this feature, see **Automatic Text Classification** on page 213.

### To analyze the text of all retrieved segments using WordStat:

- Click the  button. QDA Miner will minimize itself and run WordStat, allowing you to perform a quantitative content analysis or apply text-mining techniques on the retrieved text segments. WordStat will also allow you to perform keyword comparisons across any other column in the table. To include additional data for comparison analysis, use the **Add Variables** option.

### To display the coding table:

- Click the  button. A dialog box will appear allowing you to code retrieved text segments using drag and drop operations, remove existing codings and attach comments to codes.

For more information on this feature, see **Coding Text Segments Using the Coding Table** on page 222.

### To export the table to disk:

- Click the  button. A Save File dialog box will appear.
- In the **Save As Type** list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file

(\*CSV), MS Word (\*.DOC), HTML file (\*.HTM; \*.HTML), XML files (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (.SAV).

- Type a valid file name with the proper file extension.
- Click the **Save** button.

### **To append a copy of the table in the Report Manager:**

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.
- To append only a single hit from the table, select the row containing the text segment you want to append to the Report Manager and click the  button.

For more information on the Report Manager, see the **Report Manager Feature** topic on page 230.

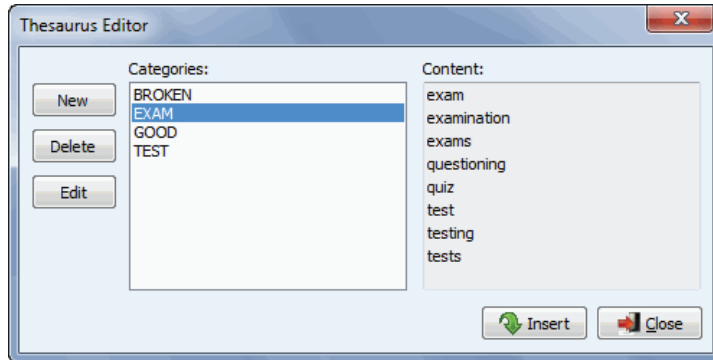
### **To print the table:**

- Click the  button.

## Thesaurus Editing

A thesaurus-based text search allows one to search for several words or phrases associated with a single thesaurus entry previously defined. The Thesaurus Editor allows one to create entries that can later be used in text search expressions to retrieve text units containing anyone of those words or phrases.

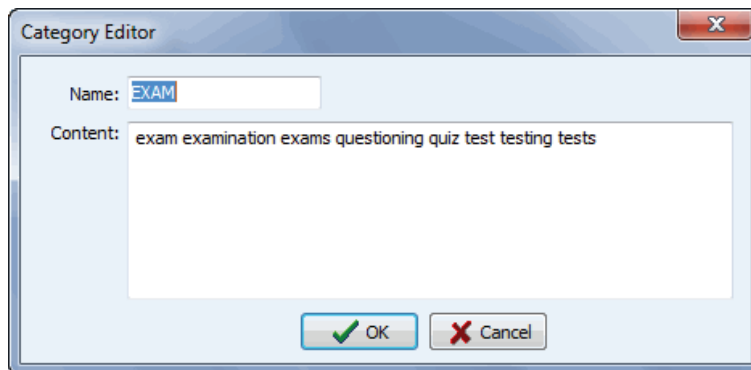
The Thesaurus Editor can be accessed from the Text Search dialog box by clicking the  icon located next to the text search expression. When activate, a dialog box similar to this one appears.



The dialog box displays a list of items currently defined. To view words and phrases associated with an entry, simply select the item in the **CATEGORIES** list. Items associated with the highlighted category are listed on the right side of the dialog box in the **CONTENT** list box.

### To create an new category:

- Click the **NEW** button. A dialog box similar to this one will appear.



- Enter the name of the category. This will be the placeholder to be used in search expression.
- Type the list of words you would like to search for when the category name is entered in a search expression. You may separate words by spaces or line breaks. Spaces in phrases should be replaced by an underscore character. For example, to search for the phrase "quality of life", you have to type "quality\_of\_life". You may also use any one of the valid wildcards supported by QDA Miner.
- Click the **OK** button to store the newly defined category.

**To edit an existing category:**

- Select the category you would like to edit.
- Click the **EDIT** button.
- Edit the list of items associated with the category.
- Click the **OK** button to store the newly defined category.

**To delete an existing category:**

- Select the category you would like to delete.
- Click the **DELETE** button.

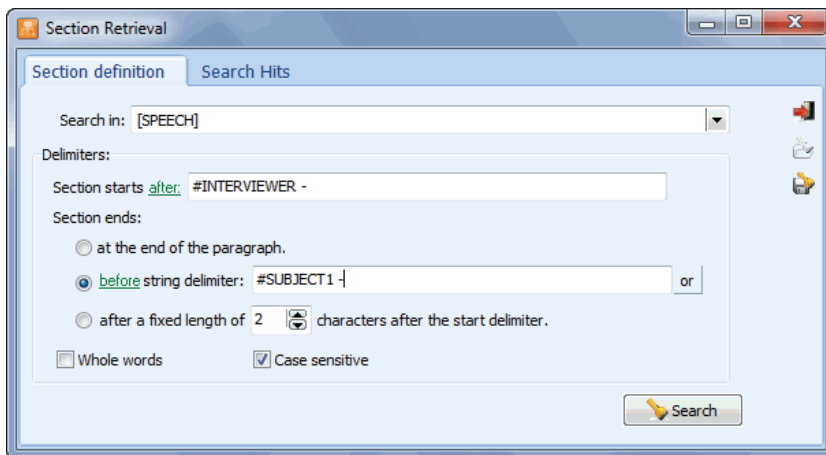
**To insert a category into a search expression:**

- Select the category you would like to insert.
- Click the **INSERT** button.

# Section Retrieval

The SECTION RETRIEVAL function searches structured documents for sections bounded by fixed delimiters. This feature is especially useful to automatically assign codes to recurring sections within a document or in multiple documents. You can search in all documents in a project or restrict the search to specific document variables.

To start the section search feature, select the SECTION RETRIEVAL command from the ANALYSIS menu. The following dialog box will appear:



On the first page, you can set the search conditions and results are displayed on the **Search Hits** page. When entering this dialog box, the second page is initially disabled. As soon as a search returns at least one hit, QDA Miner will enable this second page, allowing you to review the search hits and optionally assign codes to them.

The **SEARCH IN** option allows you to specify on which document variables the search will be performed. If the current project contains more than one document variable, you will have a choice of selecting either one or a combination of them. By default, all document variables are selected. To restrict the analysis to only a few of them, click the down arrow key at the right of the list box. You will be presented with a list of all available document variables. Select the variables on which you want the search to be performed.

The **DELIMITERS** option set allows you to specify which delimiters are used to define the starting and ending location of a section.

- The **Section starts** option specifies which constant string is used to delimit the start of the section. By default, the retrieved section begins after this constant string. To include this delimiter in the retrieved segment, click the **after** label to display a context menu and select **with**.
- The **Section ends** option allows you to select the method used to specify the length of the section. Three methods are available.

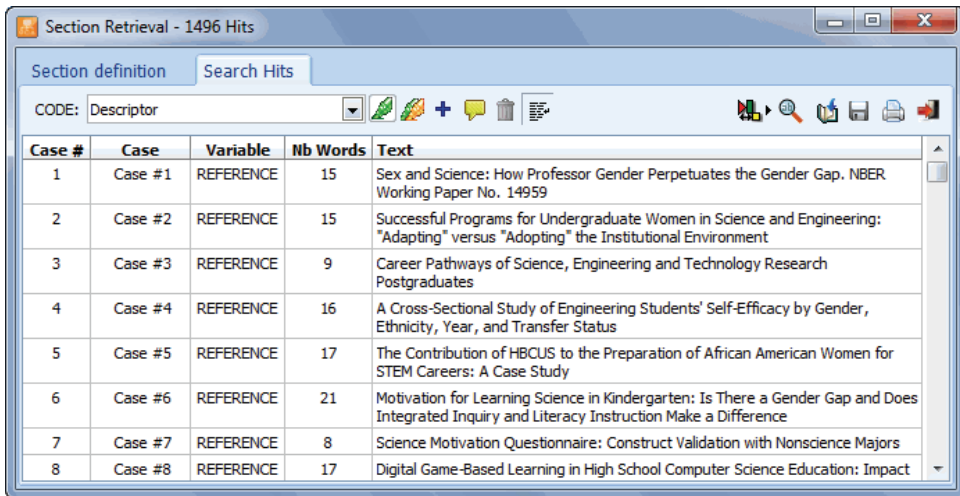
- 1) If you select **at the end of the paragraph** the program retrieves all text found from the start delimiter up to end of the paragraph.
- 2) Selecting **string delimiter** allows you to specify one or several constant strings found at the end of a section. To enter several string delimiters, type the | character or click the  button located at the right of the edit box. By default, the retrieved section does not include this delimiter and ends before it. To include this delimiter, simply click the **before** label to display a context menu and select **with**.
- 3) The third method allows you to retrieve a fixed number of characters after the end of the starting delimiter.

**CASE SENSITIVE** - This option defines whether the search is case-sensitive or not. If enabled, only string delimiters with the exact same cases will be recognized.

To store the search expression and options, click the  button and specify the name under which those query options will be saved.

To retrieve a previously saved query, click the  button and select from the displayed list the name of the query you would like to retrieve (for more information, see **Saving and Retrieving Queries** on page 108).

To perform the search, click the  button. The results of a search are displayed in a table located on the **Search Hit** page.



| Case # | Case    | Variable  | Nb Words | Text                                                                                                                                          |
|--------|---------|-----------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| 1      | Case #1 | REFERENCE | 15       | Sex and Science: How Professor Gender Perpetuates the Gender Gap. NBER Working Paper No. 14959                                                |
| 2      | Case #2 | REFERENCE | 15       | Successful Programs for Undergraduate Women in Science and Engineering: "Adapting" versus "Adopting" the Institutional Environment            |
| 3      | Case #3 | REFERENCE | 9        | Career Pathways of Science, Engineering and Technology Research Postgraduates                                                                 |
| 4      | Case #4 | REFERENCE | 16       | A Cross-Sectional Study of Engineering Students' Self-Efficacy by Gender, Ethnicity, Year, and Transfer Status                                |
| 5      | Case #5 | REFERENCE | 17       | The Contribution of HBCUS to the Preparation of African American Women for STEM Careers: A Case Study                                         |
| 6      | Case #6 | REFERENCE | 21       | Motivation for Learning Science in Kindergarten: Is There a Gender Gap and Does Integrated Inquiry and Literacy Instruction Make a Difference |
| 7      | Case #7 | REFERENCE | 8        | Science Motivation Questionnaire: Construct Validation with Nonscience Majors                                                                 |
| 8      | Case #8 | REFERENCE | 17       | Digital Game-Based Learning in High School Computer Science Education: Impact                                                                 |

The table provides basic information on each hit, such as the case number and label, as well as the document variable in which it was found, the hit location, etc. To sort this list of hits in ascending order on any column values, simply click this column header. Clicking a second time on the same column header sorts the rows again in descending order.

Selecting an item in this table, either by clicking it or by using the keyboard cursor keys such as the UP or DOWN arrow keys, automatically displays the corresponding case and document in the main window. This feature provides a way to examine the retrieved segment in its surrounding context. You can also assign an existing code to the highlighted segment.

### To assign a code to a retrieved section:

- In the table of search hits, select the row corresponding to the section you want to code.
- Use the CODE drop-down list located above this table to select the code you want to assign.
- Click the  button to assign the selected code to the highlighted section.

### To assign a code to all retrieved sections:

- Use the CODE drop-down list located above this table to select the code that you want to assign.
- Click the  button to assign the selected code to all sections meeting the search expression.

Prior to the assignment of a code to all search hits, you may want to remove selected hits that do not correspond to what you were looking for. To remove a search hit from the list, select its row and then click the  button.

### To display the automatic text classification panel:

- Click the  button. A panel will appear at the bottom of this hit list, allowing you to apply an automatic text classification model on the retrieved text segments. Clicking this button again, will hide the panel.

For more information on this feature, see **Automatic Text Classification** on page 213.

### To analyze the text of all retrieved segments using WordStat:

- Click the  button. QDA Miner will minimize itself and run WordStat, allowing you to perform a quantitative content analysis or apply text-mining techniques on the retrieved text segments. WordStat will also allow you to perform keyword comparisons across any other column in the table. To include additional data for comparison analysis, use the **Add Variables** option.

### To display the coding table:

- Click the  button. A dialog box will appear allowing you to code retrieved text segments using drag and drop operations, remove existing codings and attach comments to codes.

For more information on this feature, see **Coding Text Segments Using the Coding Table** on page 222.

### To export the table to disk:

- Click the  button. A **Save File** dialog box will appear.
  - In the **Save As Type** list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file (\*.CSV), MS Word (\*.DOC), HTML file (\*.HTM; \*.HTML), XML files (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (.SAV).
  - 
  - Type a valid file name with the proper file extension.
- Click the **Save** button.

### To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.
- To append only a single hit from the table, select the row containing the text segment you want to append to the Report Manager and click the  button.

For more information on the Report Manager, see on page 230.

### To print the table:

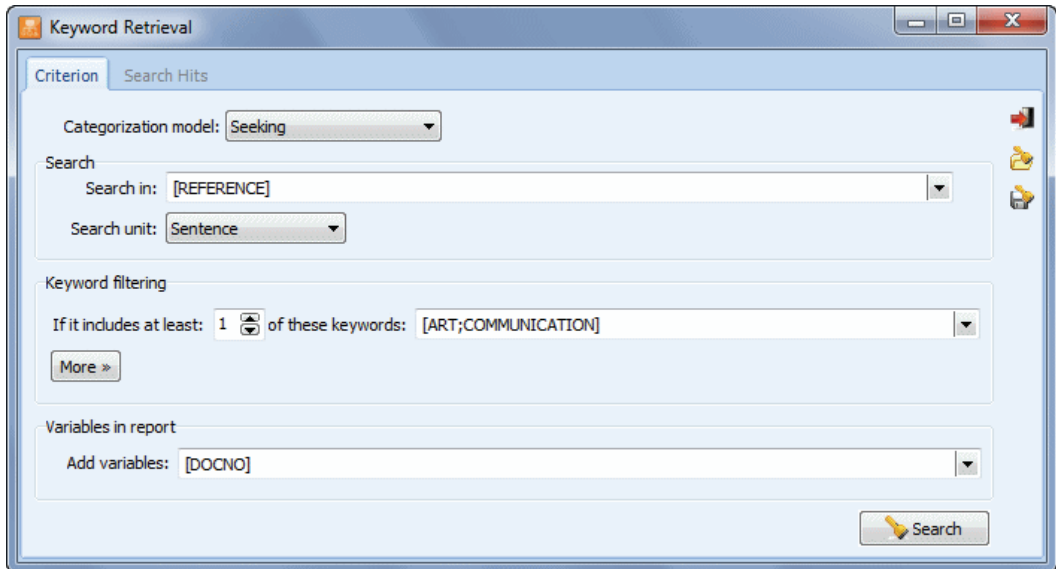
- Click the  button.

# Keyword Retrieval

The KEYWORD RETRIEVAL feature can be used to retrieve any document, paragraph, sentence, or coded segment containing a specific keyword or a combination of keywords associated with a QDA Miner code or stored in a WordStat content analysis dictionary. Text units corresponding to the search criteria are returned in a table on the **Search Hits** page. This table may then be printed or saved to disk. It may also be used to create tabular or text reports as well as to attach codes to the retrieved text segments.

QDA Miner allows one to attach to any code in the codebook a list of words, word patterns and key phrases that are likely to be found in segments where such a code may apply. For example, if you have a code named Globalization, you may enter words like "Free Trade", "World Trade", "NAFTA", or "WTO". Once these keywords have been entered, the Keyword Retrieval feature may be used to search for any text segment that contains any one of those words. In a sense, a list of keywords is a kind of predefined search that is stored with a code within the codebook.

Keywords may also be stored externally in categorization model files. Categorization models are advanced content analysis processes created by WordStat 5.0 or later and saved to disk in a .wcat file. A WordStat categorization model may involve various text processing such as stemming, lemmatization and word exclusions. It also involves the categorization of words, word patterns, and phrases and may also include complex coding rules involving boolean and proximity operators. Such rules may be used to perform disambiguation of words or coding of complex actions. In order to be used within QDA Miner, categorization models should be stored in a Models subfolder located under the main program folder.



QDA Miner comes with a sample categorization model developed from a content analysis of the Seeking project file, allowing users to test this feature without the need to install WordStat. This categorization model consists of 13 content categories with items such as APPEARANCE, ART, COMMUNICATION, FAMILY, etc.

The **CATEGORIZATION MODEL** option lists all WordStat categorization models available in the Models subfolder. If keywords have been stored with the current project codebook, the list will also include the **<Internal keyword>** item. The first step one needs to do in order to access the keyword retrieval features is selecting a categorization model. Once a model has been selected, the other search and formatting options become available.

The **SEARCH IN** option allows you to specify on which document variables the search will be performed. If the current project contains more than one document variable, you will have a choice of selecting either one or a combination of them. By default, all document variables are selected. To restrict the analysis to only a few of them, click the down arrow key at the right of the list box. You will be presented with a list of all available document variables. Select the variables on which you want the search to be performed.

The **SEARCH UNIT** option determines the search unit on which the search will be performed as well as what will be retrieved. You can select three different search units:

- If you select **Documents** as the search unit, QDA Miner will apply the search expression on each document associated with a specific case. If a specific document meets the search condition, its location will be returned.
- When selecting **Paragraphs** as the search unit, QDA Miner will return any paragraph meeting the search condition.
- Select **Sentences** to instruct QDA Miner to return sentences meeting the search condition.
- Selecting **Coded Segments** as the search unit allows restricts the search to the text segments that have already been coded. When a coded segment meets the search condition, its entire text is returned, no matter whether it is a single word or several paragraphs. To restrict the search to specific codes, select the **Selected** radio button to the right side of this option and select the codes from the drop-down checklist by clicking the arrow button to the right end of this list. These codes can also be selected from a tree representation of the codebook by clicking the  button. To perform the search on all codes, choose the **All** radio button.

**KEYWORD FILTERING** - This group of options allows one to select the keyword category on which the retrieval will be based. Setting the numerical value of **the Include at least** option allows one to retrieve text units containing a minimum number of keywords from a selected list. To select the keyword categories, click the arrow button to show all available categories and then click the desired items. The minimum number of keyword categories a specific unit must contain in order to be retrieved is specified using a small edit box with spin buttons on the right. Setting this numerical value to the total number of keyword categories selected will force the program to retrieve only those units containing all those categories. Setting this number to a lower value will retrieve all text units containing at least this number of categories from the selected list. In the example shown above, any unit containing keywords from at least one of the two categories ART and NIGHTLIFE will be retrieved.

To enter a second filtering condition, click the  button. You can choose to link the two filtering conditions using either one of the three Boolean expressions: AND, OR or NOT. Choosing AND will retrieve all text units fulfilling both criteria; selecting OR will result in a retrieval of text

units meeting either the first or the second condition, or both, while choosing the NOT Boolean operator will retrieve text units meeting the first condition but not the second one.

To limit the filtering conditions to a single statement, click the  button.

**ADD VARIABLES** - This drop-down checklist box may optionally be used to add the values stored in one or more variables to the table of retrieved segments for the specific case from which a text segment originates. This feature is especially useful for identifying differences in how keywords are used or how specific topics are expressed by different group of individuals. It may also be used to examine the relationship between these values and the keywords.

To store the search expression and options, click the  button and specify the name under which those query options will be saved.

To retrieve a previously saved query, click the  button and select from the displayed list the name of the query you would like to retrieve (for more information, see **Saving and Retrieving Queries** on page 108).

Once all the search options have been set properly, simply click the  button to retrieve the selected text units.

## Working with the Retrieved Text Units

The retrieved text units are displayed in a table found on the SEARCH HITS page. This table contains both the case number and the variable from which the segment originates, as well as the value of all additional variables selected by the user (see the ADD VARIABLES option above). When searching for paragraphs, sentences or coded segments, the table also displays the text associated with the retrieved unit and its location (its paragraph and sentence number). By using arrow keys or by clicking a row the associated text is displayed in a separate window at the bottom of the screen with all keywords in bold.

To sort the table of retrieved units in ascending order in any column, simply click the column header. Clicking the same column header a second time sorts the rows in descending order. Tables may also be printed, stored as a text report, or exported to disk in various file formats such as Excel, ASCII, MS Word, HTML or XML.

### To remove a search hit from the hit list:

- Select its row and then click the  button.

### To assign a code to a specific search hit:

- In the table of search hits, select the row corresponding to the text segment you want to code.
- Use the **CODE** drop-down list located above this table to select the code to assign.
- Click the  button to assign the selected code to the highlighted text segment.

### To assign a code to all search hits:

- Use the **CODE** drop-down list located above this table to select the code to assign.
- Click the  button to assign the selected code to all text segments matching the search expression.

### To create a report of retrieved segments:

- Click the  button. The sort order of the current table is used to determine the display order in the report. This report is displayed in a text-editing dialog box and may be modified, stored on disk (in RTF, HTML or plain text format), printed, or cut-and-pasted into another application. Graphics and tables may also be inserted anywhere in this report.

### To display the automatic text classification panel:

- Click the  button. A panel will appear at the bottom of this hit list, allowing you to apply an automatic text classification model on the retrieved text segments. Clicking this button again, will hide the panel.

For more information on this feature, see **Automatic Text Classification** on page 213.

### To analyze the text of all retrieved segments using WordStat:

- Click the  button. QDA Miner will minimize itself and run WordStat, allowing you to perform a quantitative content analysis or apply text-mining techniques on the retrieved text segments. WordStat will also allow you to perform keyword comparisons across any other column in the table. To include additional data for comparison analysis, use the **Add Variables** option.

### To display the coding table:

- Click the  button. A dialog box will appear allowing you to code retrieved text segments using drag and drop operations, remove existing codings and attach comments to codes.

For more information on this feature, see **Coding Text Segments Using the Coding Table** on page 222.

### To export the table to disk:

- Click the  button. A **Save File** dialog box will appear.
- In the Save As Type list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file (\*.CSV), HTML file (\*.HTM; \*.HTML), XML file (\*.XML), MS Word document (\*.DOC), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).
- Type a valid file name with the proper file extension.
- Click the **Save** button.

### To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.
- To append only a single hit from the table, select the row containing the text segment you want to append to the Report Manager and click the  button.

For more information on the Report Manager, see the **Report Manager Feature** topic on page 230.

### To print the table:

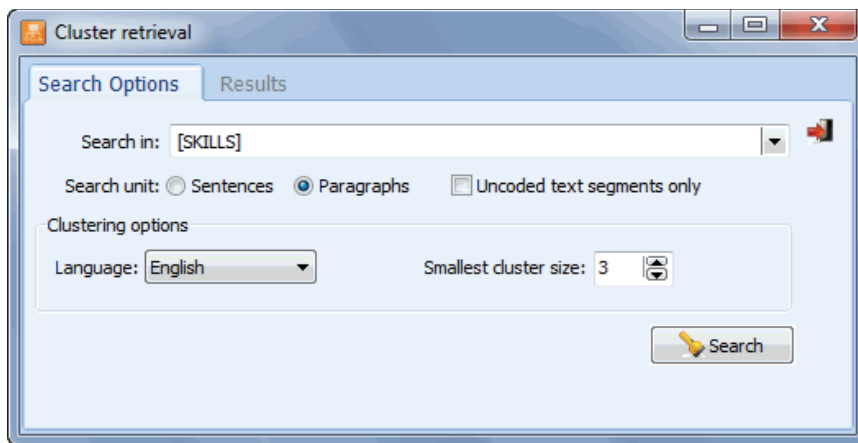
- Click the  button.

# CLUSTER EXTRACTION

The CLUSTER EXTRACTION tool is a unique retrieval feature that allows one to quickly identify all similar paragraphs or sentences in a large collection of documents. This tool relies on a patent-pending algorithm that can quickly compare up to 32,000 text segments and group the most similar ones into clusters. The amount of items that may be clustered depends on the amount of memory available. If all extracted segments cannot fit into memory, QDA Miner will offer the possibility to select a random sample to be clustered. The grouping performed is resistant to spelling errors and may also bring together various inflected forms of nouns and verbs. Once computed, clusters of items may then be coded or modified using intuitive drag-and-drop operations. Un-clustered items will also be suggested based on their level of similarity to existing clusters.

Such a tool can speed up the coding process of large quantities of open-ended responses, Twitter feeds or short comments by a factor that may vary from two to 50 times faster or more, compared to what someone would achieve by manually coding those responses one at a time.

To start the clustering process, select the CLUSTER EXTRACTION command from the ANALYSIS menu. The following dialog box will appear:



The **Search In** option allows you to specify which document variables to search. If the current project contains more than one document variable, you will have a choice of selecting either one or more, or a combination of them. To select different document variables, click the Down arrow at the right of the list box. You will be presented with a list of all available variables. Select the desired document variables.

The **Search Unit** option determines the search unit on which the clustering will be performed as well as what will be retrieved. You can select two different search units:

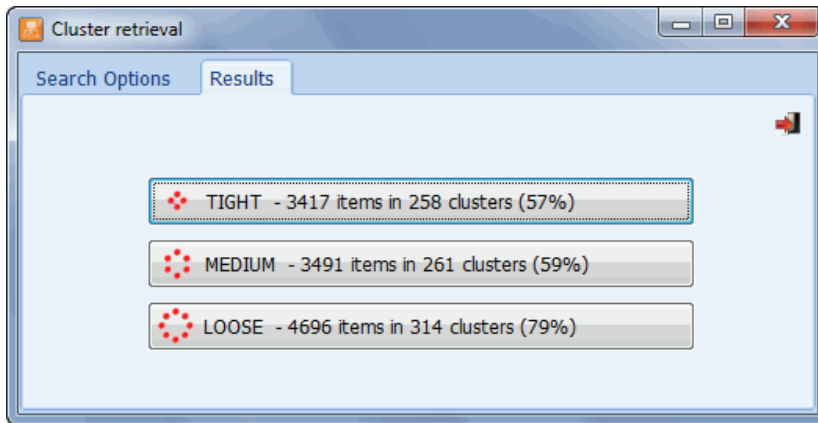
- **Paragraphs**, where QDA Miner will cluster all paragraphs meeting the search condition, and
- **Sentences**, to instruct QDA Miner to select sentences meeting the search condition.

One may restrict the retrieved text segments to those that have not been previously coded by enabling the **Uncoded Text Segment Only** check box.

The **Language** option instructs the program to optimize the clustering by taking into account some linguistic features of the language used in the project documents. If the language used is not on the list, simply set this option to **<none>**.

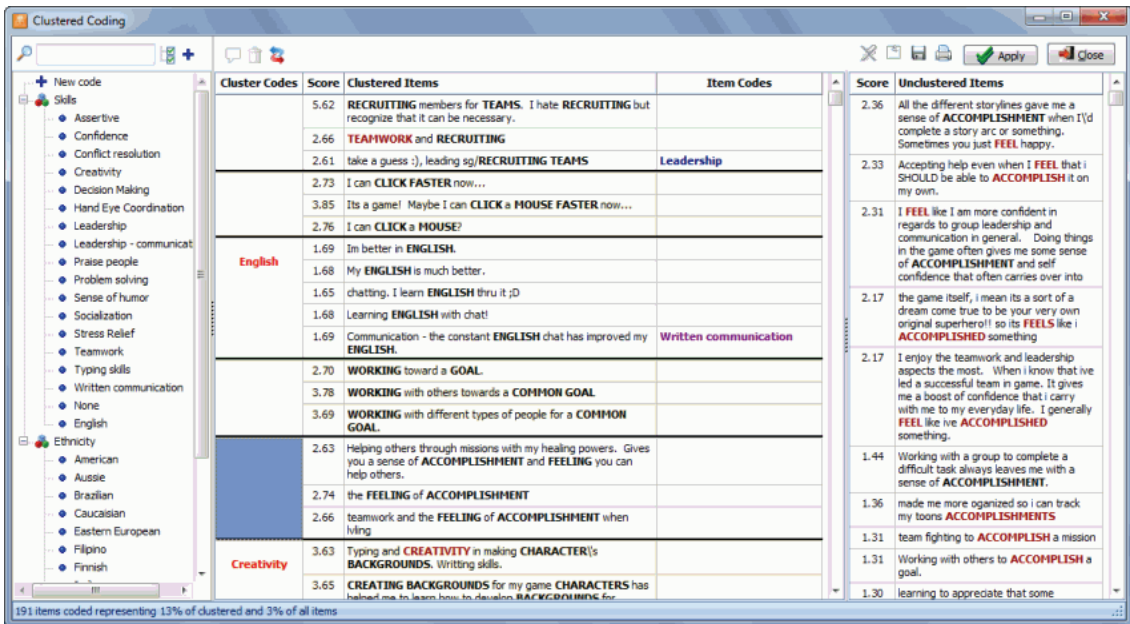
The **Smallest Cluster Size** option may be used to specify the minimum number of items that should be clustered together in order to be retrieved and displayed. The minimum value for this option is 2. Increasing this numeric value reduces the number of clusters retrieved, but allows one to focus on the groups of items that are the most frequent.

When clicking the  button, the software splits long documents into either paragraphs or sentences and compares all extracted text segments to each other and then groups together those that are the most similar. It then displays a **Results** page (offering a choice between three solutions) that looks like this:



The **Tight** cluster shows only clusters of items that are very similar to each other. In the above example, such a solution grouped 3,417 items into 258 clusters, representing 57 percent of all items searched. Choosing the **Loose** option typically shows a larger number of items grouped into clusters containing less-similar items. In the above example, a total of 4,696 items grouped into 314 clusters were found, representing 79 percent of all extracted text segments. The **Medium** button gives access to an intermediary solution.

Clicking any one of these buttons presents a coding dialog box similar to this one, allowing one to code and modify a chosen solution:



The dialog box shown above is the main coding tool. It provides numerous possibilities to adjust clusters by adding or removing some text segments, assigning codes to whole clusters or single segments, creating new codes, attaching comments, etc.

The following section provides a detailed description of this dialog box and its different sections:

- The left panel contains all existing codes that may be assigned to clusters and individual items.
- The middle panel contains a large table where all clusters of text segments are displayed. Clusters are delineated by thick horizontal lines that spread across the entire table, while the text segments they contain are delineated by thinner lines, spreading from the second column up to the right edge of the spreadsheet. This table contains four columns:

**Column 1** is where codes that apply to the entire cluster can be stored.

**Column 2** displays a numerical value that indicates how similar the text segment in this row is to the other text segments in the same cluster. The higher the score, the more representative the associated text segment is to its cluster.

**Column 3** contains the text of the clustered item. Keywords that are shared by most of the cluster are in uppercase and bold characters. Related words, including different inflections as well as misspellings, are also highlighted in red.

**Column 4** is a placeholder where codes associated to a single item can be viewed or stored.

- The right panel displays a table containing all unclustered items that share some similarities to the currently selected cluster. Like clustered items in the main table, keywords are highlighted. A score displayed on the left of the text segment indicates its level of similarity with items in the selected cluster. Those suggested unclustered items are sorted in descending order of similarity.

## Cluster Editing and Coding Operations

### To remove an item from its cluster:

- Select the item to be removed then perform one of the following four actions:
  1. Drag and drop the selected row into the unclustered items section
  2. Right-click and select the REMOVE command.
  3. Click the  button located in the top tool bar.
  4. Press the CTRL-R key combination.

### To add an unclustered item to a cluster:

- Select the unclustered items and then perform one of the following actions:
  1. Drag them to the destination cluster.
  2. Right-click and select MOVE TO CLUSTER.
  3. Press the CTRL-M key combination.

Suggested unclustered items are displayed in descending order of similarity to the selected cluster. For this reason, QDA Miner expects the dragged items to be dropped into this selected cluster. If the item is dropped into another cluster, the software will ask you to confirm this move to another cluster.

### To move a clustered item to a different cluster:

- Select the row to be moved.
- Drag and drop it anywhere in the new cluster.

### To split a cluster in two:

Sometimes, a cluster contains items that are related to two different topics and need to be coded separately. Because of the hierarchical clustering process, items related to a single topic are often near each other. To split a cluster:

- Select the row where the cluster should be split.
- Right-click then select either the SPLIT ABOVE or SPLIT BELOW command.

### To merge two clusters:

- Select the first column of the cluster.
- Drag and drop it anywhere on top of the second cluster.

Merging to an adjacent cluster can also be performed by right-clicking to display a pop-up menu and selecting the MERGE WITH PRIOR or MERGE WITH NEXT command. When codes have been associated with either one or both of the clusters, the resulting cluster inherits all the codes.

### To assign a code to a cluster:

- Select the code in the left panel and drop it anywhere in the first column of the cluster to be coded.
- or
- Select the first column of the cluster to be coded, drag it toward the left panel and drop it on the desired code.

### To assign a code to a single text segment:

- Select the code in the left panel and drop it in the corresponding row (anywhere from the second to the fourth column)
- or
- Select the row containing the item to be coded, drag it to the left panel and drop it over the appropriate code.

### To assign a new code to a cluster or segment:

- Drag the selected cluster or text segment to be coded toward the left panel and drop it on the NEW CODE item located at the top of the section or right-click and choose the ADD TO A NEW CODE command. A code editing box will be displayed with a suggested name.
- Edit the suggested name or type a new name, then click the **OK** button.

Alternatively, one may create the code beforehand by clicking the  button and then use previously described methods for assigning existing codes to clusters or single text segments.

### To remove an existing code:

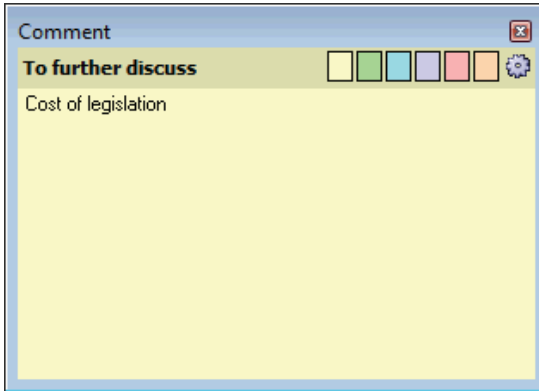
- Select the text segment or cluster containing the code you would like to remove.
- Click the  button located in the toolbar or right-click and select REMOVE. A list of all codes assigned to the current row or cluster will be displayed.
- Select the code that you would like to remove.

### To assign or remove codes using check marks:

- Click the  button to display check boxes.
- Select the text segment or the cluster you would like to code.
- Click the check boxes beside the codes you would like to assign to the selected row or cluster, or clear any existing check marks beside codes you would like to remove..

## To attach a comment to a coded segment or edit an existing one:

- Click the  button.
- Select the code to which you want to attach a comment. A small window like this will appear:
- Enter the text that you want to associate with the selected code, or edit the existing text.



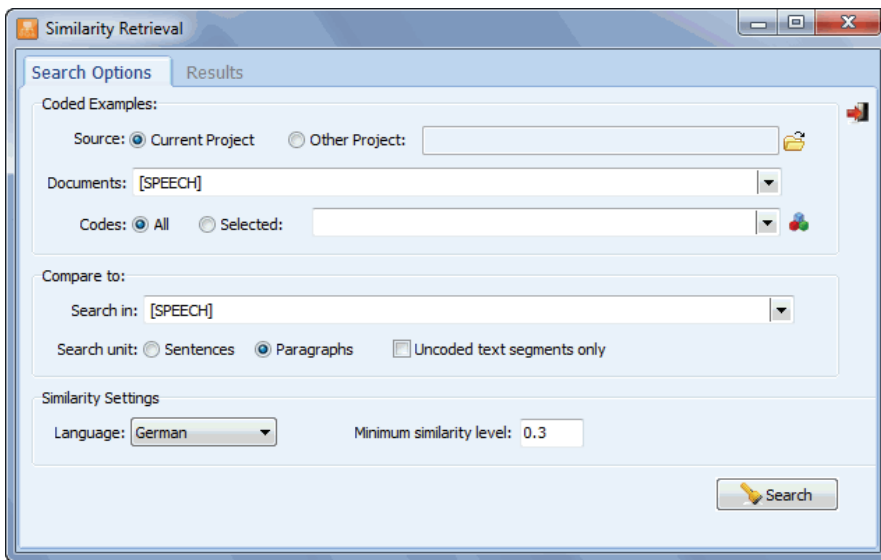
- Click the X button in the upper-right corner to save the comment and to close the window.
- When a note has been assigned to a coded segment, an exclamation point between square brackets is displayed on the left of the code name. To edit an existing comment, follow the above steps. To remove a comment, simply open the Note Editor and delete all text in the editing window.

Changes made using the coding table are not implemented immediately but are kept in memory. To implement all changes made, click the  button. It is recommended to click this button regularly after making several changes. To close the dialog box and cancel all changes, click the [PICTURE CloseBtn.bmp] button.

# Code Similarity Retrieval

The Code Similarity tool allows one to quickly identify sentences or paragraphs that share some similarities with existing coded segments in the current project or those stored in another coded project. Such a tool is especially useful on partially coded projects to quickly identify and code text segments in un-coded documents. It may also be useful to improve the consistency of the coding on fully coded projects by identifying items that may have been overlooked. One may also recall all coding performed on another project and identify sentences or paragraphs that are similar to coded segments in this other project, whether or not the associated code exists in the current one. Using an external project as a reference set is especially appropriate when both projects contain similar types of documents to be coded in a similar way.

To use this feature, run the CODE SIMILARITY command from the RETRIEVAL menu. A dialog box similar to this one will appear:



The Source option allows you to specify whether the coded examples are in the **Current Project** or consist of coded segments stored in another project. When **Other Project** is selected, an **Open File** dialog box appears, allowing you to browse through your computer to choose another project. The external project may later be changed by clicking the  button and choosing another project file.

The **Document** option allows you to specify which document variables to search. If the selected project contains more than one document variable, you will have a choice of selecting one or a combination of more than one of them. To select different document variables, click the down arrow at the right of the list box. A list of all available variables will be displayed. Select the document variables on which you want the search to be performed.

The **Search Unit** option determines the search unit on which the clustering will be performed as well as what will be retrieved. You can select two different search units:

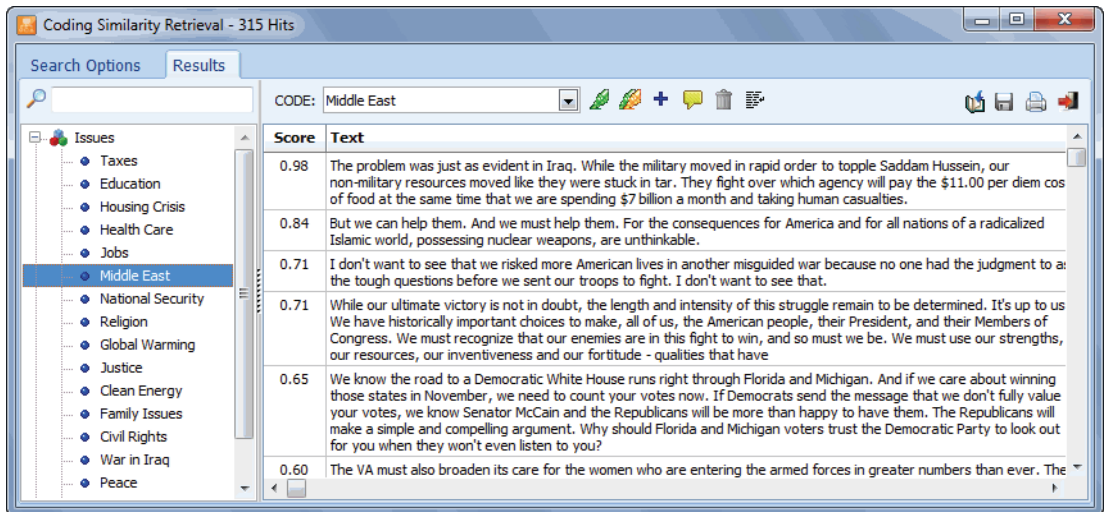
- Paragraphs where QDA Miner will extract all paragraphs in the current project; and
- Sentences where QDA Miner will extract sentences rather than paragraphs.

One may restrict the retrieved text segments to those that have not been previously coded by enabling the Uncoded Text Segment Only check box.

The **Language** option instructs the program to optimize the code similarity search by taking into account some linguistic features used in the project documents. If the language used is not on the list, simply set this option to **<none>**.

The **Minimum Similarity Level** option can be used to restrict the display of items to text segments with a similarity value equal to or above the specified value. The similarity score may vary from 0, when items are totally different, to 1 when segments are identical. The lowest value that may be set for this option is 0.1.

Once the appropriate options have been set, click the  button. QDA Miner will retrieve all coded segments, index those and then retrieve and compare all those to extract text segments. Once completed, QDA Miner will display the results in a dialog box similar to this one:



Shown on the left of the dialog box is the list of all codes in the current or in the other project. In the table to the right is the list of all extracted segments that are somewhat similar to the selected code, shown in descending order of similarity. The similarity scores vary between zero and one. Only those scores meeting the **Minimum Similarity Level** option are displayed. To display text segments similar to another code, simply select this new code from the left panel.

### To assign a code to the currently selected hits:

- In the table of search hits, select the rows corresponding to the text segments you want to code.
- Use the CODE drop-down list located above this table to select the code you want to assign.

- Click the  button to assign the selected code to the highlighted text segments.

### To assign a code to all search hits:

- Use the CODE drop-down list located above this table to select the code that you want to assign.
- Click the  button to assign the selected code to all text segments meeting the search expression.

Prior to the assignment of a code to all search hits, you may want to remove selected hits that do not correspond to what you were looking for. To remove a search hit from the list, select its row and then click the  button.

### To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.
- To append only a single hit from the table, select the row containing the text segment you want to append to the Report Manager and click the  button.

For more information on the Report Manager, see the **Report Manager Features** topic on page 230.

### To export the table to disk:

- Click the  button. A **Save File** dialog box will appear.
- In the **Save As Type** list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file (\*.CSV), MS Word (\*.DOC), HTML file (\*.HTM; \*.HTML), XML files (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).
- Type a valid file name with the proper file extension.
- Click the **SAVE** button.

### To print the table:

- Click the  button.

# Query by Example

The QUERY BY EXAMPLE search tool is an alternative to a keyword-based search. Rather than specifying a long list of keywords joined by Boolean operators, this query method looks for documents or text segments similar to a given example and presents them in an initial list of search hits in descending order of similarity. The user then marks specific items as relevant or irrelevant and then performs a new query. The system adjusts the initial query by taking into account the provided relevance information and often provides better search results. After a single or a few iterations, the user can then select relevant items for further processing.

QDA Miner's Query by Example feature allows one to provide as the starting example, a typed text - which can consist of a sample paragraph or sentence, a list of words or phrases, a selected text segment, or even all text segments in the project currently associated with one or several codes.

To start the query by example search process, select the QUERY BY EXAMPLE command from the ANALYSIS menu. The following dialog box will appear:

Query by Example

Search Criterion   Search Results   Selected Hits

Retrieve

Unit: Paragraphs   In: [SPEECH]

Starting Example:

☒ Sample text   ☐ Coded segments   ☐ Other project:

We must restore this nation's military power

Options

☒ Fuzzy string matching

☐ Do not retrieve already tagged segments

Minimum criteria: 0.001   Maximum suggestions: 200

Language: English

Add variables: [CANDIDATE]

Search

Another way to initiate the QUERY BY EXAMPLE is to highlight a text segment in the main text editor, right-click and select the FIND SIMILAR command. You will be presented with a submenu allowing you to search for similar sentences, paragraphs or documents. Selecting any one of those three items will automatically start a search and display the search results. Selecting the OPTIONS command will bring you to the first page of the dialog box, and allow you to adjust the various search options.

The **UNIT** option determines the search unit on which the search will be performed as well as what will be retrieved. You can select three different search units:

If you select **Documents** as the search unit, QDA Miner will apply the search expression on each document associated with a specific case. If a specific document meets the search condition, its location will be returned.

When selecting **Paragraphs** as the search unit, QDA Miner will return any paragraph meeting the search condition.

Select **Sentences** to instruct QDA Miner to return sentences meeting the search condition.

The **IN** option allows you to specify the document variables QDA Miner is to search. If the current project contains more than one document variable, you will have a choice of selecting either one or a combination. By default, all document variables are selected. To restrict the analysis to only a few of them, click the down arrow key at the right of the list box. You will be presented with a list of all available document variables. Select the variables for which you want the search to be performed.

The **Starting Example** section allows you to choose whether the starting example will be a **Sample Text** typed or pasted from the clipboard in an edit box or whether it will be based on existing **Coded Segments** associated with one or several codes in the current project. One may also select coded segments stored in an external project by selecting the **Other project** option. When this last option is chosen, two lists of codes are displayed. The **Examples** list lets you select one or several codes whose segments will be used as the starting examples on which the search for similar segments will be based. The **Non-Examples** list may be used to indicate coded segments that do not correspond to what you are looking for. This second set of coded segments allows the program to focus more precisely on items similar to the starting examples. For this reason, it may be a good idea to provide non-examples that are close in content to the starting examples. For example, if you are looking for paragraphs about globalization of economy, providing only globalization examples may retrieve paragraphs that are associated not only with globalization but also paragraphs containing words related to economy or foreign policy (names of other countries). If you provide non-examples, sample text segments about local economy or foreign policy that are not related to globalization per se, the program will apply less weight to words more specific to the economy and foreign policy and less related to the topic of globalization and more weight to words specifically about globalization.

The following additional search options are available:

**FUZZY STRING MATCHING** - This option allows one to match documents even if they do not necessarily share the exact same relevant words. Computation of similarity is normally achieved by looking at shared words and requires exact matches of those words. Such an approach can perform poorly on documents containing misspelled words and may also fail to retrieve relevant documents using slightly different word forms. If the starting example has the word "aggression", QDA Miner may fail to match text segments containing words like "aggressive" or "aggressiveness" and will also fail to match other segments containing misspelled forms such as "agression". Enabling the Fuzzy String Matching algorithm relaxes the matching condition for words and retrieves more text segments than are likely to be relevant. Please note however, that this option may also increase the number of irrelevant results retrieved.

**DO NOT RETRIEVE ALREADY TAGGED SEGMENTS** - When searching sentences or paragraphs, this option restricts the retrieved text segments to those that have not been previously coded. Such an option is especially useful when you are attempting to find, among uncoded text segments those that are similar to existing coded segments and that could potentially receive the same code.

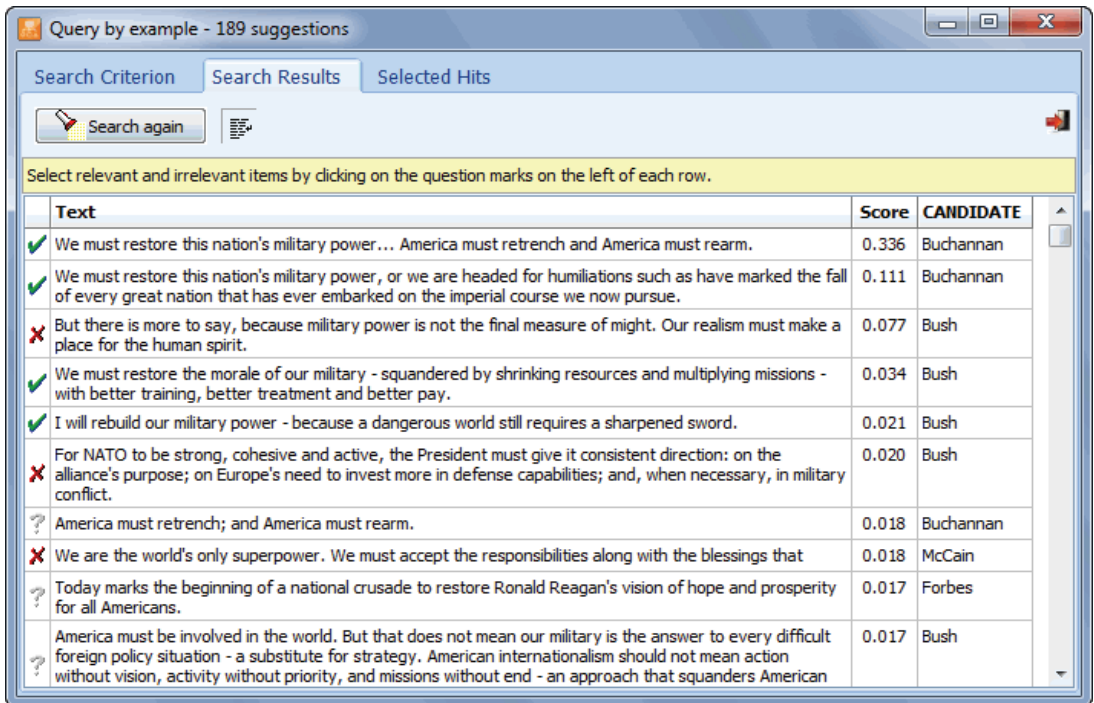
**MINIMUM CRITERIA** - A similarity score is computed for each retrieved segment. This score varies between 0.0 for totally different segments to 1.0 for segments identical to the starting example. This option allows you to set a minimum value for this similarity score. In order for a text segment to be retrieved and displayed, it should obtain a similarity score equal to or higher than this minimum value. By default, this option is set to 0.001.

**MAXIMUM HITS** - This option allows you to restrict the number of retrieved segments to a specified number. By default, up to 200 items are retrieved.

**LANGUAGE** - This option instructs the program to optimize the search query by taking into account some linguistic features of the language used in the project documents. If the language used is not on the list, simply set this option to **<none>**.

**ADD VARIABLES** - This drop-down checklist box may optionally be used to add values stored in one or more variable to the table of retrieved segments for the specific case from which a text segment originates.

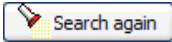
To perform the search, click the  button. The results of a search are displayed in a table located on the **Search Results** page, similar to the one below:

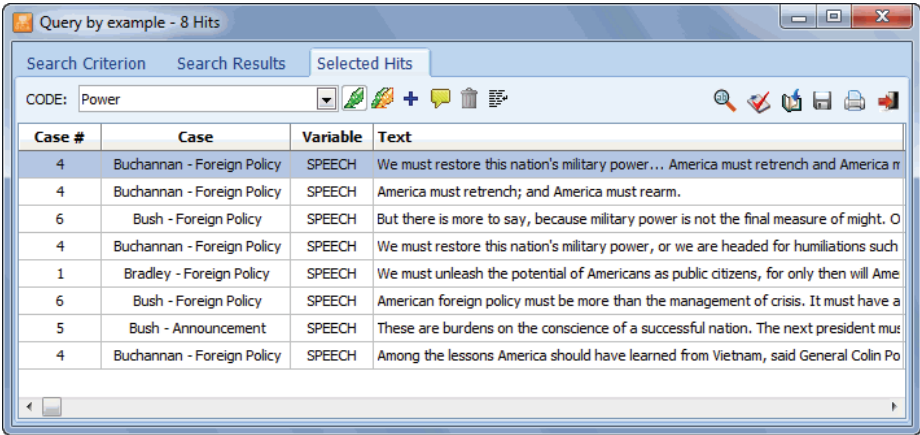


The screenshot shows a window titled "Query by example - 189 suggestions". It has three tabs: "Search Criterion", "Search Results", and "Selected Hits". The "Search Results" tab is active. Below the tabs, there is a "Search again" button and a list icon. A yellow instruction bar says: "Select relevant and irrelevant items by clicking on the question marks on the left of each row." Below this is a table with four columns: a checkbox column, a "Text" column, a "Score" column, and a "CANDIDATE" column. The table contains 10 rows of search results, each with a checkbox (green checkmark for relevant, red X for irrelevant, or question mark for undecided), a text snippet, a similarity score, and a candidate name.

|                                     | Text                                                                                                                                                                                                                                                                                                                         | Score | CANDIDATE |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-----------|
| <input checked="" type="checkbox"/> | We must restore this nation's military power... America must retrench and America must rearm.                                                                                                                                                                                                                                | 0.336 | Buchanan  |
| <input checked="" type="checkbox"/> | We must restore this nation's military power, or we are headed for humiliations such as have marked the fall of every great nation that has ever embarked on the imperial course we now pursue.                                                                                                                              | 0.111 | Buchanan  |
| <input type="checkbox"/>            | But there is more to say, because military power is not the final measure of might. Our realism must make a place for the human spirit.                                                                                                                                                                                      | 0.077 | Bush      |
| <input checked="" type="checkbox"/> | We must restore the morale of our military - squandered by shrinking resources and multiplying missions - with better training, better treatment and better pay.                                                                                                                                                             | 0.034 | Bush      |
| <input checked="" type="checkbox"/> | I will rebuild our military power - because a dangerous world still requires a sharpened sword.                                                                                                                                                                                                                              | 0.021 | Bush      |
| <input type="checkbox"/>            | For NATO to be strong, cohesive and active, the President must give it consistent direction: on the alliance's purpose; on Europe's need to invest more in defense capabilities; and, when necessary, in military conflict.                                                                                                  | 0.020 | Bush      |
| <input type="checkbox"/>            | America must retrench; and America must rearm.                                                                                                                                                                                                                                                                               | 0.018 | Buchanan  |
| <input type="checkbox"/>            | We are the world's only superpower. We must accept the responsibilities along with the blessings that                                                                                                                                                                                                                        | 0.018 | McCain    |
| <input type="checkbox"/>            | Today marks the beginning of a national crusade to restore Ronald Reagan's vision of hope and prosperity for all Americans.                                                                                                                                                                                                  | 0.017 | Forbes    |
| <input type="checkbox"/>            | America must be involved in the world. But that does not mean our military is the answer to every difficult foreign policy situation - a substitute for strategy. American internationalism should not mean action without vision, activity without priority, and missions without end - an approach that squanders American | 0.017 | Bush      |

All results are presented in a list in descending order of similarity. This table is used to select relevant items for further processing or to improve the query results by providing relevance feedback. Initially, all items are marked with a question mark ( ? ) icon on the far left side. Clicking once on this question mark changes the question mark into a check mark ( ✓ ) indicating that this item is relevant. Using this check mark also selects it for further processing. Clicking it again changes it to a ✗ mark which indicates that this result is not relevant and that further queries should ignore items like this one. Clicking it again will change it back to the question mark.

Once a few relevant and non-relevant items have been marked, clicking the  button performs another search and attempts to retrieve more relevant items. After one or more iterations, you will find that no significant improvement is made by providing additional feedback. When this occurs, examine entries on the list and make sure a ✓ mark appears beside relevant ones. Then move to the **Selected Hits** page.



| Case # | Case                      | Variable | Text                                                                                          |
|--------|---------------------------|----------|-----------------------------------------------------------------------------------------------|
| 4      | Buchanan - Foreign Policy | SPEECH   | We must restore this nation's military power... America must retrench and America must rearm. |
| 4      | Buchanan - Foreign Policy | SPEECH   | America must retrench; and America must rearm.                                                |
| 6      | Bush - Foreign Policy     | SPEECH   | But there is more to say, because military power is not the final measure of might. O         |
| 4      | Buchanan - Foreign Policy | SPEECH   | We must restore this nation's military power, or we are headed for humiliations such          |
| 1      | Bradley - Foreign Policy  | SPEECH   | We must unleash the potential of Americans as public citizens, for only then will Ame         |
| 6      | Bush - Foreign Policy     | SPEECH   | American foreign policy must be more than the management of crisis. It must have a            |
| 5      | Bush - Announcement       | SPEECH   | These are burdens on the conscience of a successful nation. The next president mus            |
| 4      | Buchanan - Foreign Policy | SPEECH   | Among the lessons America should have learned from Vietnam, said General Colin Po             |

The table in this **Selected Hits** page displays a list of each text segment or document that was tagged as relevant in the **Search Results** page. This list includes additional information such as the case number and case label, as well as the document variable in which it was found, the hit location, etc. To sort this list of hits in ascending order on any column values, simply click the column header. Clicking a second time on the same column header sorts the rows again in descending order.

Selecting an item in this table, either by clicking it or by using the keyboard cursor keys such as the UP or DOWN arrow keys, automatically displays the corresponding case and document in the main window. If the search unit was set to **Paragraphs** or **Coded Segments**, selecting a specific hit will also highlight the corresponding text segment. This feature provides a way to examine the retrieved segment in its surrounding context. You can also assign an existing code to the highlighted segment.

### To assign a code to a specific search hit:

- In the table of search hits, select the row corresponding to the text segment you want to code.
- Use the CODE drop-down list located above this table to select the code you want to assign.

- Click the  button to assign the selected code to the highlighted text segment.

### To assign a code to all search hits:

- Use the CODE drop-down list located above this table to select the code that you want to assign.
- Click the  button to assign the selected code to all text segments meeting the search expression.
- Prior to the assignment of a code to all search hits, you may want to remove selected hits that do not correspond to what you were looking for. To remove a search hit from the list, select its row and then click the  button.

### To display the automatic text classification panel:

- Click the  button. A panel will appear at the bottom of this hit list, allowing you to apply an automatic text classification model on the retrieved text segments. Clicking this button again, will hide the panel.

For more information on this feature, see **Automatic Text Classification** on page 213.

### To analyze the text of all retrieved segments using WordStat:

- Click the  button. QDA Miner will minimize itself and run WordStat, allowing you to perform a quantitative content analysis or apply text-mining techniques on the retrieved text segments. WordStat will also allow you to perform keyword comparisons across any other column in the table. To include additional data for comparison analysis, use the **Add Variables** option.

### To display the coding table:

- Click the  button. A dialog box will appear allowing you to code retrieved text segments using drag and drop operations, remove existing codings and attach comments to codes.

For more information on this feature, see **Coding Text Segments Using the Coding Table** on page 222.

### To export the table to disk:

- Click the  button. A **Save File** dialog box will appear.

- In the **Save As Type** list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file (\*.CSV), MS Word (\*.DOC), HTML file (\*.HTM; \*.HTML), XML files (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).
- Type a valid file name with the proper file extension.
- Click the **Save** button.

### To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.
- To append only a single hit from the table, select the row containing the text segment you want to append to the Report Manager and click the  button.

For more information on the Report Manager, see the **Report Manager Feature** topic on page 230.

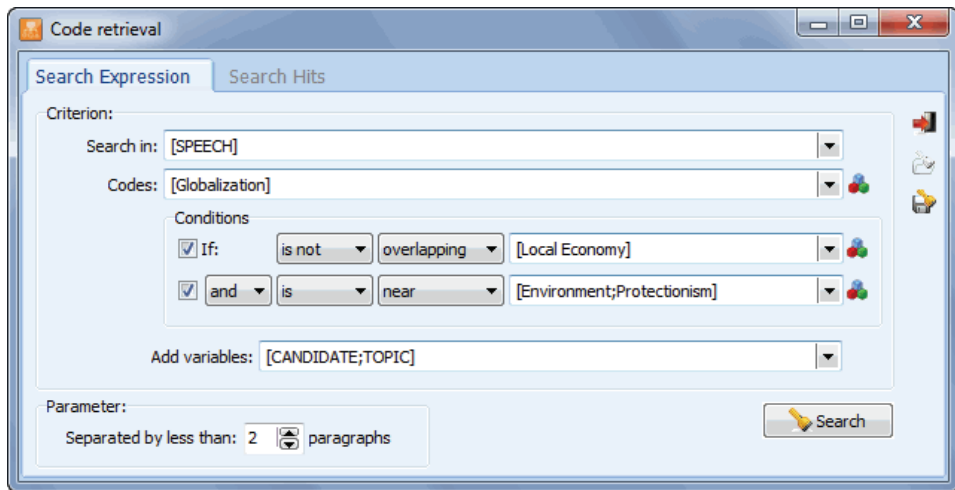
### To print the table:

- Click the  button.

# Coding Retrieval

The CODING RETRIEVAL feature lists all text segments associated with some codes or patterns of codes. Coded segments corresponding to the search criteria are returned in a table on the **Search Hits** page. This table may be printed or saved to disk. It may also be used to create tabular or text reports as well as to assign new codes to the retrieved segments.

To start the code retrieval feature, select the CODING RETRIEVAL command from the ANALYSIS menu. The following dialog box will appear:



## Performing a Simple Code Retrieval

Only the first two options at the top of this dialog box need to be set in order to retrieve coded segments associated with some codes.

**SEARCH IN** - This option allows you to specify the document variables to search. If the current project contains more than one document variable, you will have a choice of selecting either one or a combination of them. By default, all document variables are selected. To restrict the search to only a few of them, click the down arrow key at the right of the list box. You will be presented with a list of all available document variables. Select all variables on which you want the search to be performed.

**RETRIEVE** - This option allows you to select one or more codes to be retrieved. You can select the codes either from a drop-down checklist by clicking the arrow button or from a tree representation of the codebook by clicking the  button.

**ADD VARIABLES** - This drop-down checklist box may optionally be used to add the values stored in one or more variables to the table of retrieved segments for the specific case from which a text segment originates. This feature is especially useful for identifying differences in how codes are

used or how specific topics are expressed by different group of individuals. It may also be used to examine the relationship between these values and the codes.

To store the search expression and options, click the  button and specify the name under which those query options will be saved.

To retrieve a previously saved query, click the  button and select from the displayed list the name of the query you would like to retrieve (for more information, see **Saving and Retrieving Queries** on page 108).

Once the options have been set properly, simply click the  button to retrieve the selected codes.

## Performing a Complex Code Retrieval

QDA Miner can also search for more complex code patterns by providing the option of setting up to two criteria specifying the spatial relationship between the codes. The combination of codes can be defined using one of seven logical operators. The table below provides the list of the available logical conditions that may be used to characterize the code combination. In this table, **Coding A** refers to the coded segments that will be retrieved while **Coding B** refers to the coded segments based on which the condition will be tested. Please note that the reverse of these functions may also be obtained by using a negation operator (see below)

| Function    | Description                                                                                                                                                                                                        |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Equal to    | Coding A will be retrieved if it points to the same text segment as Coding B.                                                                                                                                      |
| Enclosing   | Coding A will be retrieved its text segment entirely enclosed another coded segment assigned to B.                                                                                                                 |
| Included in | Coding A will be retrieved if its text segment is entirely surrounded by another text segment assigned to B.                                                                                                       |
| Overlapping | Code A will be retrieved if its text segment completely or partially overlaps another coded segment assigned to B.                                                                                                 |
| Followed by | Coding A will be retrieved if followed by a text segment assigned to B. The two segments should not overlap and the distance separating the two segments should be less than a user specified distance.            |
| Preceded by | Coding A will be retrieved if it follows a text segment assigned to B. The two segments should not overlap and the distance separating the two segments should be less than a user specified distance.             |
| Near        | Coding A will be retrieved if it follows or precedes a text segment assigned to B. The two segments should not overlap and the distance separating the two segments should be less than a user specified distance. |

Three list boxes are used to set a retrieval condition:

The **first** list box is used to specify whether to include or exclude the coded segments meeting the specified criterion. Setting this list box to **Is** will retrieve any segments meeting the specified

criterion, while setting it to **Is Not** will exclude these segments and retrieve only those that do not meet the criterion.

The **second** list box allows you to choose the logical function that will be applied (see earlier for the list of logical functions). If one of the last three functions is selected (i.e., followed by, preceded by, or near) a new option will be displayed at the bottom of the dialog box. This new option allows you to set what should be the maximum distance, in number of paragraphs, separating the two codes.

The **third** list box is used to select the codes based on which the condition will be tested. If several codes are specified either in the list of codes to be retrieved or in this third list box, they will be treated as if they were joined by the **OR** Boolean operator. In other words, the condition will be met if any combination of codes is encountered. For example, searching for codes A and B to the condition that they must be equal the list of codes C, D and E corresponds to the following formula:

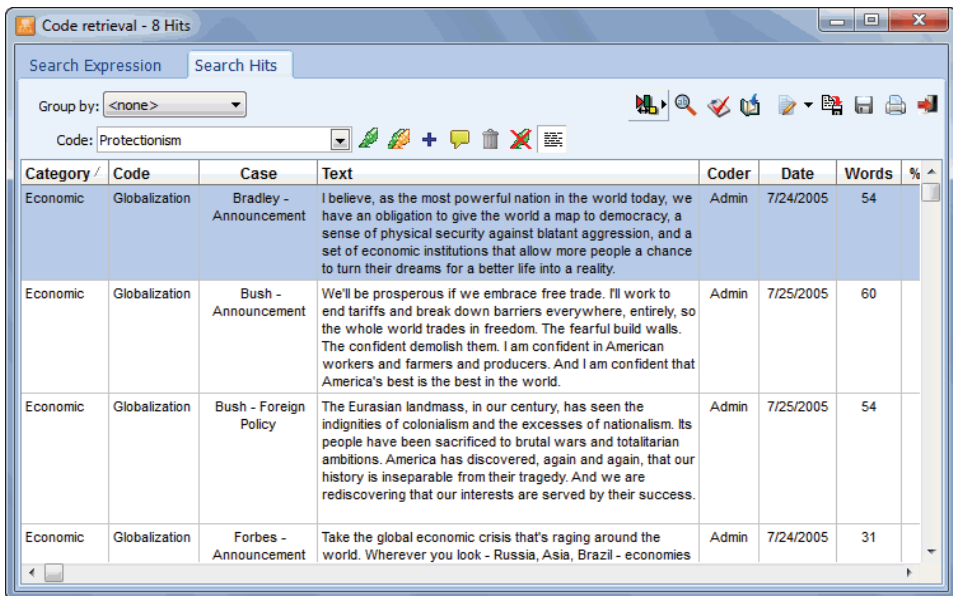
(A or B) is equal to (C or D or E)

Up to two retrieval conditions may be set and combined using a Boolean operator. If the **AND** operator is used, then only these segments meeting both criteria will be retrieved while selecting the **OR** operator will retrieve any segment meeting at least one of these two criteria.

For example, in the dialog box above, QDA Miner will retrieve any text segments coded as *Power* if this segment is within three paragraphs of another text segment coded as *Freedom*, and if it overlaps a text segment coded either as *Ethic* or *Globalization*.

## Working with the Retrieved Segments

The retrieved codes are displayed in a table found on the **Search Hits** page.



Code retrieval - 8 Hits

Search Expression: Search Hits

Group by: <none>

Code: Protectionism

| Category | Code          | Case                   | Text                                                                                                                                                                                                                                                                                                                                                             | Coder | Date      | Words | % |
|----------|---------------|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-----------|-------|---|
| Economic | Globalization | Bradley - Announcement | I believe, as the most powerful nation in the world today, we have an obligation to give the world a map to democracy, a sense of physical security against blatant aggression, and a set of economic institutions that allow more people a chance to turn their dreams for a better life into a reality.                                                        | Admin | 7/24/2005 | 54    |   |
| Economic | Globalization | Bush - Announcement    | We'll be prosperous if we embrace free trade. I'll work to end tariffs and break down barriers everywhere, entirely, so the whole world trades in freedom. The fearful build walls. The confident demolish them. I am confident in American workers and farmers and producers. And I am confident that America's best is the best in the world.                  | Admin | 7/25/2005 | 60    |   |
| Economic | Globalization | Bush - Foreign Policy  | The Eurasian landmass, in our century, has seen the indignities of colonialism and the excesses of nationalism. Its people have been sacrificed to brutal wars and totalitarian ambitions. America has discovered, again and again, that our history is inseparable from their tragedy. And we are rediscovering that our interests are served by their success. | Admin | 7/25/2005 | 54    |   |
| Economic | Globalization | Forbes - Announcement  | Take the global economic crisis that's raging around the world. Wherever you look - Russia, Asia, Brazil - economies                                                                                                                                                                                                                                             | Admin | 7/24/2005 | 31    |   |

Beside the text associated with retrieved coded segments, this table also contains the code name along with its category, the case from which this segment originates, the coder name and the date when it has been assigned. The **Words** column contains the total number of words in the text segment, while the **%Words** column shows the relative importance of this segment versus the entire document, in term of percentage of words. If a comment has been assigned to a specific segment, it will also be included in the table. Lastly, values for all variables that you have chosen to add to the table will be listed at the right end of the table.

To sort this list of retrieved segments in ascending order on any column values, simply click this column header. Clicking the same column header a second time sorts the rows again in descending order.

Hits tables may be printed, coded, stored as a text report, saved to disk either as a new project or exported in a new file format such as Excel, MS Word, ASCII, HTML, or XML format.

### To remove a search hit from the hit list:

- Select its row and then click the  button.

### To assign a code to a specific search hit:

- In the table of search hits, select the row corresponding to the text segment you want to code.
- Use the **Code** drop-down list located above this table to select the code that you want to assign.
- Click the  button to assign the selected code to the highlighted text segment.

### To assign a code to all search hits:

- Use the **Code** drop-down list located above this table to select the code that you want to assign.
- Click the  button to assigned the selected code to all text segments meeting the search expression.

### To create a report of coded segments:

- Click the  button. The sort order of the current table is used to determine the display order in the report. This report is displayed in a text editor dialog box and may be modified, stored on disk in RTF, HTML or plain text format, printed, or cut and pasted into another application. Graphics and tables may also be inserted anywhere in this report.

### To display the automatic text classification panel:

- Click the  button. A panel will appear at the bottom of this hit list, allowing you to apply an automatic text classification model on the retrieved text segments. Clicking this button again, will hide the panel.

For more information on this feature, see **Automatic Text Classification** on page 213.

## To analyze the text of all retrieved segments using WordStat:

- Click the  button. QDA Miner will minimize itself and run WordStat, allowing you to perform a quantitative content analysis or apply text-mining techniques on the retrieved text segments. WordStat will also allow you to perform keyword comparisons across any other column in the table. To include additional data for comparison analysis, use the **Add Variables** option.

## To display the coding table:

- Click the  button. A dialog box will appear allowing you to code retrieved text segments using drag and drop operations, remove existing codings and attach comments to codes.

For more information on this feature, see **Coding Text Segments Using the Coding Table** on page 222.

## To export retrieved codes to disk:

- Click the  button. A **Save File** dialog box will appear.
- In the **Save as type** list box select the file format under which you would like to save the table. The following formats are supported: ASCII file (\*.TXT); Tab delimited file (\*.TAB); Comma delimited file (\*.CSV); HTML file (\*.HTM; \*.HTML); XML file (\*.XML), MS Word document (\*.DOC), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).
- Enter a valid filename with the proper file extension.
- Click the **Save** button.

The obtained table can also be stored in a new project file, where each row becomes a new case and each column is transformed into a variable. This type of project file may be useful for performing a more detailed analysis of coded segments.

## To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.
- To append only a single hit from the table, select the row containing the text segment you want to append to the Report Manager and click the  button.

For more information on the Report Manager, see the **Report Manager Feature** topic on page 230.

### To export retrieved codes into a new project:

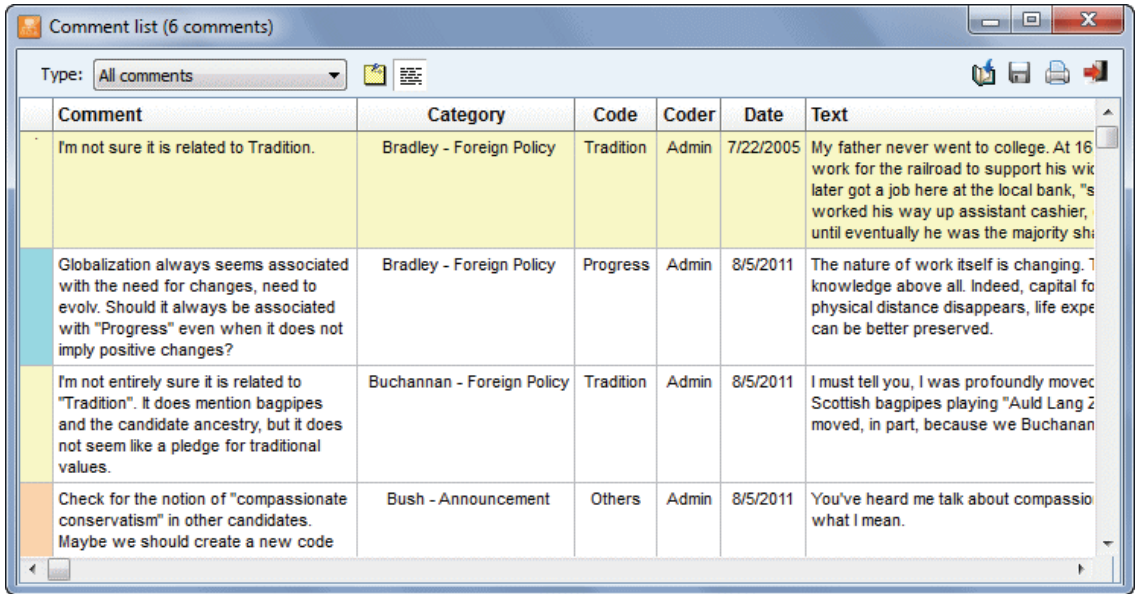
- Click the  button.
- Enter the name of the new project and click the **Save** button.

### To print the table:

- Click the  button.

# Retrieving a List of Comments

To review all comments attached to codings in a project, select the LIST COMMENTS command from the CODES menu. A dialog box similar to this one will appear:



| Comment                                                                                                                                                                       | Category                  | Code      | Coder | Date      | Text                                                                                                                                                                                                                |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|-----------|-------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| I'm not sure it is related to Tradition.                                                                                                                                      | Bradley - Foreign Policy  | Tradition | Admin | 7/22/2005 | My father never went to college. At 16 work for the railroad to support his wife. Later got a job here at the local bank, "s worked his way up assistant cashier, until eventually he was the majority shareholder. |
| Globalization always seems associated with the need for changes, need to evolve. Should it always be associated with "Progress" even when it does not imply positive changes? | Bradley - Foreign Policy  | Progress  | Admin | 8/5/2011  | The nature of work itself is changing. Technology knowledge above all. Indeed, capital for physical distance disappears, life expectancy can be better preserved.                                                   |
| I'm not entirely sure it is related to "Tradition". It does mention bagpipes and the candidate ancestry, but it does not seem like a pledge for traditional values.           | Buchanan - Foreign Policy | Tradition | Admin | 8/5/2011  | I must tell you, I was profoundly moved by Scottish bagpipes playing "Auld Lang Syne" moved, in part, because we Buchanan                                                                                           |
| Check for the notion of "compassionate conservatism" in other candidates. Maybe we should create a new code                                                                   | Bush - Announcement       | Others    | Admin | 8/5/2011  | You've heard me talk about compassionate what I mean.                                                                                                                                                               |

Selecting an item in this table, either by clicking it or by using the keyboard cursor keys such as the UP or DOWN arrow keys, automatically displays the corresponding document in the main window and highlight the corresponding coded segment. This feature provides a way to examine the retrieved segment in its surrounding context. By default, all types of comments are displayed. To display a single type of comment, set the **TYPE** list box to the desired comment type

## To edit or delete a comment:

- Select the row containing the comment you would like to edit.
- Click the  button and edit the text of the comment. To remove a comment, simply delete all the text.

## To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.

For more information on the Report Manager, see the **Report Manager Feature** topic on page 230.

### To export retrieved comments to disk:

- Click the  button. A Save File dialog box will appear.
- In the **Save as type** list box select the file format under which you would like to save the table. The following formats are supported: ASCII file (\*.TXT); tab delimited file (\*.TAB); comma delimited file (\*.CSV); HTML (\*.HTM; \*.HTML); XML (\*.XML); MS Word document (\*.DOC); Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).
- Type a valid filename with the proper file extension.
- Click the **Save** button.

### To print the list of comments:

- Click the  button.

# Link Retrieval

The LINK RETRIEVAL feature allows you to obtain a list of all hyperlinks in your project, or restrict this list to specific link types or to those meeting specific conditions. For example, one may retrieve links associated with specific codes, or to those containing some keywords or key phrases in their label or description. One may also filter links for events occurring in a specific time period or associated with a specific geographic region. Items with geographic coordinates may then be exported and viewed in Google Earth or ArcView software (with optional time animation) while time-tagged links may be displayed on an interactive timeline display.

To retrieve a list of hyperlinks, choose the LINK RETRIEVAL command from the RETRIEVAL menu. A dialog box similar to this one will appear:

The screenshot shows the 'Link Retrieval' dialog box. It has two tabs: 'Criterion' and 'Search Hits'. The 'Criterion' tab is active. Under 'Retrieve links in:', the 'Document:' dropdown is set to '[DOCUMENT]'. Below it, the 'Attached to:' radio buttons are set to 'Code marks'. The 'Link types:' section has 'Geographic location' checked. The 'Conditions:' section has 'Associated with codes:' checked and set to '[Criminal activities]', 'Geographic area within:' checked and set to '100 kilometers from latitude: 31.6115261932677 longitude: 65.7268961137171', and 'Date and time:' checked and set to 'After 01/03/20 12:00:00 AM'. The 'Add to Report' section has 'Coded segment' and 'Comment' unchecked, and 'Variables:' set to '[FILE]'. A 'Search' button is at the bottom right.

The **Document** option allows you to specify from which document variables to retrieve hyperlinks. If the current project contains more than one document variable, you will have a choice of selecting either one or a combination of them. By default, all document variables are selected. To restrict the analysis to only a few of them, click the Down arrow key at the right of the list box. You will be presented with a list of all available document variables. Select the variables on which you want the search to be performed. One may also retrieve links attached to **Text** segments only, **Code marks** only, or **Both**.

The **Link Type** option box lets you specify which types of links should be retrieved. One may choose any one of the six types available in QDA Miner or a combination thereof.

The **Condition** option box is used to filter links meeting additional conditions. Five types of filters may be applied:

- **Current Cases Only** will restrict the retrieval to links associated with documents or images in the current case.
- The **Associated with Codes** condition allows one to restrict the retrieval of links to those attached to specific codes. When this option is chosen, you will be required to click the down arrow key and select the codes from list. These codes can also be selected from a tree representation of the codebook by clicking the  button.
- One may also retrieve only links containing specific keywords or key phrases in their label or description by selecting the **Label or Description Includes** box and entering a search term. The **Case Sensitive** option defines whether the search is case sensitive, while the **Whole Word** box allows you to define whether the typed words or phrases can match part words or only whole words.
- Geographic tags make it possible to filter those tags based on their distance from a specific geographic location. To enable this feature, the **Geographic Area Within** box should be selected. One then needs to specify a geographic point using its latitude and longitude and the distance either in **miles** or in **kilometers** from this location within which the geographic tag should be located. You may click the  button to type the geographic coordinates in degrees, minutes, seconds and cardinal compass point rather than decimal values, or paste the location from Google Earth. You may also import a geographic location from a KMZ or KML file by clicking the  button, or search previously entered or imported ones by clicking the  button.
- It is possible to filter date and time as well as geographic links with time information by setting the **Date and Time** box. One then needs to select whether the associated dates should occur **Before**, **After**, or **Between** a specific date-and-time range. In the latter case, two dates need to be provided.

The **Add to Report** section allows you to append to the table of retrieved segments additional information such as the text from the original document associated with this link, any comments attached to the coded segment, or any values stored in one or more variable for the specific case from which the links originate.

Once the search conditions have been set, click the  button. The results of a search are displayed in a table located on the **Search Hits** page similar to this one:

| Link Retrieval |                     |                     |                                        |                                   |                  |                   |            |
|----------------|---------------------|---------------------|----------------------------------------|-----------------------------------|------------------|-------------------|------------|
| Criterion      |                     |                     |                                        | Search Hits                       |                  |                   |            |
| Case           | Code                | Link type           | Link/Label                             | Description                       | Latitude         | Longitude         | Date       |
| Case #2        | Academic Career     | Geographic Location | Sacramento, CA                         | First bombing                     | 38° 34' 53.66" N | 121° 29' 39.84" W |            |
| Case #2        | Criminal activities | Geographic Location | Feu à Northwestern University          | Set fire to a computer lab at NWU | 43° 4' 29.18" N  | 87° 41' 24.22" W  | 5/25/1978  |
| Case #2        | Criminal activities | Geographic Location | American Airline Flight 444            | 12 American Airlines passengers   | 41° 51' 0.11" N  | 87° 39' 0.18" W   | 11/15/1979 |
| Case #2        | Criminal activities | Geographic Location | Feu à Northwestern University          | Set fire to a computer lab at NWU | 43° 4' 29.18" N  | 87° 41' 24.22" W  | 5/25/1978  |
| Case #2        | Criminal activities | Geographic Location | Sacramento, California                 |                                   | 38° 34' 53.65" N | 121° 29' 39.84" W | 12/11/1985 |
| Case #2        | Criminal activities | Geographic Location | Salt Lake City                         | Set a bomb by mail                | 40° 45' 38.81" N | 111° 53' 27.78" W | 2/20/1987  |
| Case #2        | Criminal activities | Geographic Location | University of Yale                     |                                   | 41° 18' 44.96" N | 72° 55' 49.99" W  | 6/24/1993  |
| Case #2        | Criminal activities | Geographic Location | University of California - Tiburon, CA |                                   | 38° 36' 30.17" N | 91° 13' 8.54" W   | 6/22/1993  |
| Case #2        | Criminal activities | Geographic Location | North Caldwell, NJ                     |                                   | 40° 51' 53.36" N | 74° 15' 29.53" W  | 12/10/1994 |
| Case #2        | Relocation          | Geographic Location | Sacramento, CA                         | Last attempt                      | 38° 34' 53.66" N | 121° 29' 39.84" W | 4/24/1995  |
| Case #3        | Criminal activities | Geographic Location | Tiburon, CA                            |                                   | 37° 52' 24.74" N | 122° 27' 23.90" W | 4/28/1978  |

This table is synchronized with the main window, so moving from one row to another will move you to the case and document/image containing the selected link and will highlight the associated text or image area to which the link has been attached, allowing you to access its full context.

### To remove a link from the list of hits:

- Click the  button. Removing a link using this method does not delete the link itself but simply removes the associated row from the list of hits.

### To edit a link:

- Click the  button. For more information on editing options, see Hyperlinking.

### To delete a link:

- Click the  button. This operation will not only remove the row from the list of hits but will also delete the associated hyperlink.

### To create a timeline of links containing date-and-time information:

- Click the  button. For further information on this feature, see Creating and Editing Timelines.

### To export all geographic tags to a Google Earth KMZ file:

- Click the  button. For further information on this feature, see Displaying Geo-Tags in Google Earth.

### To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button. For more information on the Report Manager, see the **Report Manager Feature** topic on page 230.

### To export the table to disk:

- Click the  button. A Save File dialog box will appear.
- In the Save As Type list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file (\*.CSV), MS Word (\*.DOC), HTML file (\*.HTM; \*.HTML), XML files (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).
- Type a valid file name with the proper file extension.
- Click the SAVE button.

### To print the table:

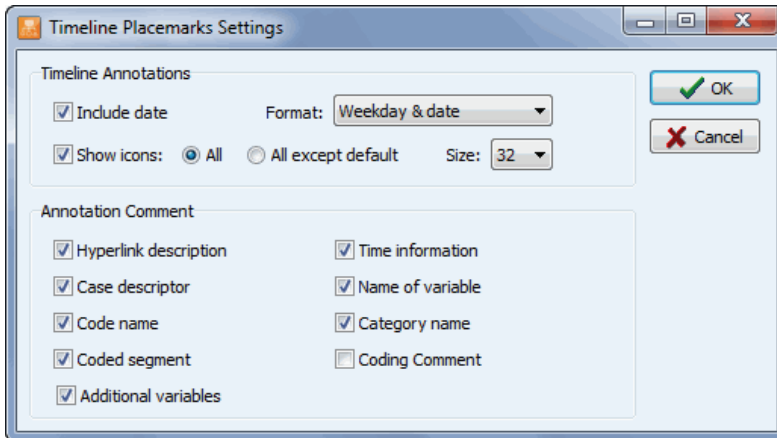
- Click the  button.

# Creating and editing timelines

QDA Miner allows one to associate date-and-time information to specific text segments or code marks. Since such events may not necessarily be presented chronologically in a project's various documents, the possibility of ordering them can represent a powerful method to identify sequences of related events. A timeline graphic fits this purpose perfectly since it is a visual representation of the sequence of events arranged in chronological order and displayed along a single line.

To create a timeline in QDA Miner, you first need to retrieve links containing time information, by running the LINK RETRIEVAL command from the RETRIEVAL menu (for further information on how to retrieve those links and on how to apply various filtering criteria, see [Link Retrieval](#), page 151).

Once links have been retrieved, click the  button located in the upper-right corner of the SEARCH HITS page. A dialog box similar to this one will appear:



A timeline in QDA Miner stores text and graphic information in two different locations: the **Timeline Annotations** and the **Annotation Comments**.

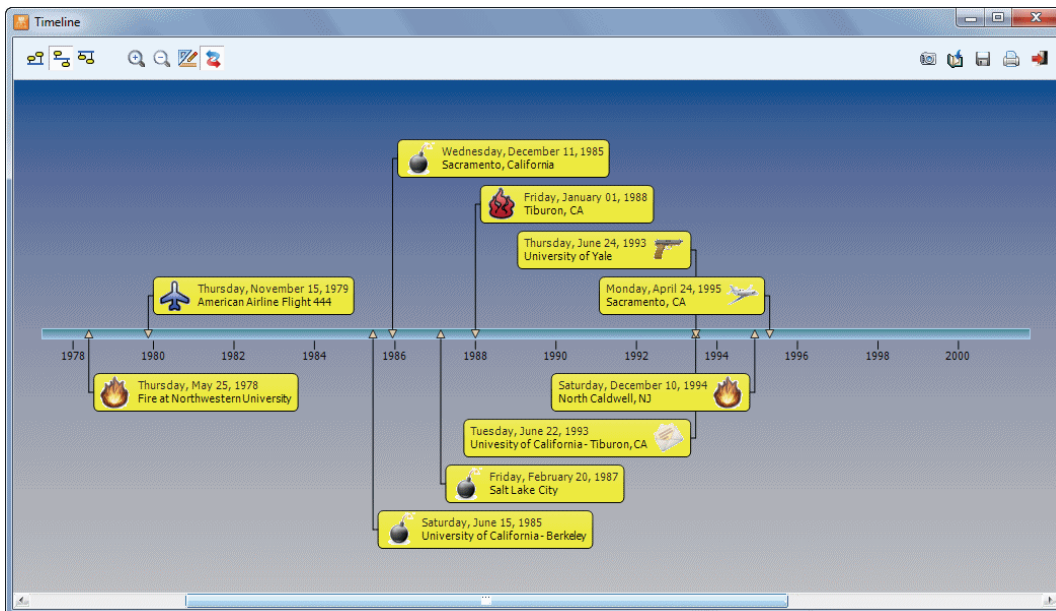
**Timeline Annotations** are small text boxes displayed above and below the timeline. They contain basic information about the event being plotted. By default, this box contains the label used to identify this time-tag or geo-tag. Selecting the **Include Date** option will add the stored date to this label. The **Format** for this date may be adjusted to include the weekday and to use either the short-date or long-date format settings of your Windows system. One may also append to this annotation the associated icon by selecting the **Show Icon** check box. By default, all icons will be shown, but one may instruct QDA Miner to display only icons that have been changed. The icon **Size** may also be adjusted by choosing a value between 8 and 64 pixels.

Each timeline annotation is also associated with a comment box containing more detailed information such as the link description, the associated code and category names, and the underlying text. This comment is displayed when the mouse cursor is moved over the timeline annotation.

The **Annotation Comment** section allows one to choose what information should be displayed in this comment box, such as the **Hyperlink Description**, the **Time Information**, the **Case Descriptor** or the **Name of the Variable** from which this time tag comes.

If the time tag has been attached to a code mark, one may also include the Code Name, its Category as well as any Comment attached to the code mark. If the time tag has been attached to a text segment, either directly or through a code mark, selecting Coded Segment will append this coded text segment to the annotation comment. Finally, if variables have been added to the link retrieval table, then selecting Additional Variables will retrieve the values of those variables for the containing case and append them to the comment.

Once the options have been set, clicking the **OK** button creates a timeline dialog box similar to this one:



The first three buttons on the top-left of the dialog box allow one to choose the location of the scale line and timeline annotations. Three options are available:



Click this button to display the scale line at the bottom of the screen. All timeline annotations will be displayed above this scale line.



Click this button to vertically centre the scale line and positions the timeline annotations above and below this scale line. The vertical position of those annotations is set automatically. One may, however, override such values by setting the vertical position of specific annotations manually (See below).



Click this button to display the scale line at the top of the screen and positions all timeline annotations below it.

By default, the time range of the timeline is spread so that all events can be displayed on the one screen. Click the  button to focus on a shorter time period. When zooming in a timeline, only a portion of the

entire time period is displayed on-screen. One can adjust the time period displayed by moving the scrollbar located at the bottom of the dialog box. Clicking the  button increases the time range, allowing one to obtain a broader view of the sequence of events.

Clicking down the  button forces QDA Miner to synchronize the graphic with the main window so that selecting one timeline annotation will cause the main window to display the corresponding case and document from which the selected time tag comes. The corresponding text or image in the original document will also be highlighted allowing one to easily view the full context of the selected event.

Clicking the  button brings a dialog box that allows one to customize the physical appearance of various elements of the time line such as the color of the background, the vertical line and the annotation, as well as the font size and style of the date labels and timeline annotation. One may also customize the display of any annotation, by selecting it, right clicking to display a popup menu and selecting the appropriate command. The available commands allow one to either **Remove** this even, change its background Color or force its **Position** relative to the timeline so that it will appear above or below it. Please note that this last option is available only when the scale line vertical position is set to be displayed in the middle of the dialog box.

The following table provides a short description of additional buttons:

| Control | Description |
|---------|-------------|
|---------|-------------|



Click this button to create a copy of the timeline to the clipboard. When this button is clicked, a pop-up menu appears, allowing one to select whether the timeline should be copied as a bitmap or as a metafile.



Click this button to append a copy of the timeline in the Report Manager. A descriptive title will be provided automatically. To edit this title or to enter a new one, hold down the Shift key while clicking this button



Click this button to save an image of the timeline on disk. Timelines may be saved in BMP, JPG or PNG graphic file formats.



Click this button to print a copy of the displayed timeline.

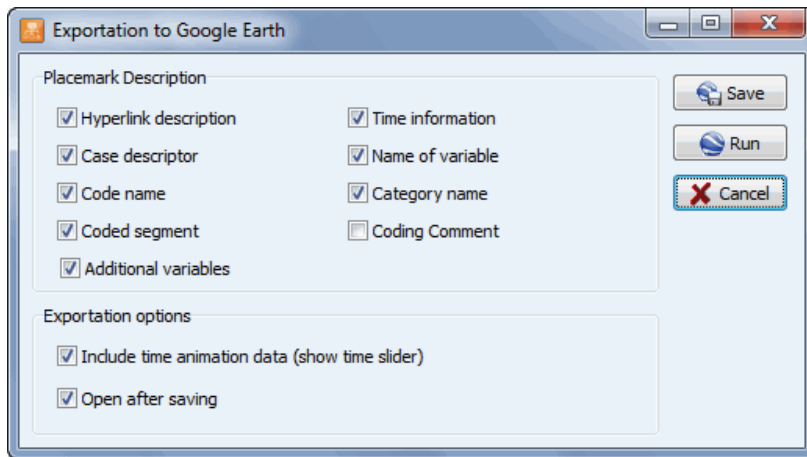


Click this button to close the timeline dialog box and return to the QDA Miner link-retrieval dialog box.

# Exporting and Viewing Geotags

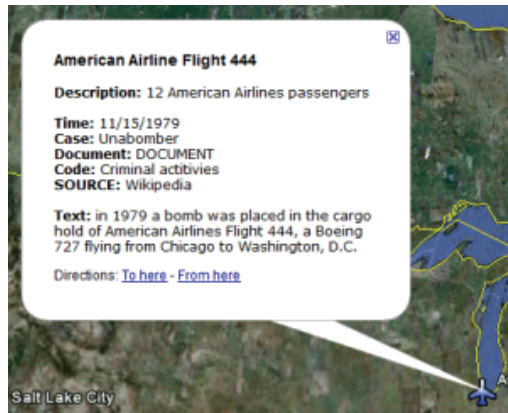
QDA Miner allows one to associate geographic coordinates with specific text segments or coded image areas. To display a specific coordinate, one simply needs to click the code mark or anywhere in the text segments while holding down the Control key. One may, however, retrieve several geographic coordinates and export those into a single KMZ or KML file and display those coordinates in Google Earth or in any other software supporting those file formats, such as ArcGIS Explorer. To create such a list of geographic locations in QDA Miner, you first need to retrieve links containing geographic coordinates by running the LINK RETRIEVAL command from the RETRIEVAL menu. Those coordinates may be filtered based on several criteria, such as their distance from a specific geographic location or their associated date in keywords stored in the label or description (for further information on how to retrieve those links and on how to apply various filtering conditions, see **Link Retrieval** on page 151).

Once links have been retrieved, click the  button located in the upper-right corner of the Search Hits pages. A dialog box similar to this one will appear:



Each placemark is associated with a description that is displayed when clicked. The **Placemark Description** section allows one to choose what information should be displayed in this comment box, such as **Hyperlink Description**, **Time Information**, **Case Descriptor** or **Name of the Variable** from which this time tag comes. If the geographic tag has been attached to a code mark, one may also include the **Code Name**, its containing **Category** and any **Comment** attached to the code mark. If the time tag has been attached to a text segment, either directly or through a code mark, or to a coded area of an image, selecting **Coded Segment** will append this coded text segment or image area to the placemark description window. Finally, if variables have been added to the link-retrieval table, then selecting **Additional Variables** will retrieve the values of those variables for the containing case and append those to the description.

The following image gives an examples of a placemark description containing not only the tag label and its description but also the associated time, the case and the variable from which this information comes, as well as the code name and the coded text segment to which this geo-tag has been attached.



The **Exportation Options** section offers two additional options. The **Include Time Animation Data** option inserts in the exported geographic placemarks their associated time stamp or time range (if any), causing Google Earth to display a time slide control like the one shown below. This slide control may be used to restrict the display of placemarks to those appearing within a specified time period or to create animations showing the change in the spatial distribution of events over time.



This slide control may be used to restrict the display of placemarks to those appearing within a specified time period or to create animations showing the change in the spatial distribution of events over time. features of the Google Earth time slider include:

1. Drag this range marker to the right or left to re-define the time range of data displayed.
2. Drag this to move the time range earlier or later.
3. Zoom in or out to shorten or lengthen the date range covered by your timeline.
4. Click this to play an animation of a sequence
5. Click this to set options for the time slider, such as the time range, the animation speed or the time format.

Selecting the **Open After Saving** option instructs QDA Miner to run Google Earth and to display the exported data.

Once the option has been set, one can either run Google Earth by clicking  or save all placemarks to a specific file by clicking the  button.

# Coding Analysis Features

QDA Miner provides several tools to assist in the coding task and to perform descriptive, comparative and exploratory analysis of codings. These tools may be used to identify regularities and patterns in coding, uncover hidden relationships between codes and other properties of the cases, etc..

The **Coding Frequency** tool allows one to obtain a list of all codes in the current codebook along with their description and various statistics such as their frequency, the number of cases in which they are found, and the total number of words in the associated text segments. This dialog box also allows one to produce bar charts and pie charts from those statistics.

The **Codes Co-occurrences** tool uses information about the proximity or the co-occurrence of codes within documents to explore potential relationships among them as well as similarities among cases. This dialog box gives access to various statistical and graphic tools, such as cluster analysis, multidimensional scaling, and proximity plots.

The **Coding Sequences** tool can be used to identify recurring sequences of codes. This feature can produce frequency lists of all sequences involving two selected sets of codes as well as the percentage of time one code follows or is followed by another one.

The **Coding by Variable** tool is useful for identifying or testing potential code similarities or differences between subgroup of cases (categorical variable) or to assess the relationship between these codes and other numerical variables. Various statistical and graphical tools are used for this purpose such as contingency tables, association statistics, bar charts and line charts, heatmaps, correspondence plot, etc.

The **Inter-Coders Agreement** tool is used to compare the consistency of coding between several coders. This tool can be useful to uncover differences in interpretation, clarify equivocal rules, identify ambiguity in the text, and ultimately quantify the final level of agreement obtained by those raters.

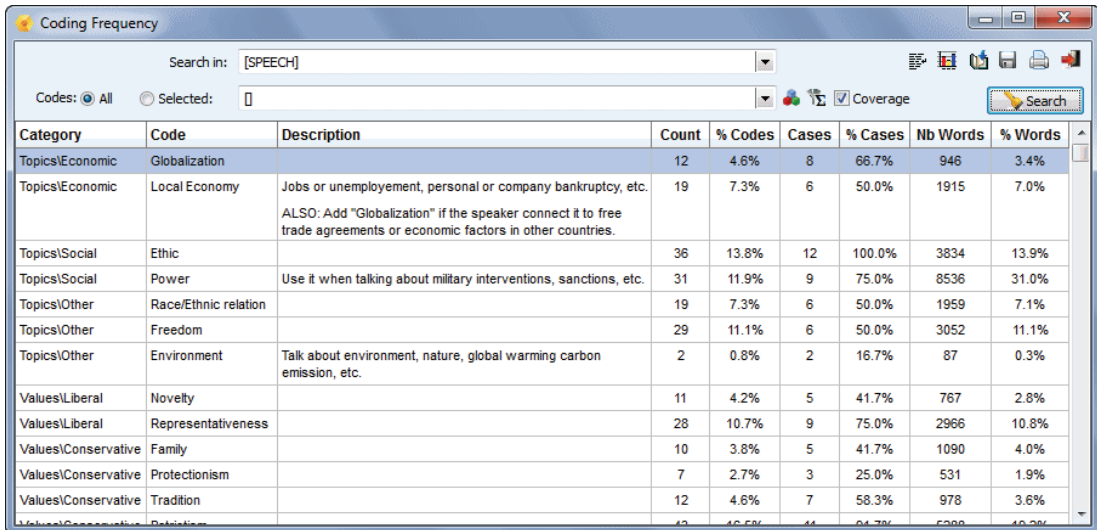
QDA Miner may also be used in conjunction with the following two quantitative analysis software tools:

**WordStat** is a powerful quantitative content analysis and text-mining software. When used as an add-on to QDA Miner, it analyzes words and phrases found in specific documents or in selected code segments. WordStat can perform simple descriptive analysis or explore in greater details the relationship between words or categories of words and other numeric or categorical variables.

**Simstat** is a comprehensive statistical analysis application. Since it uses the same file format as QDA Miner, Simstat can be used to perform statistical analysis on any numerical data stored in a QDA Miner project file. It can perform numeric and alphanumeric computation, transformation and recoding of variables, as well as advanced file management procedures, such as data file merging, file aggregation, etc.

# Coding Frequencies

You can use the CODING FREQUENCIES command from the ANALYSIS menu to obtain a list of all codes in the current codebook along with their description and the category to which they belong. This dialog box may also be used to obtain various statistics for each code such as their frequency, the number of cases in which they are found, and the total number of words in the associated text segments. Selecting this command displays a dialog box similar to the one below:



| Category            | Code                 | Description                                                                                                                                                                        | Count | % Codes | Cases | % Cases | Nb Words | % Words |
|---------------------|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|---------|-------|---------|----------|---------|
| Topics\Economic     | Globalization        |                                                                                                                                                                                    | 12    | 4.6%    | 8     | 66.7%   | 946      | 3.4%    |
| Topics\Economic     | Local Economy        | Jobs or unemployment, personal or company bankruptcy, etc.<br>ALSO: Add "Globalization" if the speaker connect it to free trade agreements or economic factors in other countries. | 19    | 7.3%    | 6     | 50.0%   | 1915     | 7.0%    |
| Topics\Social       | Ethic                |                                                                                                                                                                                    | 36    | 13.8%   | 12    | 100.0%  | 3834     | 13.9%   |
| Topics\Social       | Power                | Use it when talking about military interventions, sanctions, etc.                                                                                                                  | 31    | 11.9%   | 9     | 75.0%   | 8536     | 31.0%   |
| Topics\Other        | Race/Ethnic relation |                                                                                                                                                                                    | 19    | 7.3%    | 6     | 50.0%   | 1959     | 7.1%    |
| Topics\Other        | Freedom              |                                                                                                                                                                                    | 29    | 11.1%   | 6     | 50.0%   | 3052     | 11.1%   |
| Topics\Other        | Environment          | Talk about environment, nature, global warming carbon emission, etc.                                                                                                               | 2     | 0.8%    | 2     | 16.7%   | 87       | 0.3%    |
| Values\Liberal      | Novelty              |                                                                                                                                                                                    | 11    | 4.2%    | 5     | 41.7%   | 767      | 2.8%    |
| Values\Liberal      | Representativeness   |                                                                                                                                                                                    | 28    | 10.7%   | 9     | 75.0%   | 2966     | 10.8%   |
| Values\Conservative | Family               |                                                                                                                                                                                    | 10    | 3.8%    | 5     | 41.7%   | 1090     | 4.0%    |
| Values\Conservative | Protectionism        |                                                                                                                                                                                    | 7     | 2.7%    | 3     | 25.0%   | 531      | 1.9%    |
| Values\Conservative | Tradition            |                                                                                                                                                                                    | 12    | 4.6%    | 7     | 58.3%   | 978      | 3.6%    |
| Values\Conservative | Religion             |                                                                                                                                                                                    | 13    | 4.9%    | 11    | 84.6%   | 5908     | 21.8%   |

The **SEARCH IN** option allows you to specify which document variables to search. If the current project contains more than one document variable, you will have a choice of selecting either one or a combination of them. By default, all document variables are selected. To restrict the analysis to only a few of them, click the down arrow key at the right of the list box. You will be presented with a list of all available document variables. Select the variables on which you want the search to be performed.

The **CODES** option allows you to select which codes will be analyzed. By default, the analysis is performed on all codes in the codebook. You can select specific codes either from a drop-down checklist by clicking the arrow button or from a tree representation of the codebook by clicking the  button. You may also select codes based on a minimum frequency criteria by clicking the  button.

Clicking the  button adds four columns to the right side of the table containing summary statistics on code usage:

|                |                                                 |
|----------------|-------------------------------------------------|
| <b>Count</b>   | Number of times this code has been used.        |
| <b>Cases</b>   | Number of cases in which this code appears.     |
| <b>% Count</b> | Percentage of coding associated with this code. |
| <b>% Cases</b> | Percentage of cases containing this code.       |

Selecting the **COVERAGE** option displays two additional columns:

|                 |                                                                       |
|-----------------|-----------------------------------------------------------------------|
| <b>Nb Words</b> | Total number of words in all text segments associated with this code. |
| <b>% Words</b>  | Percentage of words tagged with this code.                            |

If graphics are also included in the search, the following columns will also become visible:

|                  |                                                                              |
|------------------|------------------------------------------------------------------------------|
| <b>Nb Pixels</b> | Total number of pixels (in K) in all graphic areas associated with this code |
| <b>% Image</b>   | Percentage of image areas tagged with this code.                             |

To spread long descriptions over more than one line, click down the  button.

When statistics are shown in the table, the  button allows one to produce bar charts or pie charts to visually display the distribution of specific codes. To produce such charts:

- Set the Sort By option to the order you want the values to be shown graphically.
- Select the rows you would like to plot (multiple but separate rows can be selected by clicking the desired rows while holding down the CTRL key).
- Click the  button.

For further information see the **Bar chart and Pie Chart** section below.

To sort this list of hits in ascending order on any column values, simply click a given column header. Clicking the same column header a second time sorts the rows in descending order.

Click the  button to append a copy of the table in the Report Manager. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.

The list of codes may be saved to disk in Excel, MS Word, plain ASCII, text delimited, HTML or XML format by clicking the  button.

You may also print this list by clicking the  button.

## Bar chart and Pie Chart

QDA Miner allows one to produce bar charts or pie charts to visually display the distribution of specific codes. To produce such charts:

- Run the CODING FREQUENCIES command from the ANALYSIS menu.
- Click the **Show Statistics** button.
- Sort the table to the desired graphic order of the values.
- Select the rows you would like to plot (multiple but separate rows can be selected by clicking while holding down the CTRL key)
- Click the  button.

Three types of charts may be used to depict the distribution of keywords or content categories:



The vertical bar chart is the default chart used to display absolute or relative frequencies of keywords or content categories.



The horizontal bar chart displays the same information as the vertical one but is especially useful when the number of keywords is high and their labels cannot be displayed entirely on the bottom axis.



The pie chart is useful to display the relative frequency of each code and compare individual values to other values and to the whole. Numerical values displayed in pie charts are always expressed in percentages of either the total frequency or case occurrences.



The tag cloud shows the relative frequency of each code. This graphic is useful for quickly perceiving the most prominent code.

The **Plot** option allows one to select the values that will be used as the scale for the length of bars in bar charts or as the percentage base for pie charts. For bar charts the options are:

|           |                                                 |
|-----------|-------------------------------------------------|
| FREQUENCY | Number of occurrences of the code               |
| NB CASES  | Number of cases where this code appears         |
| NB WORDS  | Total number of words associated with this code |
| % CODES   | Percentage of codes associated with this code   |
| % CASES   | Percentage of cases where this keyword appears  |
| % WORDS   | Percentage of words associated with this code   |

If images have been coded, the following options may be available:

|           |                                                          |
|-----------|----------------------------------------------------------|
| NB PIXELS | Number of pixels in kilobytes associated with this code. |
| % PIXELS  | Percentage of pixels associated with this code.          |

For pie charts, three options are available to specify how percentages will be computed:

|           |                                                 |
|-----------|-------------------------------------------------|
| FREQUENCY | Number of occurrences of the code               |
| NB CASES  | Number of cases where this code appears         |
| NB WORDS  | Total number of words associated with this code |

and if images have been coded:

|           |                                                         |
|-----------|---------------------------------------------------------|
| NB PIXELS | Number of pixels in kilobytes associated with this code |
|-----------|---------------------------------------------------------|

The **View Others** option displays an additional bar or slice representing all items in the frequency table that have not been selected.

The following table provides a short description of available buttons and controls:

| Controls                                                                            | Description                                                                                                                                                                                                                         |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | Press this button to retrieve a chart previously saved on disk.                                                                                                                                                                     |
|    | Press this button to append a copy of the graphic in the Report Manager. A descriptive title will be provided automatically. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button. |
|    | Press this button to save a chart on disk. Charts are saved in a proprietary format and may be edited and customized using the Chart Editor.                                                                                        |
|  | Pressing this button allows you to print a copy of the displayed chart.                                                                                                                                                             |
|  | Click this button to turn on/off the 3D perspective for the current chart.                                                                                                                                                          |
|  | This button allows you to edit various features of the chart such as the left and bottom axis , the chart and axis titles, the location of the legend, etc.                                                                         |
|  | This button is used to create a copy of the chart to the clipboard. When this button is clicked, a pop-up menu appears allowing you to select whether the chart should be copied as a bitmap or as a metafile.                      |
|  | Pressing this button closes the chart dialog box and returns to QDA Miner’s Code Frequency dialog box.                                                                                                                              |

## Customizing bar charts and pie charts

Clicking the  button on the chart dialog box gives access to a dialog box to customize the appearance of bar charts and line charts. The options available in this dialog box represent only a small portion of all settings available.

To further customize the chart, modify data points, value labels, or series order, click the  button located on the right side of the dialog box.

### LEFT OR BOTTOM AXIS

**Minimum / Maximum** - QDA Miner automatically adjusts the vertical axis scale to fit the range of values plotted against it. To manually set these values, type the desired minimum and maximum. **Increment** Increasing or decreasing this value affects the distance between numbers as well as tick marks. Horizontal grid lines are also affected by modification of this value.

**Horizontal Grid** This option turns horizontal grid lines on and off. Grid lines extend from each tick mark on an axis to the opposite side of the graph. To increase or decrease the number of grid lines or the distance between those lines, change the Increment value of the axis. A list box also allows a choice among five different line styles to draw those grid lines.

### LEGEND

**Location** - This option positions the legend. Legends may be placed at Top, Left, Right and Bottom side of the chart.

**From top** - When the legend is displayed on the left or the right side of the chart, this option specifies the legend's top position in percent of total chart height.

**From left** - When the legend is displayed on the top or the bottom chart, this option specifies the legend's top position in percent of total chart width.

### TITLES

Proper titles and axis labels are of utmost importance when describing the information displayed in a chart. By default, QDA Miner uses variable names and labels as well as other predefined settings to provide such descriptions.

The title page allows one to modify the top title, as well as the labels on the left, bottom and right axis. To edit the title, select the proper radio button. Enter several lines of text for each title by pressing the **ENTER** key at the end of a line before entering the next line.

The Font button on the right side of the edit box allows to change the font size or style of the related title.

## 3D VIEW

**Orthogonal** - Turning this option off disables the free elevation and rotation of the 3D chart.

**Zoom** - This option zooms the whole chart. Expressed as a percentage, increasing the value positively will bring the chart towards the viewer, increasing the overall chart size as the Zoom value increases.

**3D Percent** - The 3D Percent property indicates the size ratio between chart dimensions and chart depth by specifying a percent number from 1 to 100.

**Perspective** - Use this property with Orthogonal unchecked to modify the 3D perspective of the Chart. Larger values add more depth perspective.

**Bar shadow** - Enabling this option dark shades to the sides of 3D bars. Turning it off will color the sides of the bar the same as the front.

**Bar width** - This option determines the percent of total bar width used. Setting this value to 100 makes joined bars.

**Bar depth** - Use this property to limit the depth that each bar series uses. By default, bars will take up the part proportional to the number of bar series in the chart so that the back of a bar will join the front of the bar immediately behind it. To insert a gap between series of bars, decrease this value.

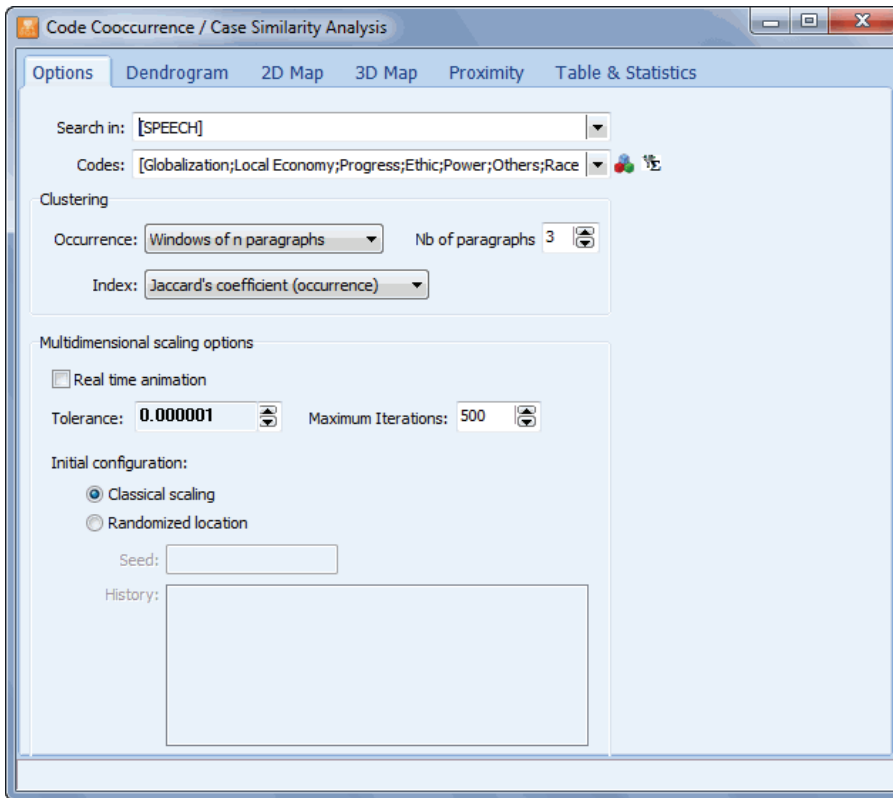
**Pie depth** - Use this property to change the thickness of the pie chart.

To further customize the chart, edit data points, value labels, etc., click the  button located at the right side of the dialog box.

# Coding Co-occurrences

QDA Miner allows you to further explore the relationship among codes by providing various graphic tools to assist the identification of related codes. These tools are obtained through the computation of similarity or co-occurrences index and the application of hierarchical cluster analysis and multidimensional scaling on all or selected codes. Results are displayed in the form of dendrograms, concept maps and proximity plots. Cases may also be clustered based on their content similarity using the same statistical and graphic tools.

To perform this type of analysis, select the CODING CO-OCCURRENCES command from the ANALYSIS menu. A dialog box similar to the one below will appear:



The first page of the dialog box is used to restrict the analysis to specific document variables or specific codes. It specifies whether clustering should be performed on codes or on cases and can be used to set various analysis and display options for both enters of analysis.

**SEARCH IN** - This option allows you to specify on which document variables the analysis will be performed. If the current project contains more than one document variable, you will have a

choice of selecting either one of them or a combination of them. By default, all document variables are selected. To restrict the analysis to a few of them, click the down arrow key at the right of the list box. You will be presented with a list of all available document variables. Select all variables on which you want the search to be performed.

**CODES** - This option allows you to select which codes will be analyzed. By default, the analysis is performed on all codes in the codebook. You can select specific codes either from a drop-down checklist by clicking the arrow button or from a tree representation of the codebook by clicking the  button.

You may also select codes based on a minimum frequency criteria by clicking the  button.

## Clustering Cases

When the clustering is set to be performed on cases, the distance matrix used for clustering and multidimensional scaling consists of cosine coefficients computed on the relative frequency of the various codes. The more similar two cases will be in term of the distribution of codes, the higher will be this coefficient.

**DATA** - This option allows you to select which indicator will be used to compare the importance of codes in the various cases. Four indicators are available:

- Code Occurrence
- Code Frequencies
- Number of Words
- Percentage of Words

Selecting **Code Occurrence** will result in a comparison based on which codes appear in each case, without taking into account the number of times each code appears. To take into account the number of times a code has been used, select **Code Frequencies**. Selecting **Number of Words** will base the comparison on the total number of words that have been assigned to different codes, while the **Percentage of Words** option will divide this value by the total number of words found in documents associated with each case.

## Clustering Codes

When clustering codes, several options are available to define co-occurrence and select which similarity index will be computed from the observed co-occurrences.

**CO-OCCURRENCE** - This option allows you to specify how a co-occurrence will be defined. By default, a co-occurrence is said to happen every time two codes appear in the same document (within document option). You may also restrict the definition of co-occurrence to codes that are

separated by no more than a specified number of paragraphs (window of n paragraphs option), or to codes that overlap each other (segment overlap option). Lastly, you can also restrict the definition of co-occurrence even further by limiting it to instances where codes have been assigned to the exact same segment.

**INDEX** - The Index option allows the selection of the similarity measure used in clustering and in multidimensional scaling. Four measures are available. The first three measures are based on the mere occurrences of specific codes in a case and do not take into account their frequency. In all these indexes, joint absences are excluded.

**Jaccard's coefficient** - This coefficient is calculated from a fourfold table as  $a/(a+b+c)$ , where  $a$  represents cases where both items occur, and  $b$  and  $c$  represent cases where one item is found but not the other. In this coefficient equal weight is given to matches and non matches.

**Sorensen's coefficient** - This coefficient is similar to Jaccard's but matches are weighted double. Its formula is  $2a/(2a+b+c)$ , where  $a$  represents cases where both items occur, and  $b$  and  $c$  represent cases where one item is present but the other one is absent.

**Ochiai's coefficient** - This index is the binary form of the cosine measure. Its formula is  $\text{SQRT}(a^2/((a+b)(a+c)))$ , where  $a$  represents cases where both items occur, and  $b$  and  $c$  represent cases where one item is present but not the other one.

The last coefficient takes into account not only the presence of a code in a case, but also how often it appears in this case.

**Cosine theta** - This coefficient measures the cosine of the angle between two vectors of values. It ranges from -1 to +1.

The options below apply whether clustering cases or codes. They will affect either the calculation or the display of multidimensional scaling charts.

**REAL-TIME ANIMATION** - When this option is enabled, the multidimensional plots are updated after every iteration allowing the user to monitor the progress made during the analysis at the cost of higher computing time.

**TOLERANCE** - This option specifies the tolerance factor that is used to determine when the algorithm has converged with a solution. Reducing the tolerance value may produce a slightly more accurate result, but will increase the number of iterations and the running time.

**MAXIMUM ITERATIONS** - This option allows you to specify the maximum number of iterations that are to be performed during the fitting procedure. If the solution does not converge with the limit specified by the TOLERANCE option before the maximum number of iterations is reached, the process is stopped and the results are displayed.

**INITIAL CONFIGURATION** - This option allows you to specify whether multidimensional scaling will be applied on a random configuration of points or on the result of classical scaling.

By selecting the **Classical Scaling** option, you perform a classical scaling first on the similarity matrix, and then use the derived configuration as initial values for the ordinal multidimensional scaling analysis.

Selecting the **Randomized Location** option performs multidimensional scaling analysis on a random configuration of points. By default, QDA Miner initializes the random routine before each analysis with a new random value. The seed value used for the creation of this initial configuration is stored along with the final stress value in the **history** listbox, located at the bottom of the dialog box. The **Seed** option may be used to specify a starting number that will initialize the randomization process and produce a fixed random sequence. To recall a specific seed value used previously, double-click the proper line in the **history** list box.

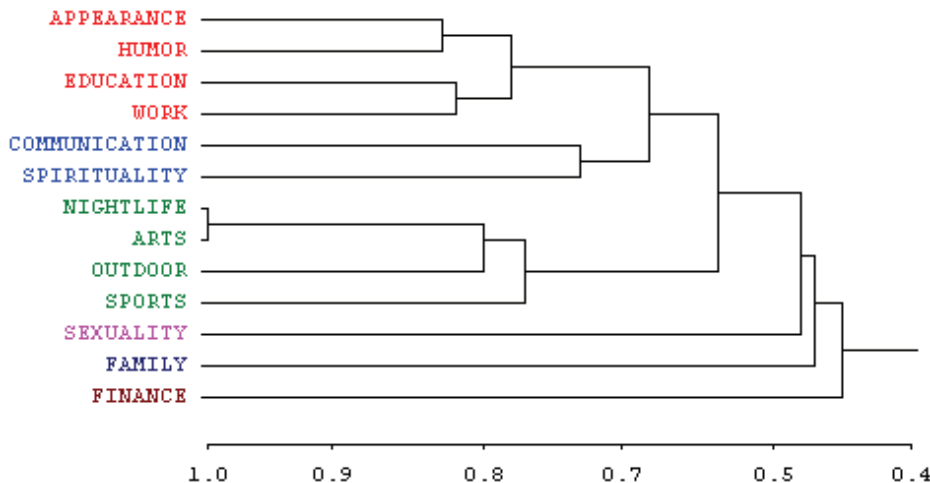
To store the analysis options, click the  button and specify the name under which those query options will be saved.

To retrieve a previously saved options, click the  button and select from the displayed list the name of the query you would like to retrieve (for more information, see **Saving and Retrieving Queries** on page 108).

Once the various options have been set, move to the appropriate page to perform the requested analysis.

# Dendrogram

QDA Miner uses an average-linkage hierarchical clustering method to create clusters from a similarity matrix. The result is presented in the form of a dendrogram (see below), also known as a tree graph. In this type of graph, the vertical axis is made up of the items, and the horizontal axis represents the clusters formed at each step of the clustering procedure. Codes that tend to appear together are combined at an early stage, while those that are independent from one another or those that do not appear together tend to be combined at the end of the agglomeration process.



It is important to remember that a dendrogram only specifies the temporal order of the branching sequence. Any cluster can be rotated around each internal branch on the tree without in any way affecting the accuracy of the dendrogram. The best analogy here is to think of a Calder mobile. Different photos for this type of mobile will yield different distances between hanging objects.

The following controls are found at the top of the dendrogram page:

**Nb Clusters** This option sets how many clusters the clustering solution should have. Different colors are used both in the dendrogram and in the 2D and 3D maps to indicate membership of specific items to different clusters.

**Display** This option lets you choose whether the vertical lines of the dendrogram represents the agglomeration schedule or the similarity indices.



When clustering codes co-occurrences, pressing this button displays bars beside each dendrogram item to represent their relative frequencies.



Use this button to increase the dendrogram font size and focus on a smaller portion of the tree



Use this button to reduce the dendrogram font size and view a larger portion of the tree.



To retrieve text segments associated with a specific cluster, click anywhere on a cluster to select it (the selected cluster is displayed with thicker black lines), and then click on this button to retrieve the associated documents. This operation retrieves all pairs of coded segments associated with the selected cluster.



Press this button to append a copy of the graphic in the Report Manager. A descriptive title will be provided automatically. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button



This button allows storing the displayed dendrogram into a graphic file. QDA Miner supports three different file formats: .BMP (Windows bitmap files), .PNG (Portable Network Graphic compress files) and .JPG (Jpeg compressed files).

Note: Clustering using other similarity or distance measures or agglomeration methods may be achieved using Simstat and the MVSP multivariate analysis add-on module.

## 2D and 3D Concept Maps

The concept maps are graphic representations of the proximity values calculated on all included items (codes or cases) using multidimensional scaling (MDS). When clustering codes, each point represents a code and the distances between pairs of points indicate how likely these codes tend to appear together. In other words, codes that appear close together on the plot usually tend to occur together, while codes that are independent from one other or that do not appear together are located on the chart far from each other. The interpretation of multidimensional scaling plot is somewhat different when analyzing cases. In this case, each point represents a case and the distance between pairs of points indicate how similar the two cases are. Cases with similar patterns of codes will tend to appear close each other, while dissimilar cases will be plotted far from each other. Colors are used to represent membership of specific items to different partitions created using hierarchical clustering. The resulting maps are useful to detect meaningful underlying dimensions that may explain observed similarities between items.

Please note that since multidimensional scaling attempts to represent the various points into a two- or three-dimensional space, some distortion may result, especially when this analysis is performed on a large number of items. As a consequence, some items that tend to appear together or that are parts of the same cluster may still be plotted far from each other. As well, performing a multidimensional scaling on a large number of items usually produces a cluttered map that is hard to interpret. For these reasons, interpretation of concept maps may be feasible only when applied to a relatively limited number of items.

## 2D and 3D Map controls

**NO. CLUSTERS** - This option is used to set how many clusters the clustering solution should have. Different colors are used both in the dendrogram and in the 2D and 3D maps to indicate membership of specific items to different clusters.



The actual orientation of axes in the final solution is arbitrary. The map may be rotated in any way desired; the distances between items remain the same. The rotating knob can be used to adjust the final orientation of axes in the plane or space in order to obtain an orientation that can be most easily interpreted.



Clicking this button enables you to zoom in on a plot. To zoom in on a area of the plot, hold the left mouse button down and drag the mouse down and to the right. A rectangle indicates the selected area. Release the left mouse button to zoom in.



Clicking this button restores the original viewing area of the plot.



Clicking this button performs another multidimensional scaling on a new random configuration of points. This button is visible only when the initial configuration is set to **Random Location**.



Press down this button creates a constrained multidimensional scaling. This clustering algorithm now allows one to preserve the clustering structure in multidimensional scaling plots, making the interpretation of 2D and 3D MDS maps a lot easier and more consistent with the clustering solutions. Enabling this option allows one to use the MDS module to create maps of concepts similar to those suggested by Trochim, in its Concept Mapping procedure.



Pressing down this button displays lines to represent relationships between data points of the multidimensional scaling plot. When the button is down, a cursor will appear in a tool panel below the plot, allowing you to select the minimum association strengths to be displayed.



This button is used to create a copy of the chart to the clipboard. When this button is clicked, a shortcut menu appears in which you can select whether the chart should be copied as a bitmap or as a metafile.



Press this button to append a copy of the graphic in the Report Manager. A descriptive title will be provided automatically. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.



This button allows editing of various features of multidimensional scaling plots such as the appearance of value labels and data points, chart and axes titles, the location of the legend, etc. (see Multidimensional Scaling Plot Options below)



This button allows storing the displayed multidimensional scaling plot in a graphic file. QDA Miner supports four different file formats: .BMP (Windows bitmap files), .PNG (Portable Network Graphic compress files) and .JPG (Jpeg compressed files) as well as .WSX a proprietary file format (QDA Miner Chart file). Charts stored in the latter format may be opened, further edited and customized using the external Chart Editor utility program.



Clicking this button prints a copy of the displayed chart.



Clicking this button closes the **Dendrogram & Concept Map** dialog box and returns to QDA Miner's main window.

## 3D Map buttons



This button shows or hides left, bottom and back walls.



Clicking this button draws anchor lines from the floor to the data point to better locate data points in all 3 dimensions.



Clicking this button switches the data of the X, Y and Z axes.



Clicking this button allows changing the viewing angle of the chart. To rotate the chart, make sure this button is selected, click any area of the chart, hold the mouse button down and drag the mouse to apply the desired rotation.



Clicking this button draws anchor lines from the floor to the data point to better locate data points in all 3 dimensions.



Locating a data point on the depth dimension of a 3D plot can be very difficult, especially when the plot remains static. One often has to rotate this plot constantly on the various axes to obtain an accurate idea of where the data point is located on this third axis. Clicking this button forces QDA Miner to rotate the plot automatically. To disable automatic rotation, click a second time.



This button is used to create an AVI movie file of the rotating chart. It behaves as both a record and a stop button. To start recording, simply click this button. Click again to stop the recording and to save the AVI movie file to disk.

## Multidimensional Scaling Plot Options Dialog Box

The various options in this dialog box are used to customize the appearance of multidimensional scaling plots. These options represent only a small portion of all settings available.

To further customize the chart, edit data points, value labels, etc., click the  button located at the right side of the dialog box.

### DATA POINTS

**VIEW DATA POINTS** - By default, multidimensional scaling plots display both the data points and their associated labels. This option toggles on and off the display of the data points.

**STYLE** - This option defines the shape used to display the data points. Nine different pointer shapes may be used to represent data point (e.g. square, triangles, dots, cross, etc.).

**SHADOW** - This option toggles on and off the shade on the sides of data points. This option only affects square, triangle and down triangle data points when viewed in three-dimension.

**SIZE** - This option specifies the width and height of data points in pixels.

## DATA LABELS

**TRANSPARENT LABELS** - This option allows you to specify whether data labels are to be displayed on an opaque background or whether the background should be invisible.

**BACKGROUND** - This option allows you to select the background color of data labels when not transparent

**BORDER** - This option enables or disables the rectangle surrounding the label background.

**VERTICAL GAP** - The vertical gap property determines the vertical distance in pixels between the top of the data point and the bottom of its label.

**FONT** - Click this button to adjust the font properties used to display data labels (i.e., style, size, and color).

## WALLS AND FRAME PAGE

**WALL VISIBLE** - Use this option to toggle the display of left, back and bottom walls to simulate a 3D effect.

**TRANSPARENT** - The Transparent property controls whether the wall backgrounds will be opaque or transparent.

**DARK 3D** - This option shades the sides of the walls.

## TITLE

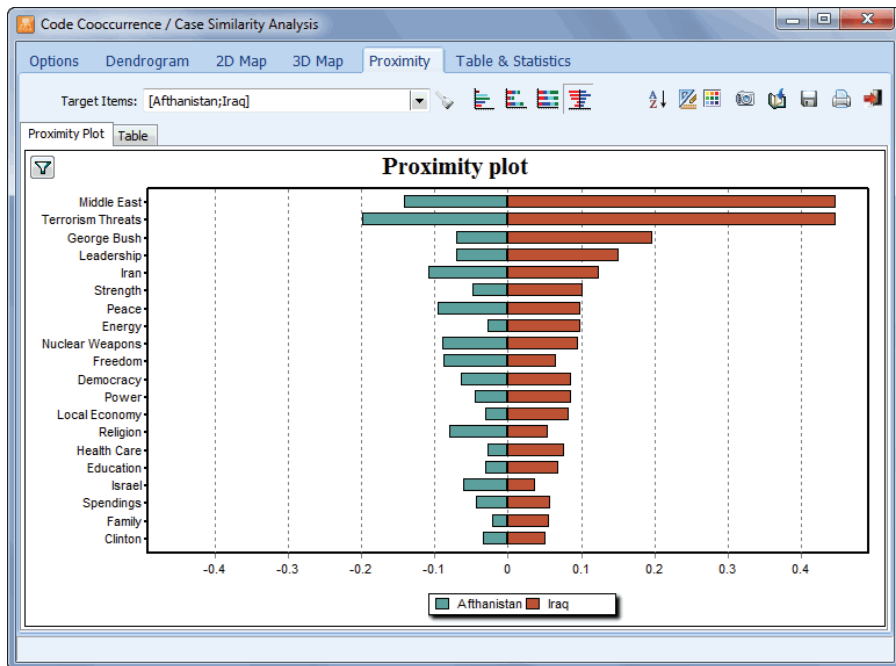
**SHOW TITLE** - By default, multidimensional scaling plots have no title. To display a title, enable the Show Title option and enter the desired title in the edit box to the right of this check box. Enter several lines of text by pressing the **<Enter>** key at the end of a line before entering the next line.

The **FONT** button at the bottom of this page is used to change the font size or style of this title.

## Proximity Plot

Cluster analysis and multidimensional scaling are both data reduction techniques and may not accurately represent the true proximity of codes or cases to each other. In a dendrogram, while codes that co-occur or cases that are similar tend to appear near each other, one cannot really look at the sequence of codes as a linear representation of these distances. You have to remember that a dendrogram only specifies the temporal order of the branching sequence. Consequently, any cluster can be rotated around each internal branch of the tree without in any way affecting the accuracy of the dendrogram. The best analogy here is to think of a Calder mobile. Different photos for this type of mobile will yield different distances between hanging objects. While multidimensional scaling is a more accurate representation of the distance between objects, the fact that it attempts to represent the various points in a two- or three-dimensional space may result in some distortion. As a consequence, some items that tend to appear together or to be very similar may still be plotted far from each other.

The proximity plot is the most accurate way to graphically represent the distance between objects by displaying the measured distance from a selected object to all other objects on a single axis. It is not a data reduction technique, but a visualization tool to help you extract information from the huge amount of data stored in the distance matrix at the origin of the dendrogram and the multidimensional scaling plots. In this plot, all measured distances are represented by the distance from the left edge of the plot. The closer an object is to the selected one, the closer it will be to the left.



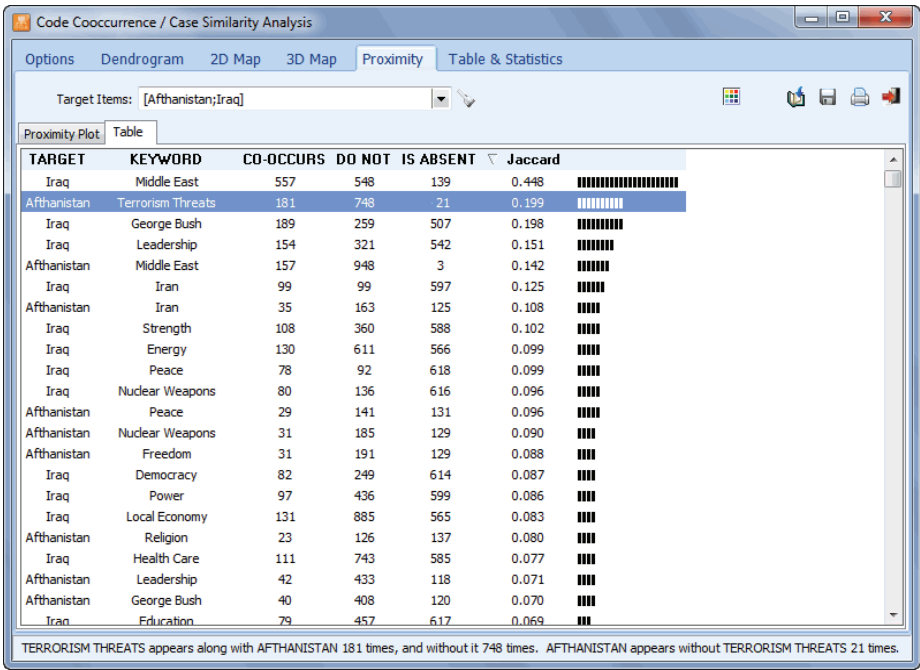
To select a code or a case that will be used as the point of reference, you can choose from the Target Items drop-down checklist located at the top of the page. You can also freely browse through different codes or cases by double-clicking the bar in the Proximity Plot. The co-occurrence or similarity to more than one target item may be displayed in a single chart allowing easy comparisons. When several target

items are selected, the proximity plot may consist of bars clustered side by side (clustered bars) or stacked (stacked bars), representing either the total amount or the relative distribution of scores (100-percent stacked bars). When two target items are selected, it is also possible to display the bars on both sides of a central axis, like the sample chart shown below (mirrored bars).

By default, the chart displays the proximity of the target items to a maximum of 30 related items. Clicking the  button located in the upper left-hand corner of the chart displays a dialog box that allows one to either manually choose items to be plotted or to automatically select a specific number of items.

When looking at code co-occurrences, selecting a bar enables the  button. Clicking this button retrieves every pair of coded segments co-occurring, allowing one to further explore the factors that may explain this co-occurrence. Right-clicking any existing bar displays a menu that allows one to remove the selected item, to move it to the list of target items by adding it to the existing bars, or to replace one of them. One may also retrieve documents or text segments using this pop-up menu.

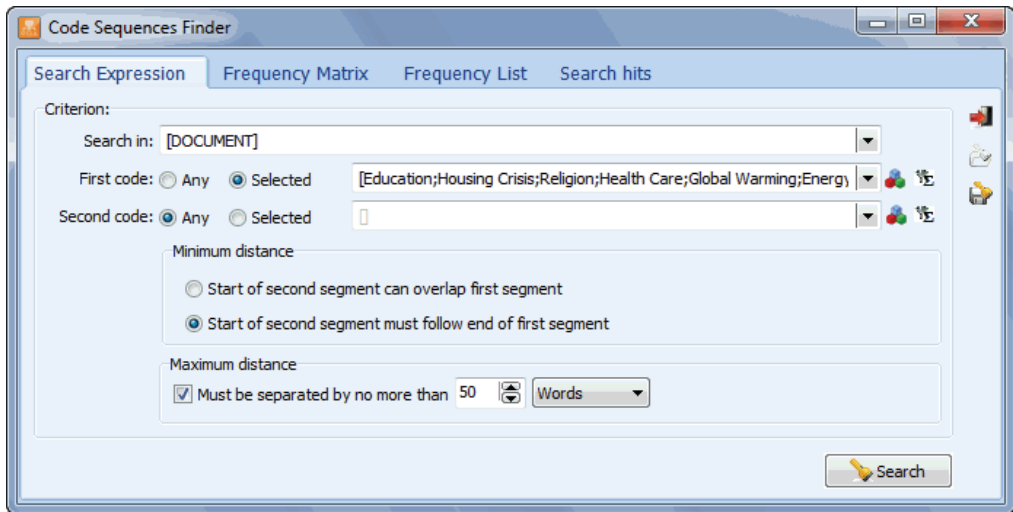
The table view (see below) allows you to examine in more detail the numerical values behind the calculation of these plots. When the distance measure is based on co-occurrences, the table provides detailed information such as the number of times a given code co-occurs with the selected one (CO-OCCURS) and the number of times that it appears in the absence of this selected code (DO NOT). Such a table also includes the number of times the selected code appears in the absence of the given code (IS ABSENT). In the example below, we can see on the first line of the table that the INSTRUMENTAL BEHAVIOR code co-occurs 3,727 times with SOCIAL BEHAVIOR, but on 3,376 occasions this code is encountered without the presence of SOCIAL BEHAVIOR, while this last code is found to occur 1,182 times in the absence of INSTRUMENTAL BEHAVIOR.



# Code Sequence Analysis

While coding co-occurrences analysis looks at the concomitant presence of codes regardless of their order of appearance in the documents, the Code Sequences finder is used to identify codes that not only co-occur, but do so in a specific order and under specific conditions. This command list the frequency of all sequences involving two selected sets of codes as well as the percentage of time one code follows or is followed by another one.

To search for code sequences, select the CODING SEQUENCES analysis from the ANALYSIS menu. A dialog box similar to the one below will appear:



The first page of the dialog box is used to restrict the analysis to specific document variables or specific codes and to set the conditions that must be met to determine whether a code follows another one.

**SEARCH IN** - This option allows you to specify the document variable to search. If the current project contains more than one document variable, you will have a choice of selecting either one or more of them. By default, all document variables are selected. To restrict the analysis to a few of them, click the down arrow key at the right of the list box. A list of all available document variables will appear. Select all variables on which you want the search to be performed.

**FIRST CODES** and **SECOND CODES** - These two options allow you to select which specific code sequences will be analyzed. By default, the analysis is performed on all codes in the codebook, providing you with a comprehensive list of all sequences found in the documents. The list of sequences can however be limited by restricting either the leading or the following codes. For example, if someone is interested in the reactions of a therapist to client verbalizations, he may set the **First Codes** option to include only the codes used to describe client verbalizations and limit the codes in the **Second Codes** option to those used to describe the therapist's reaction. You can select specific codes for these two options either from the corresponding drop-down checklist by clicking the arrow button or from a tree representation of the codebook by clicking the  button. You may also select codes based on a minimum frequency criteria by clicking the  button.

**MINIMUM DISTANCE** - This option is used to specify whether two codes will be considered to form a sequence when they partially overlap. If the first option is selected, a code can be considered as following the first code despite the fact that the text segment to which it was assigned starts before the end of the segment associated with the first code. However, if the second radio button is selected, QDA Miner will not consider these overlapping codes to be code sequences. For example, in the following example:



If the option is set to **Start of second segment can overlap the first segment**, then QDA Miner will treat the *Local Economy* and *Free Trade* codes as a valid sequence. If the **Minimum Distance** option is set so that the start of the second code must follow the end of the first one, then the same pair of codes will not be considered as a sequence and will not be included in the results table.

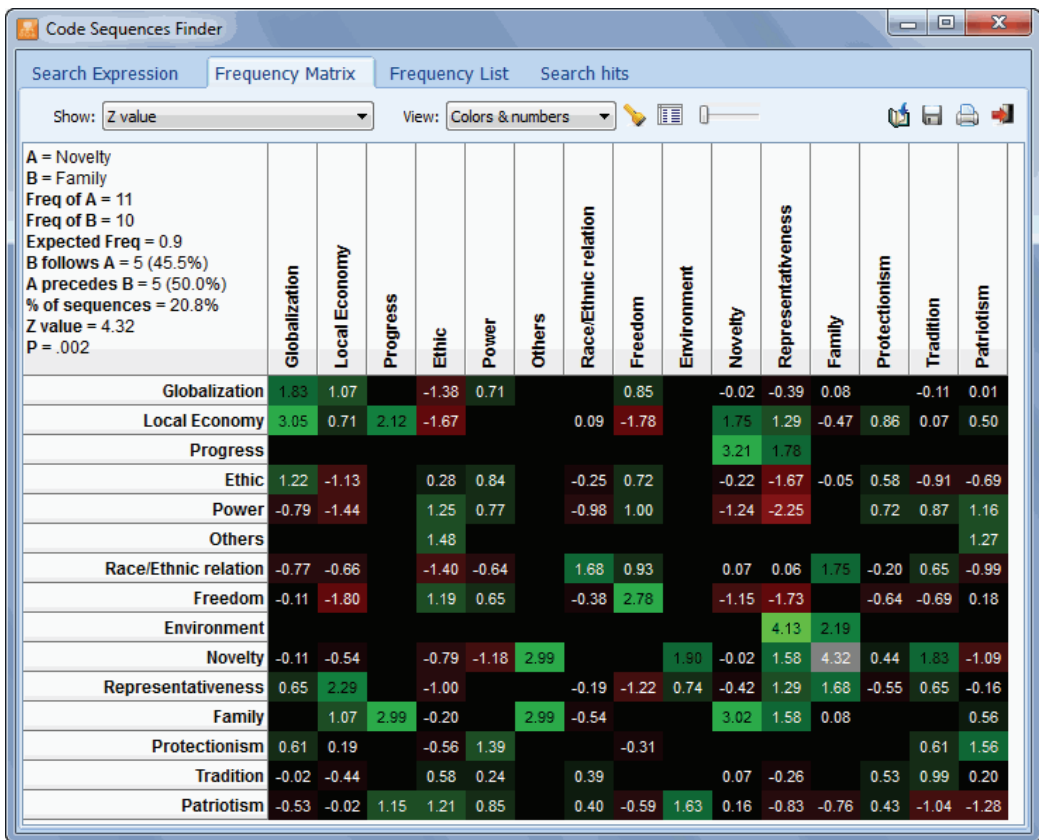
**MAXIMUM DISTANCE** - This option is used to specify what should be the maximum distance separating two codes in order to consider them to form a sequence. If disabled, QDA Miner will consider a code to follow another one as long as it is in the same document and meets the minimum distance criteria (see above). When enabled, code sequences are restricted to codes separated by no more than a predefined distance. This distance is calculated from the end of the first coded segment to the beginning of the second coded segment and can be measured either in number of characters, words, or paragraphs, or alternatively by setting a maximum number of codes that can occur between them. If the minimum distance is set to zero code, then only the codes following immediately the first code will be considered to be a sequence.

To store the search expression and options, click the  button and specify the name under which those query options will be saved.

To retrieve a previously saved query, click the  button and select from the displayed list the name of the query you would like to retrieve (for more information, see **Saving and Retrieving Queries** on page 108).

Once the criteria have been set properly, click the  button to find all sequences meeting those criteria. If at least one sequence is found, the **Frequency Matrix** and **Search Hits** page become enabled.

The **Frequency Matrix** page shows a matrix of all code sequences found in the document. The starting code (code A) in the A-B sequence is shown on the left side, while the code that follows (code B) is displayed at the top.



By default, the matrix displays the Z values of all sequences. In this example, colors are used to represent deviation from the expected frequency. Cells in green represent sequences occurring more often than expected, while cells in red represent sequences occurring less often. The brighter the color is, the farther it is from the expected frequency and the less likely such a sequence can be attributed to chance.

The **Show** list box allows you to display other statistics. The following sequence statistics are available:

- Frequency of A followed by B
- Frequency of B followed by A
- Percentage of A followed by B
- Percentage of B followed by A
- Percentage of sequences equal to AB
- Z-value
- Probability

The  control on the toolbar allows you to adjust the brightness of the color pattern, and thus increase or decrease the contrast between specific values.

The **View** list box may also be used to remove the color pattern and display only numerical values, or hide numerical values and display only color patterns.

The top left cell of the matrix displays for the currently selected cell (displayed with a gray background) the following information:

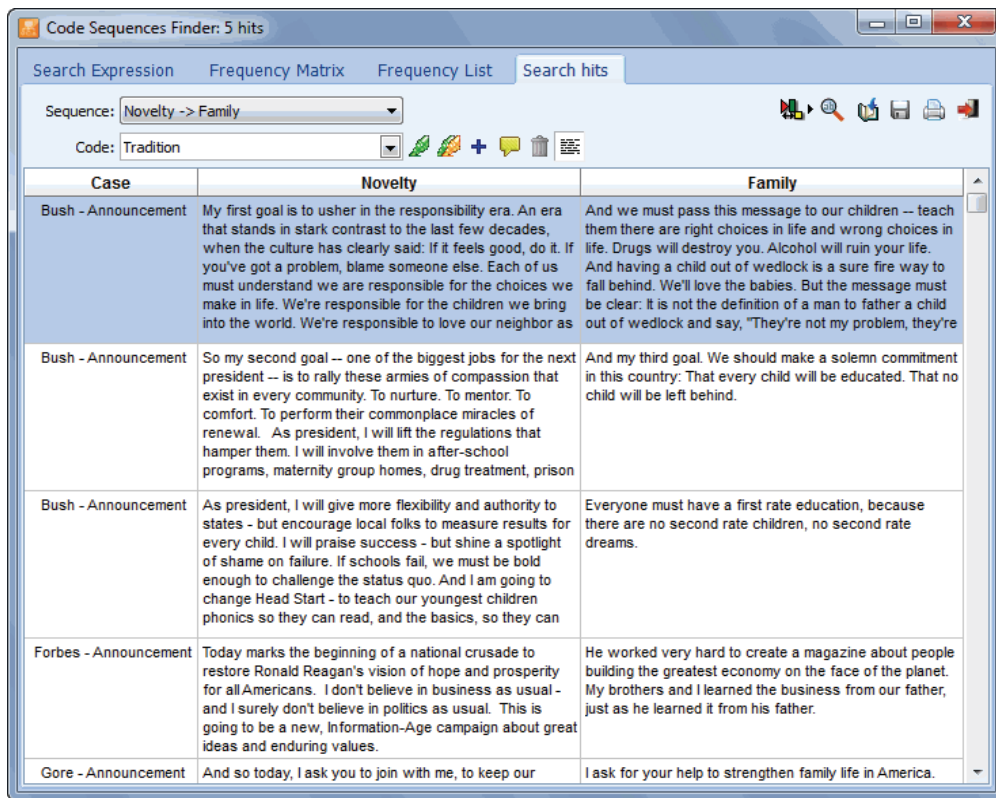
The table includes the following information:

| Column         | Interpretation                                                                          |
|----------------|-----------------------------------------------------------------------------------------|
| A=             | The name of the first code in the sequence                                              |
| B=             | The name of the second code in the sequence                                             |
| Freq of A      | The number of times that the first code is followed by this second code.                |
| Freq of B      | The percentage of time that the first code is followed by this second code.             |
| B follow A     | The number and percentage of times that the second code follows this first code.        |
| A precedes B   | The number and percentage of times that the first codes appears before the second code. |
| % of sequences | The percentage of all sequences equal to the current sequence.                          |
| Z values       | A standardized measure of the deviation from the expected frequency.                    |
| P              | The probability that this deviation occurs by chance.                                   |

To obtain a table containing those statistics for all observed sequences, click the  button. A new dialog box will appear with a large table of statistics that may be exported, printed or saved to disk. Clicking any column header sorts the rows in ascending order of values found in this column. Clicking the same column header a second time sorts the table in descending order.

 button. Double-clicking this cell will also moves you to the **Search Hit** page and display the associated text segments.

The **Search Hits** page allows you to examine in more detail the text segments associated with specific code sequences.



To select the code sequence to display, select its name from the **Sequence** list box. Selecting an item in this table, either by clicking its row or by using keyboard cursor keys such as the **UP** or **DOWN** arrow keys automatically displays the corresponding case and document in the main window. This feature provides a way to examine the retrieved segments in their surrounding context.

### To remove a search hit from the hit list:

- Select its row and then click the  button.

### To assign a code to a specific search hit:

- In the table of search hits, select the row corresponding to the text segment you want to code.
- Use the **CODE** drop-down list located above this table to select the code to assign.
- Click the  button to assign the selected code to the highlighted text segment.

## To assign a code to all search hits:

- Use the **CODE** drop-down list located above this table to select the code to assign.
- Click the  button to assign the selected code to all text segments matching the search expression.

The **Frequency Matrix** or the **Search Hits** list may be saved to disk in Excel, plain ASCII, text delimited, MS Word, HTML or XML format by clicking the  button.

They may also be appended in the Report Manager by clicking the  button. To edit the title used to describe the table, hold down the SHIFT keyboard key while clicking this button.

To append only a single hit from the Search Hits table, select the row containing the text segment you want to append to the Report Manager and click the  button.

To print any one of the tables, click the  button

# Coding by Variables

The CODE FREQUENCY dialog box is used to explore the relationship between codes assigned to documents and subgroups of cases defined by values of a numeric or categorical variable. When performing these kind of analyses, QDA Miner displays a contingency table containing either codes frequencies, code occurrences (present or absent), or either absolute or relative importance (in number or percentage of words). This dialog box also gives you access to several graphical and statistical tools to help you assess or visualize the strength of the relationship between code usage and the selected variable.

**Coding by Variable - CANDIDATE**

Search in: [SPEECH]

Codes: ☒ All ☐ Selected:

Tabulate with: CANDIDATE

Search

Table content

Count: Code frequency Display: Column percent

Statistic: Student's F

|                      | Bradley | Buchanan | Bush  | Forbes | Gore  | McCain | Student's F | P value |
|----------------------|---------|----------|-------|--------|-------|--------|-------------|---------|
| Globalization        | 2.4%    | 5.1%     | 4.3%  | 10.5%  | 5.8%  | 2.8%   | 0.400       | .833    |
| Local Economy        | 9.5%    | 10.3%    | 5.7%  | 0.0%   | 9.6%  | 2.8%   | 0.387       | .842    |
| Progress             | 2.4%    | 0.0%     | 0.0%  | 0.0%   | 0.0%  | 0.0%   | 1.000       | .489    |
| Ethic                | 7.1%    | 12.8%    | 20.0% | 15.8%  | 11.5% | 13.9%  | 1.800       | .247    |
| Power                | 0.0%    | 12.8%    | 15.7% | 15.8%  | 11.5% | 16.7%  | 0.674       | .659    |
| Race/Ethnic relation | 11.9%   | 5.1%     | 1.4%  | 0.0%   | 17.3% | 2.8%   | 0.870       | .551    |
| Freedom              | 0.0%    | 0.0%     | 11.4% | 10.5%  | 19.2% | 25.0%  | 1.079       | .456    |
| Environment          | 2.4%    | 0.0%     | 0.0%  | 0.0%   | 1.9%  | 0.0%   | 0.800       | .588    |
| Novelty              | 2.4%    | 0.0%     | 8.6%  | 10.5%  | 3.8%  | 0.0%   | 1.192       | .412    |
| Representativeness   | 28.6%   | 7.7%     | 5.7%  | 21.1%  | 7.7%  | 2.8%   | 1.157       | .425    |
| Family               | 9.5%    | 0.0%     | 5.7%  | 5.3%   | 1.9%  | 0.0%   | 0.945       | .514    |
| Protectionism        | 0.0%    | 10.3%    | 2.9%  | 0.0%   | 0.0%  | 2.8%   | 0.733       | .625    |
| Tradition            | 7.1%    | 12.8%    | 2.9%  | 5.3%   | 0.0%  | 2.8%   | 1.600       | .290    |
| Patriotism           | 16.7%   | 23.1%    | 15.7% | 5.3%   | 9.6%  | 27.8%  | 6.354       | .022    |

**SEARCH IN** - This option allows you to specify on which document variables the analysis will be performed. If the current project contains more than one document variable, you will have a choice of selecting either one of them or a combination of them. By default, all document variables are selected. To restrict the analysis to a few of them, click the down arrow key at the right of the list box. A list of all available document variables will appear. Select all variables on which you want the search to be performed.

**CODES** - This option allows you to select which codes will be analyzed. By default, the analysis is performed on all codes in the codebook. You can select specific codes either from a drop-down checklist by clicking the arrow button or from a tree representation of the codebook by clicking the  button. You can also combine values of two variables by selecting the <combine...> list

box option. When this option is selected, you will be presented with a dialog box that will let you choose the two variables you wish to combine. You may also select codes based on a minimum frequency criteria by clicking the  button.

**TABULATE WITH** - This option list all numeric and categorical variables available in the project. Selecting any variable name will display a contingency table so that you can assess the relationship between this variable and the codes assigned to text segments.

Once these options are set properly or when they are modified, click the  button to perform the analysis.

To display column labels vertically, click the  button located on the right of the Tabulate With list box.

To store the search expression and options, click the  button and specify the name under which those query options will be saved.

To retrieve a previously saved query, click the  button and select from the displayed list the name of the query you would like to retrieve (for more information, see **Saving and Retrieving Queries** on page 108).

The remaining options are used to specify the information that should be displayed in the table.

**COUNT** - This option allows you to select which indicator will be used to quantify the importance of codes. Four indicators are available:

- Code Occurrence (case)
- Code Frequency.
- Word Count
- Percentage of Words

Selecting **Code Occurrence** allows you to compare the number of cases in which at least one instance of this code appears. If a code is used more than once in the same case, it will be counted only once. To count the total number of times a code has been used, select **Code Frequency**. Selecting **Word Count** will retrieve the total number of words that have been assigned to a specific code, while the **Percentage of Words** option divide this value by the total number of words found in those documents.

**DISPLAY** - The DISPLAY list box allows you to specify the information displayed in the table. The following options are available:

- Count
- Row Percent
- Column Percent
- Total Percent

When the COUNT option is set to **Code occurrence**, an additional statistic become available:

- Category percent (percentage of cases or individuals in this subgroup)

**STATISTIC** - This list box allows you to choose among 11 association measures to assess the relationship between the numeric or categorical variable and the codes.

Nominal level statistics

- Chi-square
- Student's F

Ordinal or internal level statistics

- Tau-a
- Tau-b
- Tau-c
- Somers' D (symmetric)
- Somers' Dxy (asymmetric)
- Somers' Dyx (asymmetric)
- Gamma
- Spearman's Rho
- Pearson's R

**PROBABILITY** - The probability option allows you to select whether the probability value should be calculated using a **1-tailed** or **2-tailed** test. Probabilities of Chi-square, Likelihood ratio, and Student's F are always calculated using a 2-tailed test.

After a first crosstabulation table has been obtained, additional buttons located on top of this dialog box will become available. These buttons are used to create barcharts or line charts representations of data in selected rows, perform correspondence analyses, or create heatmaps. You will find below short descriptions of the steps needed to access these features. A more detailed description of all these features will be presented later.

## Creating Barcharts or Line Charts

- Barcharts or line charts are useful for visually comparing the distribution of specific codes among values of an independent variable such as subgroup of individuals (male vs. female) or time periods. To produce these types of charts:
- Set the TABULATE WITH, COUNT and DISPLAY options so that the information to visualize is displayed in the table.
- Using the mouse, select the rows that you would like to display. Multiple but separate rows can be selected by clicking while holding the **CTRL** key down.
- Click the  button.

For more information, see **Barchart and Line Chart** on page 187.

## Creating Bubble Charts

Bubble charts are graphic representations of contingency tables where relative frequencies are represented by circles of different diameters.

- Set the TABULATE WITH, COUNT and DISPLAY options so that the information to view is displayed in the table.
- Click  button.

For more information, see **Bubble Charts** on page 200.

## Creating Heatmaps with Clustering of Rows and Columns

A heatmap plot is a graphic representation of crosstab tables where cell frequencies are represented by different color brightnesses or tones. When combined with clustering of rows and/or columns, this exploratory tools allows you to identify functional relationships between specific codes and subgroups defined by values of the independent variable. To create a heatmap:

- Set the TABULATE WITH, COUNT and DISPLAY options so that the information to visualize is displayed in the table.
- Click the  button to access the heatmap dialog box.

For more information on heatmaps, see **Heatmap Plot** on page 191.

## Performing Correspondence Analysis

Correspondence analysis is an exploratory technique that provides a graphic overview of relationships in large crosstabulation tables of frequency. To perform a correspondence analysis:

- Set the TABULATE WITH, COUNT and DISPLAY options so that the information to visualize is displayed in the table.
- Click the  button to access the correspondence analysis dialog box.

For more information on correspondence analysis, see **Correspondence Analysis** on page 196.

## Printing and Exporting a Crosstabulation Table

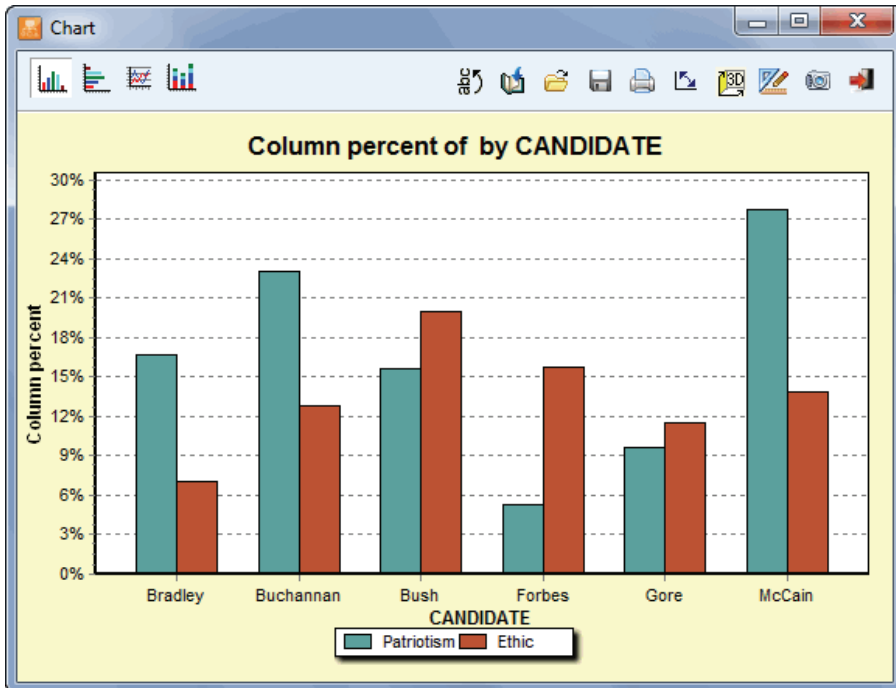
The crosstabulation table may be saved to disk in Excel, MS Word, plain ASCII, text delimited, HTML, or XML format by clicking the  button.

To append the table to the Report Manager, click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.

You may also print this table by clicking the  button.

# Barchart or Line Chart

The chart window allows you to graphically examine the relationship between codes and values of an independent variable. The barchart should be used to display the distribution of various categories within subgroups as defined by a nominal independent variable, while the line chart may be preferred for the examination of the relationship between these categories and an ordinal or quantitative variable.



## To Search for Selected text segments:

- Right-click on the bar corresponding to the code and the specific value for which you would like to retrieve associated coded segments.
- Select the RETRIEVE SEGMENTS command. All retrieve segments will be displayed in a separate dialog box. For more information on the various features available from this dialog box, see **Using the Coding Retrieval Dialog Box** on page 143.

## To Remove specific series:

- Right-click on any bar corresponding either to the code or to the specific value you would like to remove.
- The popup menu will display two REMOVE commands, one of them for removing a specific code, while the other one for removing all bars associated with a specific value. Select the one associated with what you would like to delete.

# The Toolbar

The following table provides a short description of the available buttons and controls:

| Control                                                                             | Description                                                                                                                                                                                                                         |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | Press this button to vertically display the labels on the bottom axis                                                                                                                                                               |
|    | Press this button to append a copy of the graphic in the Report Manager. A descriptive title will be provided automatically. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button. |
|    | Click this button to retrieve a chart previously saved on disk.                                                                                                                                                                     |
|    | Click this button to save a chart on disk. Charts may be saved in BMP, JPG or PNG graphic file format or may be saved in a proprietary format (.WSX file extension) that can later be edited and customized using the Chart Editor. |
|    | Clicking this button prints a copy of the displayed chart.                                                                                                                                                                          |
|    | Clicking this button causes the values represented on the bottom axis to be switched with those represented by different lines or bars (legend).                                                                                    |
|    | Click this button to turn on/off the 3D perspective for the current chart.                                                                                                                                                          |
|    | This button allows the editing of various features of the chart such as the left and bottom axes, the chart and axes titles, the location of the legend, etc. (see <b>Barchart and Line Chart Options</b> below)                    |
|  | This button creates a copy of the chart to the clipboard. When this button is clicked, a shortcut menu appears allowing you to select whether the chart should be copied as a bitmap or as a metafile.                              |
|  | Pressing this button closes the <b>chart</b> dialog box and returns to QDA Miner's main screen.                                                                                                                                     |

## Barchart and line chart options dialog box

The various options in this dialog box are used to customize the appearance of barcharts and line charts. The options available in this dialog box represent only a small portion of all settings available.

To further customize the chart, edit data points, value labels, or series order, click the  button located on the right side of the dialog box.

## LEFT AXIS

**MINIMUM / MAXIMUM** - QDA Miner automatically adjusts the vertical axis scale to fit the range of values plotted against it. To manually set these values, enter the desired minimum and maximum.

**INCREMENT** - Increasing or decreasing this value affects the distance between numbers as well as tick marks. Horizontal grid lines are also affected by changing this value.

**HORIZONTAL GRID** - This option turns horizontal grid lines on and off. Grid lines extend from each tick mark on an axis to the opposite side of the graph. To increase or decrease the number of grid lines or the distance between these lines, change the Increment value of the axis. A list box also allows a choice among five different line styles to draw grid lines.

## LEGEND

**LOCATION** - This option positions the legend. Legends may be placed at the Top, Left, Right and Bottom side of the chart.

**FROM TOP** - When the legend is displayed on the left or the right side of the chart, this option specifies the legend's top position in percentage of total chart height.

**FROM LEFT** - When the legend is displayed on the top or the bottom chart, this option specifies the legend's top position in percent of total chart height.

## TITLES

Proper titles and axis labels are of utmost importance when describing the information displayed in a chart. By default, QDA Miner uses variable names and labels as well as other predefined settings to provide descriptions.

The title page allows you to modify the top title, as well as the labels on the left, bottom and right axes. To edit the title, select the proper radio button. Enter several lines of text for each title by pressing the **Enter** key at the end of a line before entering the next line.

The **Font** button on the right side of the edit box changes the font size or style of the related title.

## 3D VIEW

**ORTHOGONAL** - Turning this option off disables the free elevation and rotation of the 3D chart.

**ZOOM** - This option zooms in and out of the whole chart. Expressed as a percentage, this value, when increased positively, will bring the chart towards the viewer, increasing the overall chart size as the Zoom value increases.

**3D PERCENT** - The 3D Percent property indicates the size ratio between chart dimensions and chart depth by specifying a percent number from 1 to 100.

**PERSPECTIVE** - Use this property with Orthogonal unchecked to change the 3D perspective of the Chart. Larger values add more depth perspective.

**BAR SHADOW** - Enabling this option dark-shades the sides of 3D bars. Turning it off will paint the sides of the bar in the same color as the front of the bar.

**BAR WIDTH** - This option determines the percentage of total bar width used. Setting this value to 100 joins the bars.

**BAR DEPTH** - Use this property to limit the depth that each bar series uses. By default, bars will take up the part proportional to the number of bar series in the chart so that the back of a bar will join the front of the bar immediately behind it. To insert a gap between a series of bars, decrease this value.

# Heatmap Plot

Heatmap plots are graphic representations of crosstab tables where relative frequencies are represented by different color brightnesses or tones and on which a clustering algorithm is applied to reorder rows and/or columns. This enter of plot is commonly used in biomedical research to identify gene expressions. When used in qualitative analysis, this exploratory data analysis tool facilitates the identification of functional relationships between related codes and group of values of an independent variable by allowing the perception of cells clumps of relatively high or low frequencies or of outlier values.

The heatmap plot implemented in QDA Miner allows you to graphically examine the relationship between codes (rows) and values of an independent variable (columns). While the clustering available through the CODING CO-OCCURRENCES command (see **Clustering Cases** on page 167) is performed on individual documents, the clustering analyses used in the heatmap plot are performed directly on the crosstabulation tables. As a consequence, the similarity index computed for two codes and used for clustering do not represent their co-occurrences within cases but measures the similarity of their distribution among the various groups of the independent variable. Likewise, two subgroups defined by values on the independent variable will be considered near to each other if the distributions of codes in those two groups are similar.

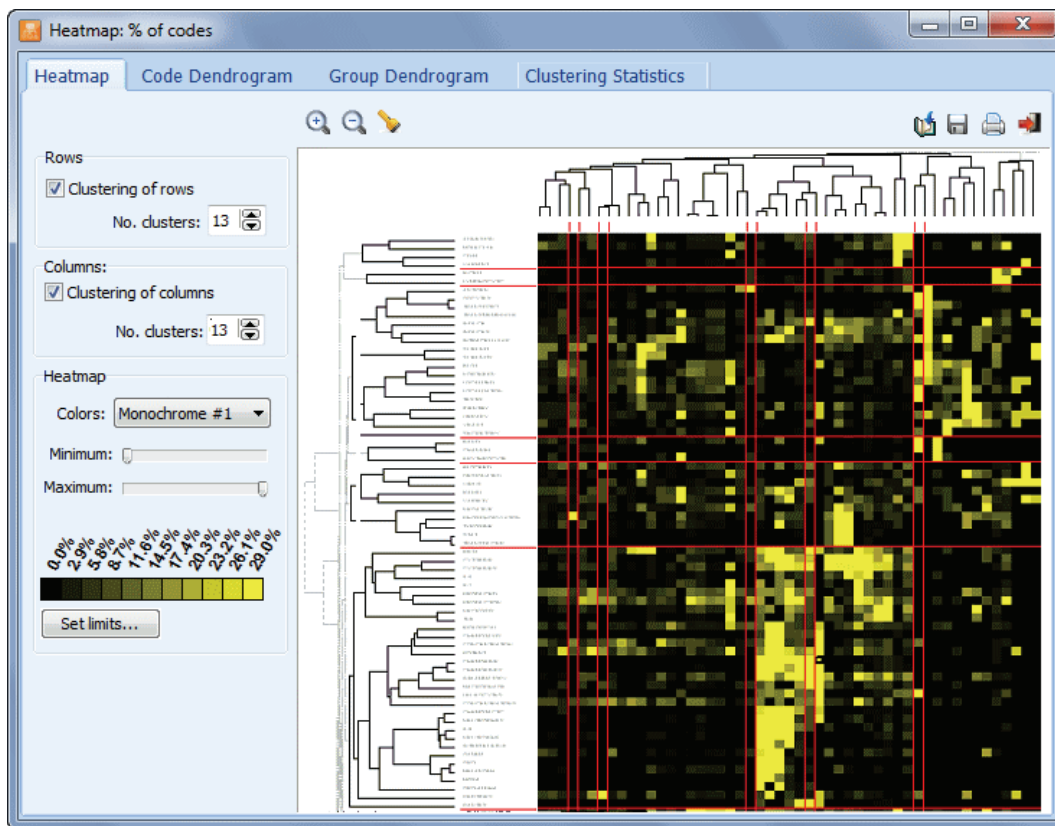
The heatmap plot can be performed on any of the available statistics (Code Occurrence, Code Frequency, Word Count, or Percentage of Words). For all these types of analyses, the observed frequency is transformed into a percentage by dividing the observed value by an appropriate statistic calculated for each subgroup of cases.

## To create a heatmap:

- Select the CODING BY VARIABLES command from the ANALYSIS menu.
- Select the codes you want to plot along with the documents to which you wish to restrict the analysis.
- Set the TABULATE WITH option to the desired independent variable.
- Set the COUNT option to either **Code Occurrence**, **Code Frequency**, **Word Count** or **% of Words**.
- Click the  button to construct the crosstab matrix.
- Click the  button.

## Heatmap Page

The main section on the right of this page, as shown in the screen shot below, displays the heatmap grid representing the relative frequencies of each cell (row and column intersection) using different brightnesses or color tones. Optional dendrograms are displayed at the top and on the left margin of this grid. The size of these dendrograms may be adjusted by moving the mouse cursor over the bottom edge (upper dendrogram) or the right edge (dendrogram on the left) of the dendrogram and dragging its limit to the desired size.



The font size used to display the row and column values may also be adjusted by clicking the  or  buttons located on the top toolbar. The size of cells and the distance between dendrogram leaves are automatically adjusted to the new font size.

## ROWS OPTIONS GROUP

**CLUSTERING OF ROWS** - Enabling this option reorders the rows based on the result of a cluster analysis and displays a dendrogram to the left of the heatmap plot. Codes that are distributed across the various subgroups in a similar way will tend to be grouped under the same cluster.

**SORT BY** - When code clustering is disabled, rows may be sorted in alphabetical order, in descending order of frequency or displayed in their original codebook order.

**NO. CLUSTERS** - This option allows you to set how many clusters the clustering solution should have. When two clusters or more are selected, horizontal red lines are drawn in the heatmap to delineate these clusters.

## COLUMNS OPTIONS GROUP

**CLUSTERING OF COLUMNS** - Selecting this option reorders the columns based on the result of a cluster analysis and display a dendrogram at the top of the heatmap. Columns with similar distribution of codes will tend to be grouped under the same cluster. When this clustering option is disabled, the columns are presented in their ascending order of values. Disabling the clustering of columns is especially useful to preserve the ordinal nature of the values. For example, when looking for a relationship between some codes and publication years, then disabling the clustering of columns will automatically sort those columns in chronological orders allowing the identification of temporal trends.

**NO. CLUSTERS** - This option allows you to set how many clusters the clustering solution should contain. When two clusters or more are selected, vertical red lines are drawn in the heatmap to delineate these clusters.

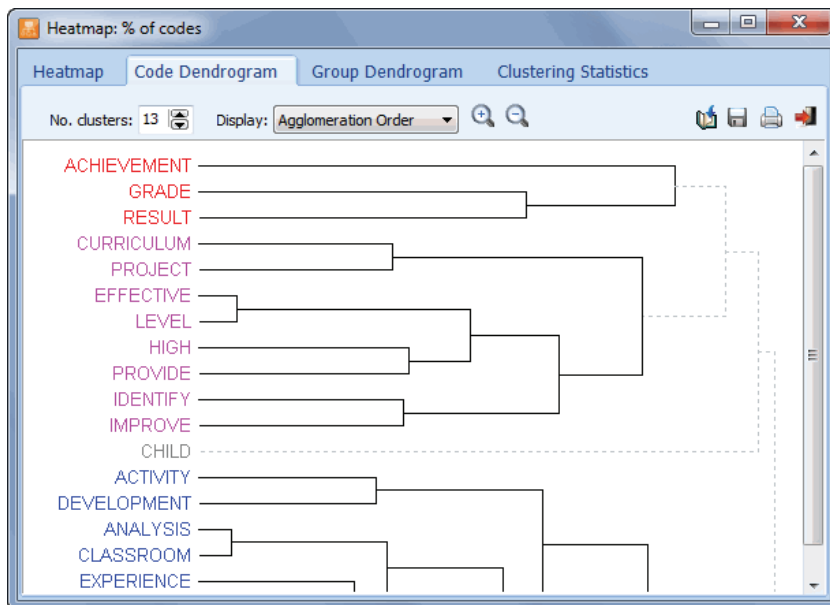
## HEATMAP OPTIONS GROUP

**COLORS** - The colors list box allows the selection of various color schemes to represent differences in percentages. Monochrome schemes will express differences using levels of brightness for a single color where black always represents the lower limit of the selected range. Multicolor spectrums may be used to represent greater nuances in the selected range of percentages.

**MINIMUM / MAXIMUM** - The minimum and maximum slide bars allows you to set the range of values that will be used to display variations of color tones or brightness. By default, the minimum value is set to zero while the maximum value is set to the highest observed percentage. To increase the contrast at a specific location of this range, minimum and maximum limits may be adjusted. All cells with values lower than the minimum will be represented with the color located at the left end of the selected color spectrum, while cells with percentages higher than the maximum limit will be displayed with the color locate at the right end of this spectrum. Reducing the range of values increases the contrast in a specific region of percentage values.

## Code and Group Dendrogram Pages

The second and third pages of the heatmap dialog box provide a more detailed examination of the clustering of codes (**Code Dendrogram**) and of all values on the independent variable (**Group Dendrogram**). QDA Miner uses an average-linkage hierarchical clustering method to create clusters from a similarity matrix. The result is presented in the form of a dendrogram (see below), also known as a tree graph. In this type of graph, the vertical axis is made up of the items and the horizontal axis represents the clusters formed at each step in the clustering procedure. Items that tend to be distributed similarly against the other variables appear together and are combined at an early stage, while those that have dissimilar distributions tend to be combined at the end of the agglomeration process.



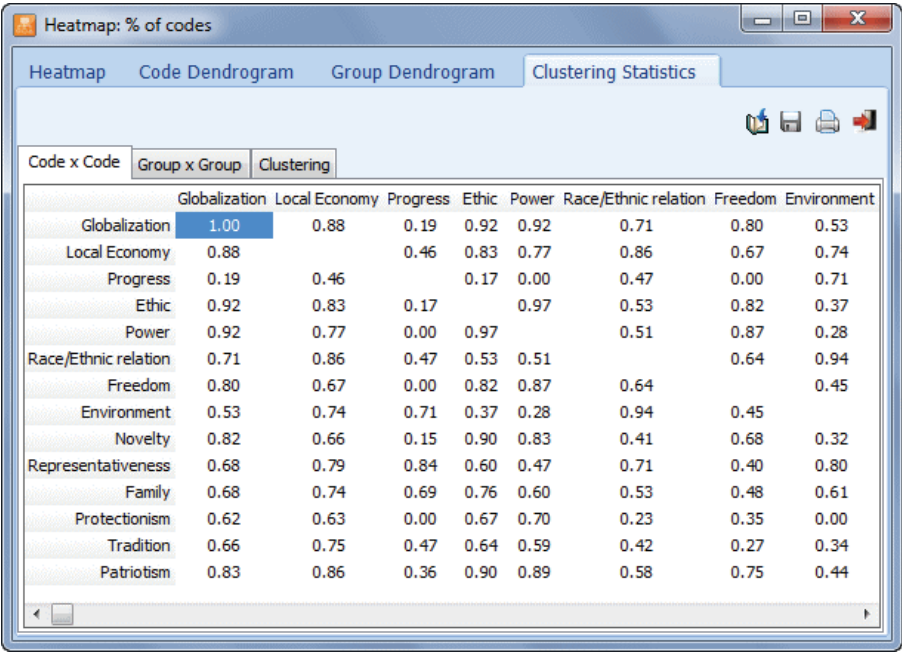
**NO. CLUSTERS** - This option sets how many clusters the clustering solution should have. Different colors are used in the dendrogram to indicate membership of specific items in different clusters.

**DISPLAY** - This option sets whether the vertical lines of the dendrogram represents the agglomeration schedule or the similarity indices.

At times, you will want to see some areas of a dendrogram. The  and  buttons may be used to zoom in or out of the dendrogram.

# Clustering Statistics Page

The fourth page of the heatmap dialog box allows the close examination of matrices of similarities among codes or among groups as well as the statistics associated with the agglomeration processes of both the codes and the groups.



# Correspondence Analysis

Correspondence analysis is a descriptive and exploratory technique designed to analyze relationships among entries in large frequency crosstabulation tables. Its objective is to represent the relationship among all entries in the table using a low-dimensional Euclidean space such that the locations of the row and column points are consistent with their associations in the table. The correspondence analysis procedure implemented in QDA Miner allows you to graphically examine the relationship between assigned codes and subgroups of an independent variable. The results are presented using a two- or three-dimensional map. Correspondence analysis statistics are also provided to assess the quality of the solution. QDA Miner currently restricts the extraction to the first three axes. This ensures that the results remain easy to interpret. This limitation may however be removed in future versions upon request from users.

The first two pages of the dialog box provides graphical views of correspondence maps. When the number codes or the number of subgroups is less than four, only the **2D Map** page can be accessed. When the comparison is restricted to two groups of individuals, only one axis can be extracted. In this situation, the axis is plotted diagonally in the two-dimensional space. This has been done mainly for readability reason: plotting all the code labels on a single horizontal axis would have produced a cluttered list of codes that would have made the graph useless.

You will find immediately below a description of the various controls in the dialog box, followed by some guidelines for interpreting correspondence plots.

## 2D Map control

**PLOT** - When more than two axes have been extracted, this control allows the selection of all possible axis combinations that can be graphed on the two axes of the plot.

## 2D and 3D Map Controls

**Codes** This checkbox displays or hides the row points (i.e., code names).

**Groups** This checkbox displays or hides the column points (i.e., subgroup labels)



Clicking this button enables to zoom in on a plot. To zoom in on a area of the plot, hold the left mouse button down and drag the mouse down and to the right. A rectangle will appear around the selected area. Release the left mouse button to zoom in.



Clicking this button restores the original viewing area of the plot.



Items distributed in a very similar way among subgroups of the categorical variables may be plotted on top of each other, making them difficult to differentiate. Clicking down this button adds some random noises to the location of individual words or keywords, allowing one to clearly identify those words that overlap. To remove the random noises, click the button again.



Moving this slider allows the adjustment of the scaling used to plot groups or classes of the categorical variable on the correspondence plot. While the positions of codes and groups relative to the origin of the axes are significant, it is important to remember that the scaling used for codes and groups are independent from each other. This slider may thus be used as a reminder of this fact. It may also be used to increase the readability of some plots, by moving group names so that they do not overlap code names.



Press this button to append a copy of the correspondence plot in the Report Manager. A descriptive title will be provided automatically. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.



This button is used to create a copy of the chart on the clipboard. When this button is clicked, a shortcut menu appears, allowing you to choose whether the chart should be copied as a bitmap or as a metafile.



Clicking this button opens a dialog box to customize the appearance of the correspondence plots.



Clicking this button prints a copy of the displayed chart.



Clicking this button closes the correspondence analysis dialog box and returns to QDA Miner's main screen.

### 3D Plot controls



This button can be used to show or hide left, bottom and back walls.



Clicking this button draws anchor lines from the floor to the data point to better locate data points in all 3 dimensions.



Clicking this button changes the viewing angle of the chart. To rotate the chart, make sure this button is selected, click any area of the chart, hold the mouse button down and drag the mouse to apply the desired rotation.



Locating a data point on the depth dimension of a 3D plot can be very difficult, especially when the plot remain static. Often this plot must be rotated constantly on the various axes to obtain an accurate idea of where the data point is located on this third axis. Clicking this button rotates the plot automatically. To disable the automatic rotation, click a second time.



This button is used to create an AVI movie file of the rotating chart. It behaves as both a record and a stop button. To start recording, simply click this button. Click again to stop the recording and to save the AVI movie file to disk.

# Interpreting Correspondence Analysis Results

Interpretation of correspondence analysis maps can be somewhat tricky and should be done with great care, especially when examining the relationship between row points and column points. Here are some basic rules that should help you interpret these maps:

## Relationship among codes (row points)

- The more similar the distribution of a code among subgroups is to the total distribution of all codes within subgroups, the closer it will be to the origin. Codes that are plotted far from this point of origin have singular distributions.
- If two codes have similar distributions (or profiles) among subgroups of the independent variable (columns), their points in the correspondence analysis plot will be close together. For example, if the codes consist of artist names and the studied subgroups represent different age groups then, if the form of the distribution of two different artists among these age groups is similar, they will tend to appear near each other. Codes with different profiles will be plotted far from each other. Note: Two points may appear close to each other on a two- or three-axis solution, but may in fact be far apart when an additional dimension is taken into account.

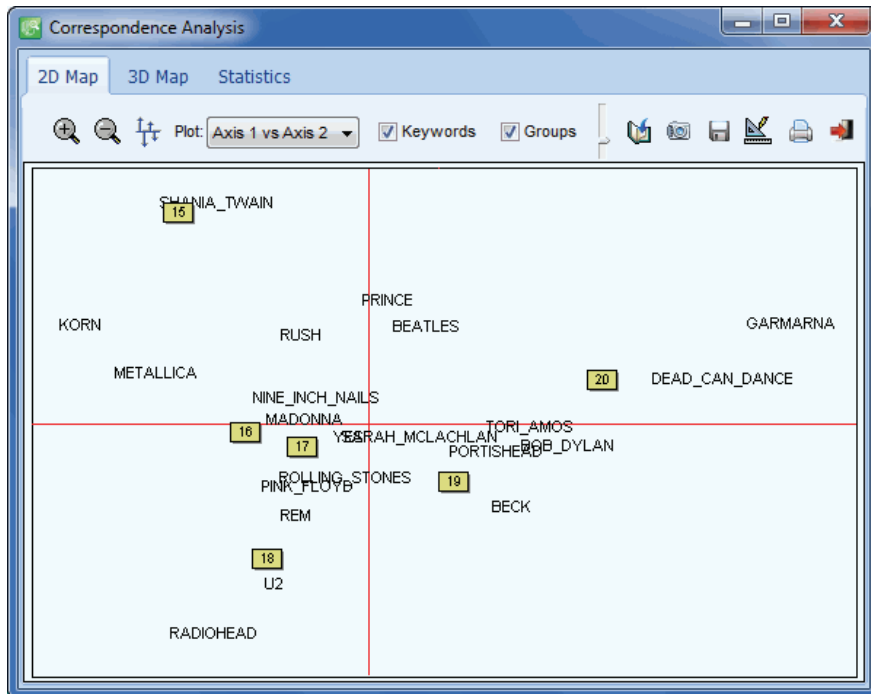
## Relationship among subgroups (column points)

- The more singular a profile of coding for a subgroup is, compared with the distribution of these codes for the entire sample, the farther this subgroup will be from the point of origin.
- If two subgroups of individuals have similar profiles of coding, they will be plotted near each other. Subgroups with different profiles will be plotted far from each other. Again, it is important to remember that two points may appear close to each other on a two- or three-axis solution, but may in fact be far apart when an additional dimension is taken into account.

## Relationship between codes and subgroups (row and column points):

- Great caution should be taken when interpreting the distances between two sets of points (row and column points). The fact that the name of a subgroup is near a specific code should not necessarily be interpreted as an indication that they are closely related.
- While the distance between codes and subgroups has no interpretable meaning, the angle between this code point and a subgroup point from the origin is meaningful:
- An acute angle indicates that the two characteristics are correlated.
- An obtuse angle, near 180 degrees, indicates that the two characteristics are negatively correlated.

In the following example, REM, U2 and RadioHead could be viewed as related to 18-year-old listeners. However, despite the fact that both REM and U2 points are closer to the 18-year-old subgroup, Radiohead should be identified as more characteristic of those listeners since it is farther from the origin.



- Codes closely associated with two subgroups will be plotted at an angle from the origin that will lie between those two groups. In the above example, the rock groups Korn and Metallica seem to be characteristic of both 15 and 16-year-old listeners.

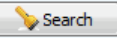
For a more complete description of this method, its calculation and applications, see Greenacre (1984).

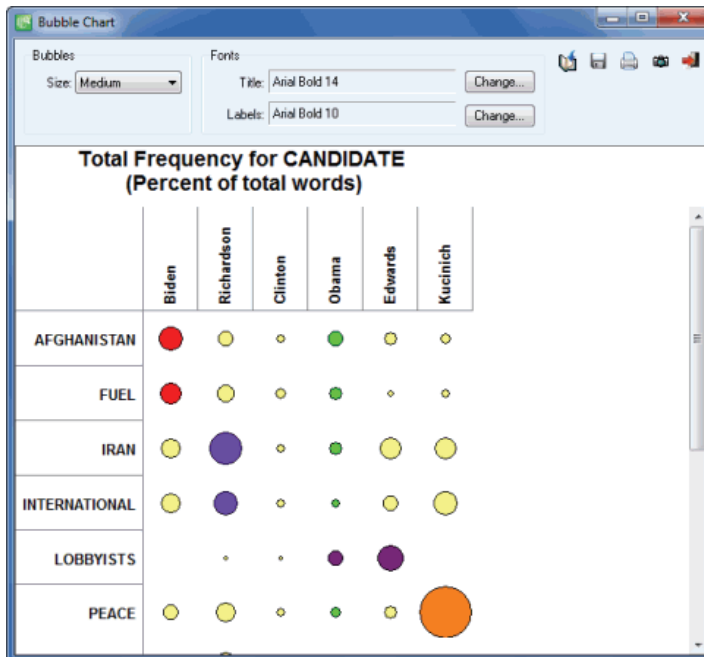
# Bubble Charts

Bubble charts are graphic representations of contingency tables where relative frequencies are represented by circles of different diameters. This type of graphic chart allows one to quickly identify high- and low-frequency cells and is thus especially useful for presentation purposes. Many features of the chart can be customized to highlight specific findings. Rows and columns can be moved freely or deleted, and one can adjust the color of each cell as well as the fonts used in the chart.

The bubble chart represents graphically the underlying table and displays the corresponding measure used (code occurrences or frequencies, word counts or percentages) and will also reflect the currently selected display settings (such as count, percentage of rows or columns, etc.).

## To create a bubble chart:

- Select the CODING BY VARIABLES command from the ANALYSIS menu.
- Select the codes that you want to plot, along with the documents to which you wish to restrict the analysis.
- Set the **Tabulate** With option to the desired independent variable.
- Set the **Count** option to either **Code Occurrence**, **Code Frequency**, **Word Count**, or **% of Words**.
- Set the **Display** option to specify how this information will be displayed.
- Click the  button to construct the frequency matrix.
- Click the  button. A dialog box similar to this one will appear:



## To adjust the size of the bubbles:

- In the **Bubbles** group box at the top of the window, adjust the **Size** option to Small, Medium or Large.

## To adjust the color of the bubbles

### Global adjustment

- To adjust the color of all bubbles, set the **Color** option to either **Codebook** or **Single Color**. Choosing the **Codebook** option will reset the color of bubbles to the color associated with each code in the codebook. Selecting **Single Color** will bring a dialog box that will allow you to select which color should be used for displaying all bubbles in the chart.

### Manual adjustment

- To adjust the color of specific bubbles, select the cell or group of cells you would like to alter.
- Right click and select the SET COLOR command. A dialog box will appear letting you choose a specific color value.

## To move a row or a column

- Click anywhere on the cell header containing the row or column label and hold the mouse down.
- Drag the mouse cursor over the desired new location and release the mouse button.

## To delete selected rows or columns

- Select a cell range covering the rows or columns you would like to delete.
- Right-click to display a popup menu.
- Select the REMOVE command and then choose SELECTED ROWS or SELECTED COLUMNS depending on your desired action.

## To adjust the font size and style of the title and labels

- Click the **Change** button beside the **Title** or **Labels** options. A Font setting dialog box will appear, letting you change the font, the font size, style, and color.

## To edit the title

- Click anywhere in the title region and type in the new title. The height of the title region is automatically adjusted to the number of lines in the title.

The following table provides a short description of additional buttons:

| Control | Description |
|---------|-------------|
|---------|-------------|



Press this button to append a copy of the graphic in the Report Manager. A descriptive title will be provided automatically. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.



Click this button to save a chart on disk. Charts may be saved in BMP, JPG or PNG graphic file format or may be saved in a proprietary format (.WSX file extension) that can later be edited and customized using the Chart Editor.



Clicking this button prints a copy of the displayed chart.



This button creates a copy of the chart to the clipboard. When this button is clicked, a shortcut menu appears allowing you to select whether the chart should be copied as a bitmap or as a metafile.



Pressing this button closes the chart dialog box and returns to QDA Miner's main screen.

# Assessing Inter-coders Agreement

When document coding is performed manually by human coders, individual differences in interpretation of codes between different coders often occur no matter how explicit, unambiguous and precise the coding rules are. Even a single coder is often unable to apply the same coding rules systematically across time. One way to ensure the reliability of the application of coding rules is to ask different raters to code the same content or to ask a single rater to code the same document at different times. The comparison of codes is then used to uncover differences in interpretation, clarify ambiguous rules, identify ambiguity in the text, and ultimately quantify the final level of agreement obtained by these raters. Unfortunately, the application of inter-coders agreement procedures often involves some requirements or assumptions that are not always compatible with the qualitative data analysis processes. At least two compatibility problems can be identified:

**1. The Codebook Problem** - Computing inter-coders agreement requires that all coders share the same codebook. They should not be allowed to create their own codes or modify existing ones. An entirely inductive approach that would give each coder the possibility of developing his own codes during the analysis would result in a situation where the end result would very likely be incommensurable, each coder ending up with entirely different categories. In this situation, it would be impossible to establish the inter-raters agreement per se. Probably one of the only types of agreements that may be achieved in this case is by judging whether their final conclusions agree. While using a predefined unmodifiable codebook is the most obvious solution to this problem, a mixed solution may also be adopted. For example, the first coder may be given the opportunity to develop his own codebook or refine an existing one by adding categories or revising existing ones. Once the coding by the first coder is finished or once he is satisfied with the codebook state of development, this codebook can then be used by other coders to code existing documents. Another possibility would be to have a codebook that would contain a portion of unmodifiable codes but leave coders the possibility of modifying another portion of them or add new ones. In the latter case, only the fixed portion of the codebook could be submitted to an assessment of intercoders agreement.

**2. The Segmentation Problem** - Another problem is related to the common absence of predefined segments to be coded. The flexibility provided by many computer qualitative data analysis software to define the length and location of a segment to be coded creates a situation where the number of coded segments, their lengths and locations will very likely differ from one coder to the next. In this situation, a strict agreement definition that compares the assignment of the same codes with the same text segment would be too stringent to apply. The most common solution to this segmentation problem has been the prior segmentation of documents to be coded. However, an alternative solution is to somewhat relax the definition of the agreement.

QDA Miner allows to circumvent some of the constraints imposed by the inter-coders agreement procedures by providing different levels of agreement and by relaxing the criterion used to establish the overlap of segments..

But first, in order to assess inter-coder reliability, some prior settings and conditions are required:

- The project should be configured to support login by more than one user (for more information on how to define multiple users, see **Security & Multi-users Settings** on page 27).
- Each user with coding rights should then log into the project using his own user name and password in order to apply codes to documents.

- To ensure the independence of the coding process, a specific user should not be allowed to view code assignments made by other users. The QDA Miner **Multi-users Settings** page provides the option of hiding all codings made by other users.
- Each user has to code a common portion of the documents using the same codebook. The agreement can only be calculated on documents that have been coded by more than one user. All other documents that have been coded by only one user will be ignored in the computation of the inter-coders agreement.

An alternative way to perform independent coding is to create multiple copies of an uncoded project, ask each coder to apply codes on their own copy of the project, and then combine all codings stored in different project files into a single master project using the project merging feature of QDA Miner (see **Merging Projects** on page 65). Each coder, however, **MUST** log on with their own user name

Only when all the above conditions have been met will it be possible to assess the agreement between coders.

## Computing Inter-coders Agreement

In order to access the inter-coders agreement procedure, one should have the right to view codes created by other users. If a user does not have this right, only the administrator can remove this restriction by changing the multi-user settings.

To assess the inter-coders agreement, select the CODING AGREEMENT procedure from the ANALYSIS menu. A dialog box similar to the one below will appear:

**Inter-coders agreement**

Analysis | List of Hits

Search in: [SPEECH]

Codes: [Globalization;Local Economy;Ethnic;Power;Environment;Freedom;Novelty;Race/Ethnic]

Coders: [Admin;Guest;Normand]

Agreement criterion: Segments must overlap

Overlap: 90 % ☐ Common absences as agreements

Statistic: Krippendorff's Alpha

**Search**

| CODE                 | AGREE ABSENT | AGREE PRESENT | DISAGREE  | PERCENT      | ALPHA        |
|----------------------|--------------|---------------|-----------|--------------|--------------|
| Environment          | 51           | 1             | 0         | 100.0%       |              |
| Ethic                | 47           | 3             | 2         | 75.0%        | 0.333        |
| Family               | 47           | 4             | 1         | 88.9%        | 0.437        |
| Freedom              | 50           | 1             | 1         | 66.7%        | 0.250        |
| Globalization        | 50           | 1             | 1         | 66.7%        | 0.250        |
| Local Economy        | 48           | 4             | 0         | 100.0%       |              |
| Novelty              | 50           | 1             | 1         | 66.7%        | 0.250        |
| Patriotism           | 39           | 11            | 2         | 91.7%        | 0.455        |
| Power                | 47           | 3             | 2         | 75.0%        | 0.333        |
| Race/Ethnic relation | 47           | 5             | 0         | 100.0%       |              |
| Representativeness   | 39           | 10            | 3         | 87.0%        | 0.425        |
| Tradition            | 47           | 3             | 2         | 75.0%        | 0.333        |
| <b>TOTAL</b>         | <b>562</b>   | <b>47</b>     | <b>15</b> | <b>97.6%</b> | <b>0.340</b> |

The following four options need to be set prior to an analysis:

**SEARCH IN** - This option allows you to specify on which document variables the analysis will be performed. If the current project contains more than one document variable, you will have a choice to select either one or more of them. By default, all document variables are selected. To restrict the analysis to a few of them, click the down arrow key at the right of the list box. A list of all available document variables will appear. Select the variables on which you want the analysis to be performed.

**CODES** - This option allows you to select which codes will be analyzed. By default, the analysis is performed on all codes in the codebook. You can select specific codes either from a drop-down checklist by clicking the arrow button or from a tree representation of the codebook by clicking the  button.

**CODERS** - This option allows you to select the coders for which you want to assess the coding agreement. At least two coders must be selected. If more than two coders are selected, the agreement table will be calculated on all possible pairs of coders provided that they have coded the same documents. By default, this option is set to include all coders. To restrict the analysis to coded segments created by a few of them, click the down arrow key at the right of the list box. A list of all coders will appear. Clear check marks beside the coders whom you want to exclude from the analysis.

**AGREEMENT CRITERION** - QDA Miner provides a way to assess four different levels of agreement. The **Agreement Criterion** option is used to select the level of agreement that will be tested by offering a choice of four criteria:

**Code occurrence** - This option checks whether coders agree on the presence of specific codes in a document, regardless of the number of times this code appears and of code location. The agreement is calculated on a dichotomous value that indicates whether the code is present or absent.

☞ Some coders may agree that something is present in a document without necessarily agreeing on what specifically are the manifestations of this "something". For example, a researcher may ask two coders to check numerous job interviews and ask them to decide whether there are some indications of anxiety from the candidate. They may very well agree in the identification of such anxiety despite their inability to agree on its specific manifestations. If a research hypothesis or an interpretation is related to the mere presence or absence of anxiety in various candidates, this level of agreement may well be sufficient. Coders do not necessarily have to agree on exactly where it occurs.

**Code frequency** - This option checks whether coders agree on the number of times specific codes appear in a document. This criterion does not take into account the specific location of the codes in the document. The agreement is calculated by comparing the observed frequency per document for each coder.

- ☞ If someone wants to examine how often a specific speaker uses religious metaphors in his speeches compared with other speakers, he may ask different coders to assign a code "Religious Metaphor" every time they encounter one in a speech. Even if they disagree somewhat on what represents a religious metaphor, they may nevertheless agree on how many of those metaphors they find in each speech. Since the research question is related to the comparison of frequency of metaphors among speakers, then an agreement on the overall frequency per speech should be sufficient.

**Code Importance** - This option allows you to assess whether coders agree on the relative importance of a code in the document. This kind of agreement is achieved by comparing for a given code the percentage of words in the document that have been tagged as instances of this.

- ☞ If a researcher makes the hypothesis that when adolescents are asked to talk about a love relationship, male adolescents devote more time than females talking about the other person's physical characteristics and less time about their own needs or feelings. With this kind of hypothesis, the researcher needs to establish an agreement on the relative importance of these topics, not necessarily on the frequency or location. Again, coders may somewhat disagree on what specific segments represent instances where the adolescent is talking about their needs or feelings, but they may nevertheless come to a close agreement on the importance of each code.

**Code Overlap** - This last level of agreement is the most stringent criterion of agreement since it requires the coders to agree not only on the presence, frequency and spread of specific codes, but also on their location. When this option is selected, you need to provide a minimum overlap criterion expressed in percentage. By default, this criterion is set to 90%. If the codes assigned by both users overlap each other on 90% of their length, then these codes will be considered to be in agreement. If the percentage of overlap is lower than 90%, the program will consider them to disagree.

- ☞ Such a level of agreement would be interesting for several reasons. A researcher may be interested in indentifying specific manifestations of a phenomenon. Another important reason is that the examination of agreements and disagreements at this level is accessory to the achievement of high agreement on any other lower level. In other words, close examination of where agreements occur or do not occur should allow one to diagnose the source of disagreement and take corrective actions to establish a shared understanding of what each code means for every coder. In a sense, checking the agreement at this level may be used in a formative way to train coders. Yet, the final summative evaluation of agreement may still be performed on a less stringent level, provided that this lower level matches the research hypothesis.

The first three criteria use documents as the unit of analysis and calculated for each document a single numerical value per code and coder. This numerical value represents either the occurrence, the frequency of the specific code or the percentage of words in the document assigned to it. This Coding Analysis Features 80 measure is then used to establish the level of agreement. For example, when assessing agreement on code frequencies, if a user assigns a specific code to a document four times while another

one uses this code only once for the same document, then the program will add 0.25 to the total number of agreements and 0.75 to the total number of disagreements. The absence of a code in the document is considered to be an agreement. Since the unit of analysis is the document and the numerical value for each code lies between 0 and 1, each document has an equal weight in the assessment of agreement, no matter how many codes are found in these documents.

When using code overlap as the criterion for assessing agreement, the unit of analysis becomes the assigned codes. For this reason, the total number of agreements and disagreements can be higher than for the other three assessment methods. In this case, the absence of a code in the document is not considered to be an agreement.

**STATISTIC** - The simplest measure of agreement for nominal level variables is the proportion of concordant codings out of the total number of codings made. Unfortunately, this measure often yields spuriously high values because it does not take into account chance agreements that occur from guessing. Several adjustment techniques have been proposed in the literature to correct for the chance factor. Three of these are currently available in QDA Miner:

**Free marginal adjustment** assumes that all categories on a given scale have equal probability of being observed. It assumes that coder decisions were not influenced by information about the distribution of the codes. This coefficient is equivalent to the Bennett, Alpert and Goldstein's (1954) S coefficient, Jason and Vegelius's (1979) C Coefficient, and Brennan and Prediger's (1981) *kn* Index (Zwick, 1988).

**Scott's pi** adjustment (Scott, 1955) yields results that are similar to those obtained using Cohen's Kappa (Cohen, 1971). Like the Kappa, it does not assume that all categories have equal probability of being observed and considers that coders decisions are influenced by this information. However, unlike the Cohen Kappa, pi treats any difference among coders in this distribution as a source of disagreement.

**Krippendorff's alpha** (Krippendorff, 2004) is similar to Scott's pi but applies a correction for small sample sizes. The nominal alpha exceeds pi by  $(1-\pi)/2n$  (where n is the number of units coded by the two coders) and asymptotically approximates pi as sample sizes become large.

To store the search expression and options, click the  button and specify the name under which those query options will be saved.

To retrieve a previously saved query, click the  button and select from the displayed list the name of the query you would like to retrieve (for more information, see **Saving and Retrieving Queries** on page 108).

Once the options has been set, click the  button to perform the analysis.

When codings by more than one user are available to assess the inter-coders agreement, two results tables will appear. The first table at the bottom of the **Analysis** page (see above) displays for each code

various frequencies as well the level of agreement obtained. The table also provides information about the overall agreement. The content of this table depends on the level of agreement that has been chosen. You will find below the information such a table can contain:

| COLUMN        | DESCRIPTION                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Code          | Displays the name of the code.                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Agree absent  | Displays the number of times that users agreed on the absence of a code in a document. This column remains empty when assessing agreement using the code overlap criterion.                                                                                                                                                                                                                                                                                 |
| Agree present | Displays the number of times users agreed on the presence of a code in a document. When assessing agreement on code frequency or code importance, this column displays the sum of all proportions of agreement calculated for each document.                                                                                                                                                                                                                |
| Disagree      | Displays the number of times users disagreed on the presence of a code in a document. When assessing agreement on code frequency or code importance, this column displays the sum of all proportions of disagreement calculated for each document.                                                                                                                                                                                                          |
| Percent       | This column reports the percentage of agreement observed between coders. This index is calculated by dividing the total number of agreements by the total number of comparisons possible (agreement + disagreement). This measure does not take into account chance agreements that occur from guessing. For this reason, the percentage of agreement often yields spuriously high values especially when assessing the less stringent levels of agreement. |

The last column of the table displays the selected agreement statistic. This measure attempts to correct the observed level of agreement to take into account chance agreements.

The second table located on the **List of Hits** page displays a list of all codings that were used for the assessment of inter-coders agreement.

Inter-coders agreement

Analysis List of Hits

View: Both

| Case                   | Variable | Start | Category            | Code       | Coder   | Agree with | Disagree with | Text                                                                                                                                                                                                                                                                       |
|------------------------|----------|-------|---------------------|------------|---------|------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Bradley - Announcement | SPEECH   | 1438  | Values\Liberal      | Freedom    | John G. |            | Admin         | I believe in a type of leadership that doesn't shine everyday. I believe in the kind of leader More than 2.8 million children in America live i poverty level. As they grow up, many of ther of "respect" and uninterested in the future. V situation                      |
| Bradley - Announcement | SPEECH   | 2058  | Values\Conservative | Family     | Admin   | John G.    |               | Today, many hard working parents who hold enough time with their children to imbue them national priority. Children need a caring adult meaning in life that is deeper than just the ac potential, America can never reach its own.                                        |
| Bradley - Announcement | SPEECH   | 2059  | Values\Conservative | Family     | John G. | Admin      |               | Today, many hard working parents who hold enough time with their children to imbue them national priority. Children need a caring adult meaning in life that is deeper than just the ac potential, America can never reach its own.                                        |
| Bradley - Announcement | SPEECH   | 2948  | Values\Conservative | Patriotism | Admin   | John G.    |               | I say, think of the unlikely miracle that was th Social Security.                                                                                                                                                                                                          |
| Bradley - Announcement | SPEECH   | 2949  | Values\Conservative | Patriotism | John G. | Admin      |               | I say, think of the unlikely miracle that was th Social Security.                                                                                                                                                                                                          |
| Bradley - Announcement | SPEECH   | 3255  | Values\Conservative | Patriotism | Admin   |            | John G.       | By living in that spirit, not only can individual assure better living conditions for most Amer observe Americans who can see deeper tha that good. When I see American companies i our companies can be that good. When I see more of us - more often - can be that good. |

The table lists the case descriptor, the code name as well as the coders' values used for the calculation of agreement. When assessing code presence, code frequency or code overlap, this value will be either **Yes** (for present) or **No** (for absent). When assessing code importance, these columns will contain the total number of words assigned to this code.

Selecting an item in this table, either by clicking its row or by using keyboard cursor keys such as the **UP** or **DOWN** arrow keys automatically displays the corresponding case and document in the main window. This feature is especially useful for revising the codings made by different users and identify sources of disagreements.

The agreement table as well as the coding tables may be saved to disk in Excel, plain ASCII, text delimited, MS Word, HTML, or XML format by clicking the corresponding  button.

They may also be appended in the Report Manager by clicking the  button. To edit the title used to describe the table, hold down the SHIFT keyboard key while clicking this button.

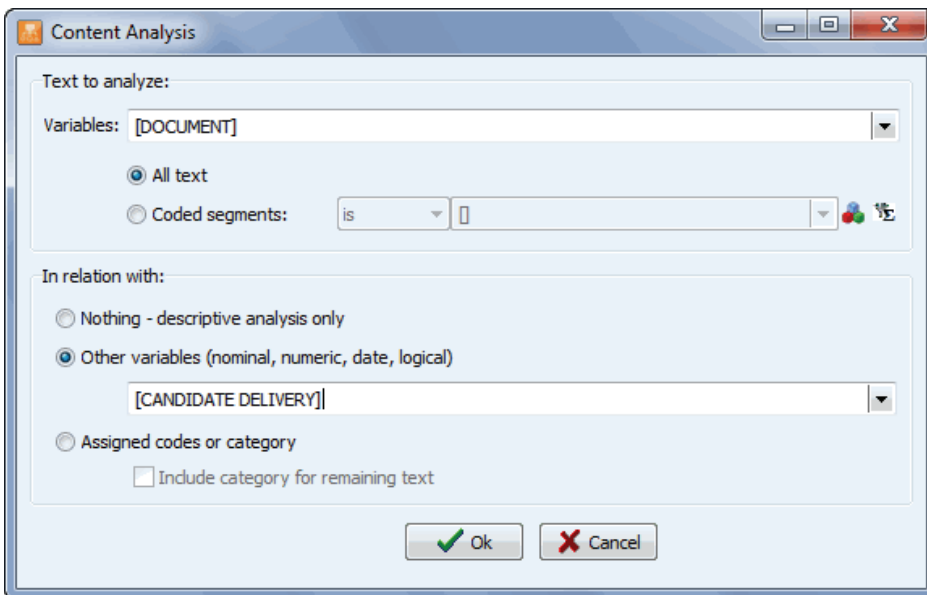
To print any table, click the  button located on the tool bar above the corresponding table.

# Content Analysis

WordStat is a quantitative content analysis add-on module that can be used to analyze words and phrases found in documents. When used with QDA Miner, WordStat can perform these analyses on entire documents or on selected code segments. It may also perform simple descriptive analysis or let you explore potential relationships between words, phrases or categories of words and other numeric or categorical variables.

There are many benefits to combining qualitative and quantitative content analysis techniques. For example, quantitative content analysis may be useful as an exploration tool prior to qualitative coding by allowing one to identify subtle differences in word usage between subgroups of individuals, or to quickly find the most common topics of phrases. Restricting the analysis to segments associated with specific codes may also be useful to identify potential words or phrases associated with those codes. You may then use the QDA Miner text retrieval tool to identify other segments to which this code may be assigned. Quantitative content analysis may also be useful after qualitative coding has been performed. For example, you may try to validate the conclusions drawn from manual coding by comparing those conclusions with the results obtained from a quantitative content analysis. Qualitative codings may also serve as the starting material to develop and validate a categorization dictionary that will allow automatic categorization of documents or cases.

To perform a quantitative content analysis on documents in a project, select the CONTENT ANALYSIS command from the ANALYSIS menu. The following dialog box will appear:



**VARIABLES** - This option allows you to specify on which document variables the analysis will be performed. If the current project contains more than one document variable, you will have a choice of selecting either one or more of them. By default, all document variables are selected. To restrict the analysis to a few of them, click the down arrow key at the right of the list box. A list of all available document variables will appear. Select all variables that you want to analyze.

By default, text analysis is performed on the entire document (**All Text** option). However, it is also possible to restrict the analysis to specific coded segments or exclude segments from the analysis by enabling the **Coded Segments** option. Set the list box immediately to the right of this option to **Is** if you want to analyze only the text found in the selected segments. Set it to **Is not** to exclude these text segments from the analysis. Finally, select the code segments to analyze or to exclude either from a drop-down checklist by clicking the arrow button at the right end of the list box or from a tree representation of the codebook by clicking the  button.

**IN RELATION WITH** - This set of options allows you to specify whether the quantitative content analysis should consider potential differences or relationship with one or more quantitative or categorical variables. If **Nothing** is selected, WordStat will allow descriptive as well as co-occurrence analysis on words or categories of words. Selecting the **Other Variables** option adds the possibility of examining in WordStat the relationship between these words or categories and the values of quantitative or categorical variables. The drop-down list box below this option is used to select the variables in relation to which the analysis will be performed. Selecting **Assigned Codes or Category** allows one to analyze the text assigned to codes and identify potential differences in words or categories of words associated with different codes. When this option is enabled, a temporary file is created containing all text segments associated with the selected codes along with a categorical variable to represent the code name. To add to this temporary file any remaining text segments not associated with any one of those codes, enable the **Include category for remaining text** option.

Once the options have been set, click the **OK** button to call WordStat.

WordStat can also be called from various dialog boxes to analyze retrieved text segments. This feature is available by clicking the  button located on the toolbar above the result table. In such a situation, the text analysis is performed on a temporary data file. For this reason, all coding features normally available in WordStat to apply codes to text segments are disabled.

# Statistical Analysis

QDA Miner shares the same file structure and format as Simstat, an easy-to-use, yet powerful statistical program. When both applications are installed on the same computer, you can call Simstat from within QDA Miner and perform statistical analysis on any numerical or categorical variable in the project without having to export data to an external file, quit QDA Miner, or even close the active project.

Simstat offers a wide range of statistical analysis and graphics. It can also perform numeric and alphanumeric computation, transformation and recoding of variables, as well as advanced file management procedures such as data file merging, file aggregation, etc.

To run Simstat from QDA Miner, simply select the STATISTICAL ANALYSIS command from the ANALYSIS menu. While Simstat is running, QDA Miner will be disabled and it will not be possible to access any of its features. You will regain access to the program as soon as you quit Simstat.

Please note that QDA Miner can produce frequency tables and crosstabulation on numerical and categorical variables and produce various charts (bar charts, pie charts, histograms, box-&-whiskers plots) without the need of run Simstat. For more information on this feature, see page 41.

# Automatic Text Classification

Automated text classification is a supervised machine-learning task by which new documents are classified into one or several predefined category labels based on an inductive learning process performed on a set of previously classified documents. This machine-learning approach of classification has been known to achieve comparable if not superior accuracy to classification performed by human coders, yet at a very low manpower cost. It has been used to automatically classify documents into proper categories or to find relevant keywords describing the content and nature of a document. It has also been used to automatically file or re-route documents or messages to their appropriate destinations, to classify newspaper articles into proper sections or conference papers into relevant sessions, to filter emails or documents (like spam filtering), or to route a specific request in an organization to the appropriate department. Automated text categorization may also be used to identify the author of a document of unknown or disputed authorship.

Developing, optimizing and testing automatic document classification models often takes time and is performed by special applications such as the WordStat content analysis and text mining module. QDA Miner by itself, however does not provide the necessary tools to develop classification models. To do so one needs the WordStat add-on module. Once developed, these classification models can be saved to disk and used for classification tasks. QDA Miner can read those saved models and apply them to text segments (sentences, paragraphs, coded segments) or whole documents. Results of classification may then be used to automatically assign codes to text segments or values to existing or new categorical variables.

In order to be accessible by QDA Miner, WordStat classification models should be stored on disk under the Provalis Research\Models subfolder.

The automatic classification option in QDA Miner is available from the result table of the following text-retrieval tools:

- Text Retrieval
- Query by Example
- Section Retrieval
- Keyword Retrieval
- Coding Retrieval
- Code Sequence Analysis

It is available from those analysis dialog boxes in a form of an optional panel appearing at the bottom of the result table. This panel can be displayed or hidden by clicking the  button and looks like the dialog box below:

Classification model: Five domains (NB)  Min Ratio: 3

| Text                                                             | Predicted class | Ratio | Economy       | Education     | Politic       | Psychology | Sociology     |
|------------------------------------------------------------------|-----------------|-------|---------------|---------------|---------------|------------|---------------|
| Our poet laureate, Robert Pinsky, calls this burden of history   | Politic         | 89.1  | 0.0001        | 0.0111        | <b>0.9888</b> | 0.0000     | 0.0000        |
| We must show American power and purpose in strong support        | Politic         | 79.7  | 0.0000        | 0.0000        | <b>0.9876</b> | 0.0000     | 0.0124        |
| All of these facts must be squarely faced. China is a competitor | Economy         | 71.4  | <b>0.9799</b> | 0.0000        | 0.0063        | 0.0001     | 0.0137        |
| Likewise, international organizations can serve the cause of p   | Politic         | 69.9  | 0.0001        | 0.0141        | <b>0.9854</b> | 0.0000     | 0.0004        |
| We, after all, are the heirs of the heroes who launched the w    | Sociology       | 58.4  | 0.0013        | 0.0000        | 0.0168        | 0.0000     | <b>0.9819</b> |
| All our goals in Eurasia will depend on America strengthening i  | Politic         | 40.1  | 0.0000        | 0.0000        | <b>0.9757</b> | 0.0000     | 0.0243        |
| I remember what it was like seven years ago. Even the innov      | Economy         | 40.0  | <b>0.9572</b> | 0.0061        | 0.0239        | 0.0000     | 0.0128        |
| Friends, it is wonderful to come home to this state that holds   | Sociology       | 37.3  | 0.0000        | 0.0008        | 0.0261        | 0.0000     | <b>0.9731</b> |
| But if there is cause for apprehension over what lies ahead, t   | Politic         | 27.4  | 0.0351        | 0.0000        | <b>0.9604</b> | 0.0000     | 0.0045        |
| Together we can be that good.                                    | Economy         | 19.6  | <b>0.9334</b> | 0.0185        | 0.0476        | 0.0000     | 0.0005        |
| God bless the Citadel, and God bless the Class of '99.           | Sociology       | 19.1  | 0.0258        | 0.0463        | 0.0445        | 0.0005     | <b>0.8829</b> |
| Over the next ten years, Milosevic started four wars - each w    | Politic         | 19.0  | 0.0497        | 0.0000        | <b>0.9472</b> | 0.0000     | 0.0031        |
| These are my new proposals to spur innovation for the 21st C     | Education       | 16.1  | 0.0585        | <b>0.9414</b> | 0.0001        | 0.0000     | 0.0000        |

The **Classification Model** list box lists all WordStat classification models stored in the Models subfolder. The first step one needs to do in order to classify retrieved text segments or documents is to select a classification model. Once a model has been selected, the other panel options become available.

You may then click the  button to obtain information about the classification model developed by WordStat, such as the number of classes to be predicted, the algorithm used for prediction, etc..

To apply the selected model, click the  button. QDA Miner will take each entry in the result table above the classification panel and will apply the classification method to the associated text and will display the classification result in a table. The **Predicted Class** column indicates which class was identified as the most likely candidate for classifying this text. The **Ratio** command computes a ratio between the similarity score obtained for the predicted class and the second-best class. For example, if for a specific row, the ratio is equal to 8x, then we could say that the probability that this text belongs to the predicted class or that the similarity of this document to other documents from this predicted class is about 8 times higher than for documents in the second most likely class. If this value is low, such as 1.1x or 1.2x, we could say that the classified text is almost as likely to belong to the second class as the predicted class, and we may decide to ignore the classification results. All the other columns on the right of this Ratio column contain relative probability or similarity scores associated with each class. The score for the predicted class is shown in blue characters and a bold typeface.

By default, the table is sorted in descending order of the ratio statistic, so that classifications for which we would be the most confident are listed at the top, while the classifications that are more uncertain are located at the bottom of the list. To sort this table in ascending order on any other column values, simply click the column header. Clicking a second time on the same column header sorts the rows again in descending order.

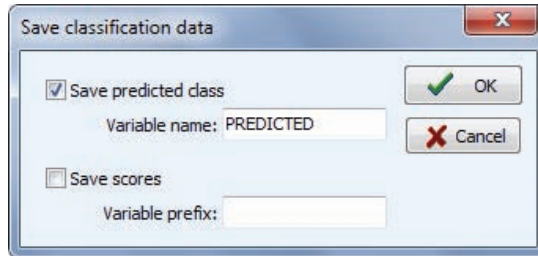
To remove a specific classification result, select its row and then click the  button.

Results of an automatic classification may be stored in the project in two different ways. If full documents are classified, one can store the value of the predicted class, and optionally, the probability or similarity score associated with each class into new numerical variables. If paragraphs or sentences are

classified, then predicted classes may be attached to the text segments as codes. If codes corresponding to the name of the classes do not already exist, QDA Miner will automatically add those to the codebook, before applying them.

### To store classification results as variables:

- Click the  button. A dialog box similar to this one will appear:



- To save the predicted class, click the check box beside the **Save Predicted Class** option and enter a variable name.
- To save the scores associated with each class and upon which the classification has been made, click the check box beside **Save Scores** and enter a variable prefix (up to 7 characters). Variable names are created by adding successive numeric values to this prefix. For example, if the edit box at the right of the Variable Prefix option is set to "CLASS\_", the variable names will be CLASS\_1, CLASS\_2, CLASS\_3 etc.
- If any one of the specified variables does not exist, QDA Miner will create new ones and store the numerical values associated with either the predicted class or the class scores. A confirmation dialog box will ask you to confirm the creation of those new variables as well as the overwriting of any existing ones.

### To store classification results as codings:

- Click the  button. QDA Miner will ask you to confirm the autocoding of all items in the table with a code corresponding to the predicted class. If you choose **Yes**, the program will browse through the entire project, select the corresponding text segment and assign the code corresponding to the predicted class. If any one of the codes is currently not in the codebook, a new category corresponding to the name of the classification model will be created, and the missing codes will be added to this new category.

### To export the classification table to disk:

- Click the  button. A Save File dialog box will appear.
- In the **Save As Type** list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file

(\*CSV), MS Word (\*.DOC), HTML file (\*.HTM; \*.HTML), XML files (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).

- Type a valid file name with the proper file extension.
- Click the **Save** button.

### To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.

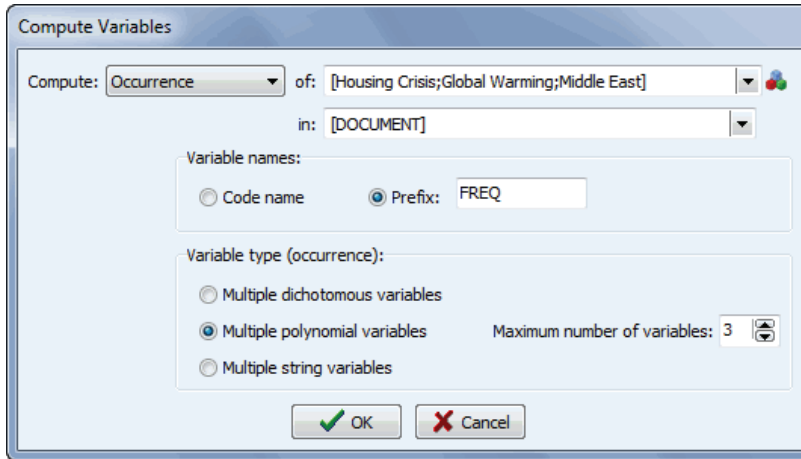
For more information on the Report Manager, see the **Report Manager** topic on page 230.

### To print the table:

- Click the  button.

# Storing Code Statistics into Numeric Variables

Most output tables created by QDA Miner may be exported to disk in various file formats including Excel, MS Word, HTML, XML, and delimited ASCII files. Code statistics may also be appended to the existing project file, allowing you to use those new variables for further numerical processing or case filtering. To append code statistics to your current project, select the **COMPUTE** command from the **VARIABLES** menu. The following dialog box will appear:



**COMPUTE** - This option allows you to specify which statistic calculated on selected codes will be stored in the numeric variables. Four statistics are available:

- Occurrence
- Frequency.
- Word Count
- Rate per 1000 words

**OF** - Select the codes on which to calculate the above statistic by clicking the arrow button at the right end of the list box and by selecting these codes. The number of variables calculated will be equal to the number of codes selected .

**IN** - This option allows you to specify on which document variables the calculation will be performed. If the current project contains more than one document variable, you will have a choice to select either one or more of them. By default, all document variables are selected. To restrict the computation to a few of them, click the down arrow key at the right of the list box. A list of all available document variables will appear. Select all the document variables that you want to analyze.

**VARIABLE NAMES** - This option sets what method should be used by QDA Miner to create new variable names. When set to **CODE NAME**, the program will attempt to use each code name as the name of a new variable. Illegal characters are automatically removed and long names are truncated to the first 10 characters. Duplicated variable names are distinguished by the

substitution of numerical digits at the end of the name. When this option is set to **PREFIX**, variable names are created by adding successive numeric values to a user-defined prefix. For example, if the edit box at the right of the prefix option is set to "FREQ", the variable names will be FREQ1, FREQ2, FREQ3, etc.

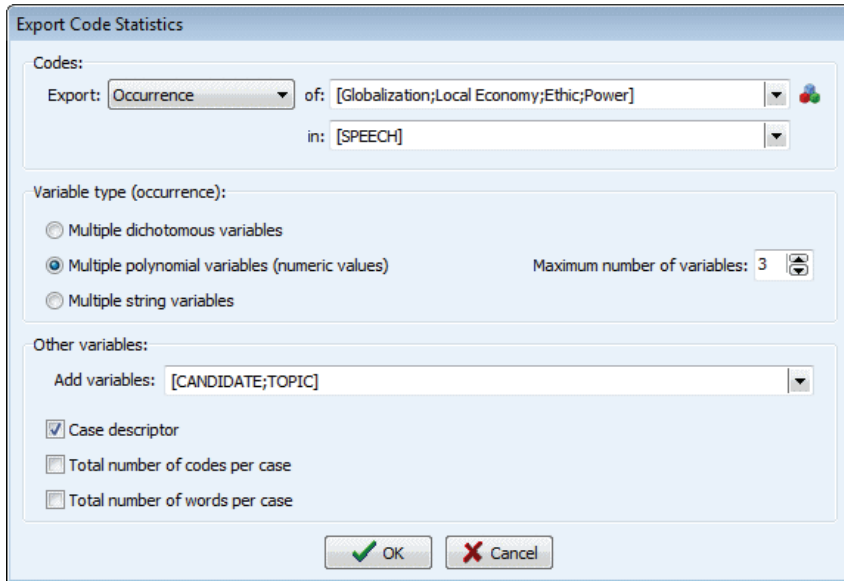
**VARIABLE TYPE** - By default, QDA Miner saves code statistics in as many variables as there are codes. For example, if your codebook contains 100 codes, then 100 variables will be necessary to store the statistics associated with each code, one code statistics per variable. When you choose to store the **occurrence** of codes, QDA Miner offers you the possibility to store the observed occurrences in a limited number of polynomial (or multinomial) variables. For example, if the maximum number of different codes per case is no more than 10, then you may instruct QDA Miner to create 10 nominal variables and store in each of those, a nominal value representing one of the codes. If less than 10 codes are found in a specific case, then the remaining variables are left empty. To store values in a limited set of nominal variables, choose the **Multiple Polynomial Variables** option and enter **the Maximum Number of Variables** that should be used for storing the code values. To store code names rather than nominal values, select **Multiple String Variables** instead. If the maximum number of codes found in a single case is higher than the specified number of variables, then a warning message will appear to let you know that some code occurrence information have been lost and to indicate what was the maximum number of codes encountered in the project. To store occurrences as zeros and ones in as many variables as there are codes, selecting the **Multiple Dichotomous Variables** option.

When you click the **OK** button, the program prompts you to confirm the calculation of the required number of variables. If variables with similar names exist in the project, the confirmation dialog box will display the number of existing variables that would be overwritten as well as the number of variables that need to be created by this command. Click the **YES** button to confirm the creation and overwriting of variables. Click **NO** to abort the procedure.

To exit from the dialog box without performing the recoding, click the **CANCEL** button.

# Exporting Code Statistics

Various code statistics may be exported to disk in different file formats including Excel, HTML, XML and delimited ASCII files, allowing those data to be further analyzed using a statistical program. To export code statistics to disk, select the EXPORT | CODE STATISTICS command sequence from the PROJECT menu. The following dialog box will appear:



The dialog box is titled "Export Code Statistics". It contains three main sections. The first section, "Codes:", has three fields: "Export:" with a dropdown menu set to "Occurrence", "of:" with a text box containing "[Globalization;Local Economy;Ethic;Power]" and a small icon, and "in:" with a dropdown menu set to "[SPEECH]". The second section, "Variable type (occurrence):", has three radio buttons: "Multiple dichotomous variables", "Multiple polynomial variables (numeric values)" (which is selected), and "Multiple string variables". To the right of these is a label "Maximum number of variables:" followed by a text box containing "3" and a small icon. The third section, "Other variables:", has a label "Add variables:" followed by a text box containing "[CANDIDATE;TOPIC]" and a small icon. Below this are three checkboxes: "Case descriptor" (checked), "Total number of codes per case" (unchecked), and "Total number of words per case" (unchecked). At the bottom right are two buttons: "OK" with a green checkmark icon and "Cancel" with a red X icon.

**EXPORT** - This option allows you to specify which statistic calculated on selected codes will be stored into numeric variables and exported. Four statistics are available:

- Occurrence
- Frequency
- Word Count
- Percentage of Words

**OF** - Select the codes on which to calculate the above statistic by clicking the arrow button at the right end of the list box and by selecting these codes. The number of variables calculated will be equal to the number of codes selected.

**IN** - This option allows you to specify on which document variables the calculation will be performed. If the current project contains more than one document variable, you will have a choice to select either one or more of them. By default, all document variables are selected. To restrict the computation to a few of them, click the down arrow key at the right of the list box. A list of all available document variables will appear. Select all the document variables that you want to analyze.

**VARIABLE TYPE** - By default, QDA Miner saves code statistics in as many variables as there are codes. For example, if your codebook contains 100 codes, then 100 variables will be necessary to store the statistics associated with each code, on code statistics per variable. When you choose to store the **occurrence** of codes, QDA Miner offers you the possibility to store the observed occurrences in a limited number of polynomial (or multinomial) variables. For example, if the maximum number of different codes per case is no more than 10, then you may instruct QDA Miner to create 10 numeric variables and store in each of those, a numeric value representing one of the codes. If less than 10 codes are found in a specific case, then the remaining variables are left empty. To store values in a limited set of nominal variables, choose **the Multiple Polynomial Variables** option and enter the **Maximum Number of Variables** that should be used for storing the code values. To store code names rather than numerical values, select **Multiple String Variables** instead. If the maximum number of codes found in a single case is higher than the specified number of variables, then a warning message will appear to let you know that some code occurrence information have been lost and to indicate what was the maximum number of codes encountered in the project. To export occurrences as zeros and ones in as many variables as there are codes, selecting the **Multiple Dichotomous Variables** option.

**ADD VARIABLES** - This drop-down checklist box may optionally be used to add the values stored in one or more variables to the exported data file along with the code statistics.

**CASE DESCRIPTOR** - Enabling this option insert in the exported data file a string variable containing the case descriptor defined by the user. For information on how to create a case descriptor see page 57.

**TOTAL NUMBER OF CODES PER CASE** - This options appends to each exported case, the total number of codes encountered in all selected documents.

**TOTAL NUMBER OF WORDS PER CASE** - This options appends to each exported case, the total number of words encountered in all selected documents.

Once the options have be set up, click the **OK** button. A File Save dialog box will appear. Select the file format you want to create using the **Save As Type** drop-down list, enter a valid filename with the proper file extension and then click the **Save** button.

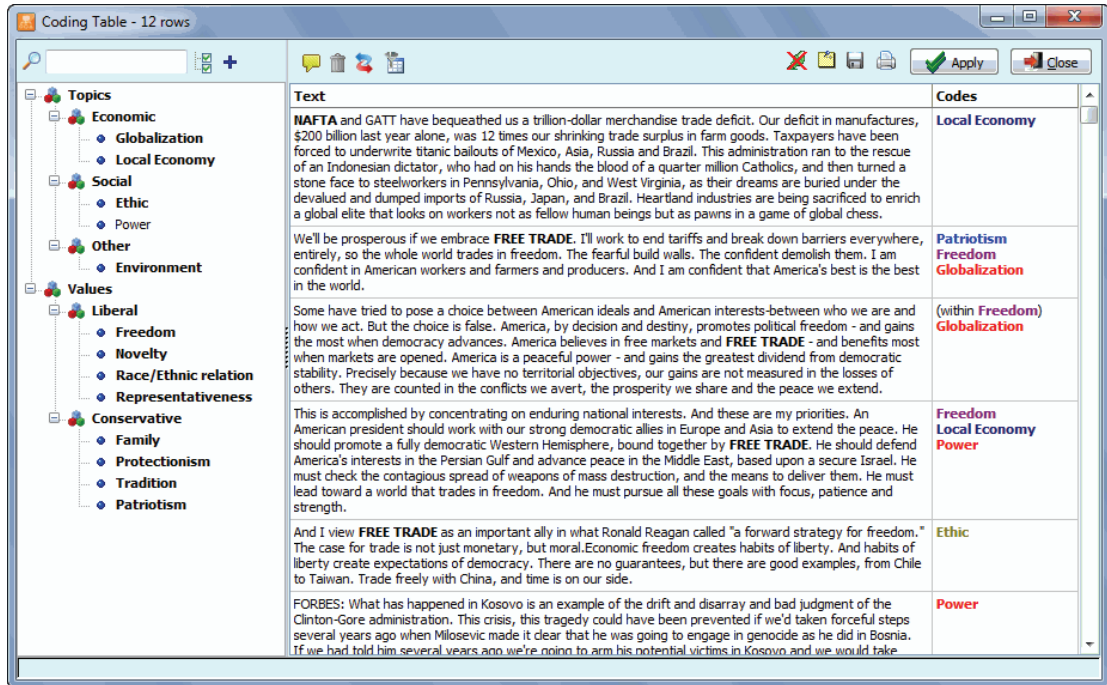
# Exporting Co-occurrence or Similarity Matrices

Matrices of co-occurrences or computed similarities may be exported to disk in Excel, MS Word, HTML, XML, and delimited ASCII files. To export such code co-occurrence as well as code or case similarity matrix:

- Select the CODING CO-OCCURRENCES command from the ANALYSIS menu.
- Set the Co-occurrence options on the first page of the dialog box (see **Coding Co-Occurrences** on page 166 for information on those options)
- Move to the STATISTICS page.
- Click the  button and select CO-OCCURRENCE MATRIX if you want to export the cross frequencies of codes, or SIMILARITY MATRIX to export the computed code or case similarity measures.
- Select the file format you want to create using the **Save As Type** drop-down list.
- Enter a valid filename with the proper file extension.
- Click the **Save** button.

# Coding Text Segments Using the Coding Table

The Coding Table view provides an efficient way of assigning different codes to specific items in a list of extracted text segments. It is especially useful when coding large amounts of short text segments like responses to an open-ended question or a long list of customer feedback responses. This dialog box can be accessed from several text retrieval and coding retrieval tools in QDA Miner by clicking the  button.



The dialog box shows the codebook on the left and a table with two columns to the right. The first column contains the text segments to be coded, one segment per row, and the second column contains the codes currently assigned to them. Codings that overlap – including, or enclosed in, text segments – cannot be modified or removed and are displayed between parentheses.

Clicking down the  button forces QDA Miner to synchronize the table with the main window so that moving from one row to another will cause the main window to display the corresponding case and document from which the selected text comes. The corresponding text in the original document will also be highlighted allowing one to easily view the full context of the selected text segment.

The assignment of codes to text segments can be performed in several ways, using either drag-and-drop operations or by clicking check marks beside codes in the codebook.

## To assign a code using drag-and-drop operations

- Click on the code you would like to assign, keeping the mouse button down.
- Drag the code over the row you would like to assign the code to and release the mouse button.

or

- Click the row you would like to code in order to select it and keep the mouse button down.
- Drag the row over the code you would like to assign and release the mouse.

While dragging, if a code cannot be assigned to a specific row, the mouse cursor will be displayed as a , indicating that the dragged item cannot be dropped at this location.

## To assign or remove codes using check marks

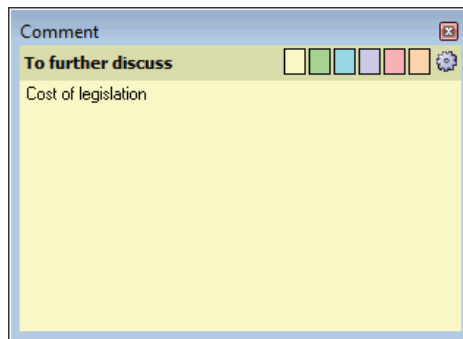
- Depress the  button to display check boxes.
- Select the text segment you would like to code.
- Click check marks beside codes you would like to assign to the selected row or clear any existing check marks beside codes you would like to remove.

## To remove a code assigned to a segment:

- Click the  button located on the toolbar. A list of all codes assigned to the current row will be displayed.
- Select the code that you would like to remove.

## To attach a comment to a coded segment or edit an existing one:

- Click the  button.
- Select the code to which you want to attach a comment. A small window like the one below will appear.



- Enter the text that you want to associate with the selected code or edit the existing text.

- Click the **X** button in the upper-right corner to save the comment and close the window.

When a note has been assigned to a coded segment, an exclamation point between square brackets is displayed on the left of the code name. To edit an existing comment, follow the above steps. To remove a comment, simply open the **Note Editor** and delete all text in the editing window.

### To remove a row from the list:

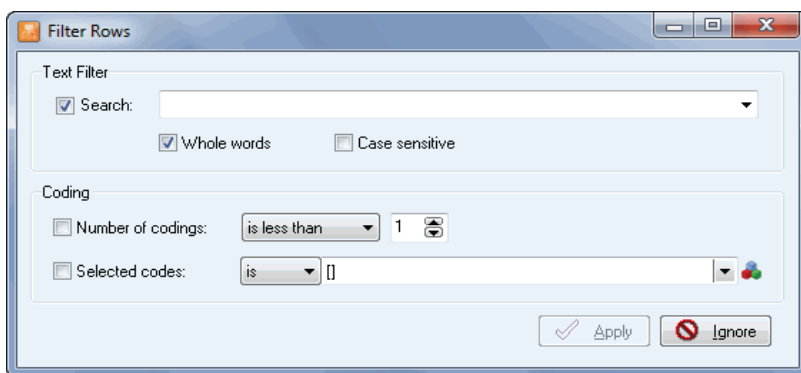
- Select its row and then click the  button.

### To add a new code to the codebook

- Click the  button.

### To filter coded segments

- Click the  button. A dialog box similar to this one will appear:



- To filter rows using a text-search expression, put a check mark beside the **Search** option and enter your search expression. The search filter supports the same syntax as the **Text Retrieval** tool and may contain Boolean, wildcards and parentheses.
- To filter row based on the number of codings assigned to text segments, put a check mark beside the **Number of Codings** option and set the associated operator and value.
- To filter rows containing specific codes, put a check mark beside the **Selected Codes** option and specify a list of codes to search for. When more than one code is specified, the table will show every row containing at least one of these codes.

### To export the table to disk:

- Click the  button. A **Save File** dialog box will appear.
- In the **Save As Type** list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file (\*.CSV), MS Word (\*.DOC), HTML file (\*.HTM; \*.HTML), XML files (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).
- Type a valid file name with the proper file extension.
- Click the **SAVE** button.

### To append a single row from the table:

- Select the row containing the text segment you want to append to the Report Manager and click the  button.

for more information on the Report Manager, see the **Report Manager Features** on page 230.

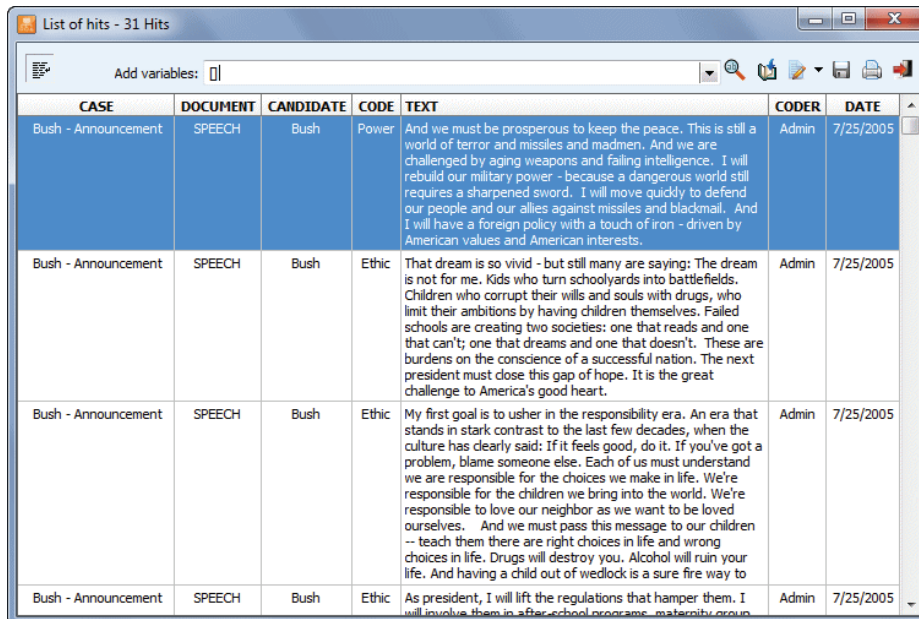
### To print the table:

- Click the  button.

Changes made using the coding table are not implemented immediately but are kept in memory. To implement all changes made, click the  **Apply** button. It is recommended to click this button regularly after making several changes. To close the dialog box and cancel all changes, click the  **Close** button.

# Using the Retrieved Segments Dialog Box

The Retrieved Segments dialog box is displayed when one searches coded segments from joint frequency or co-occurrence tables as well as from various charting tools (such as heatmaps, bubble charts, bar charts, dendrograms, proximity plots). It lists all coded segments associated with the last retrieval operation in a table that may be customized, saved or printed. The table provides basic information on each hit, including the case and the document variable in which it was found, the code name, the coder's name and coding date as well as the full text of the coded segment. When searching for co-occurrence, the texts of both coded segments are displayed in separate columns (see below).



| CASE                | DOCUMENT | CANDIDATE | CODE  | TEXT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | CODER | DATE      |
|---------------------|----------|-----------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-----------|
| Bush - Announcement | SPEECH   | Bush      | Power | And we must be prosperous to keep the peace. This is still a world of terror and missiles and madmen. And we are challenged by aging weapons and failing intelligence. I will rebuild our military power - because a dangerous world still requires a sharpened sword. I will move quickly to defend our people and our allies against missiles and blackmail. And I will have a foreign policy with a touch of iron - driven by American values and American interests.                                                                                                                                                                                                            | Admin | 7/25/2005 |
| Bush - Announcement | SPEECH   | Bush      | Ethic | That dream is so vivid - but still many are saying: The dream is not for me. Kids who turn schoolyards into battlefields. Children who corrupt their wills and souls with drugs, who limit their ambitions by having children themselves. Failed schools are creating two societies: one that reads and one that can't; one that dreams and one that doesn't. These are burdens on the conscience of a successful nation. The next president must close this gap of hope. It is the great challenge to America's good heart.                                                                                                                                                        | Admin | 7/25/2005 |
| Bush - Announcement | SPEECH   | Bush      | Ethic | My first goal is to usher in the responsibility era. An era that stands in stark contrast to the last few decades, when the culture has clearly said: If it feels good, do it. If you've got a problem, blame someone else. Each of us must understand we are responsible for the choices we make in life. We're responsible for the children we bring into the world. We're responsible to love our neighbor as we want to be loved ourselves. And we must pass this message to our children -- teach them there are right choices in life and wrong choices in life. Drugs will destroy you. Alcohol will ruin your life. And having a child out of wedlock is a sure fire way to | Admin | 7/25/2005 |
| Bush - Announcement | SPEECH   | Bush      | Ethic | As president, I will lift the regulations that hamper them. I will involve them in after-school programs, maternity group                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Admin | 7/25/2005 |

## To sort the list of hits:

- To sort in ascending order any column values, simply click this column header. Clicking a second time on the same column header sorts the rows in descending order.

## To add variables to the table:

- Use the **Add Variables** drop-down checklist box to add the values stored in one or more variable to the table of retrieved segments for the specific case from which a text segment originates.

## To analyze the text of all retrieved segments using WordStat:

- Click the  button. QDA Miner will minimize itself and run WordStat, allowing you to perform a quantitative content analysis or apply text-mining techniques on the retrieved segments. WordStat will also allow you to perform keyword comparisons across any other column in the table. To include additional data for comparison analysis, use the Add Variables option.

### To append a copy of the table in the Report Manager:

- Click the  button. A descriptive title will be provided automatically for the table. To edit this title or to enter a new one, hold down the SHIFT keyboard key while clicking this button.

For more information on the **Report Manager**, see page 230.

### To create a text report of retrieved segments:

- Click the  button. The sort order of the current table is used to determine the display order in the report. This report is displayed in a text editor dialog box and may be modified, stored on disk in RTF, HTML or plain text format, printed, or cut and pasted into another application. Graphics and tables may also be inserted anywhere in this report.

### To export the table to disk:

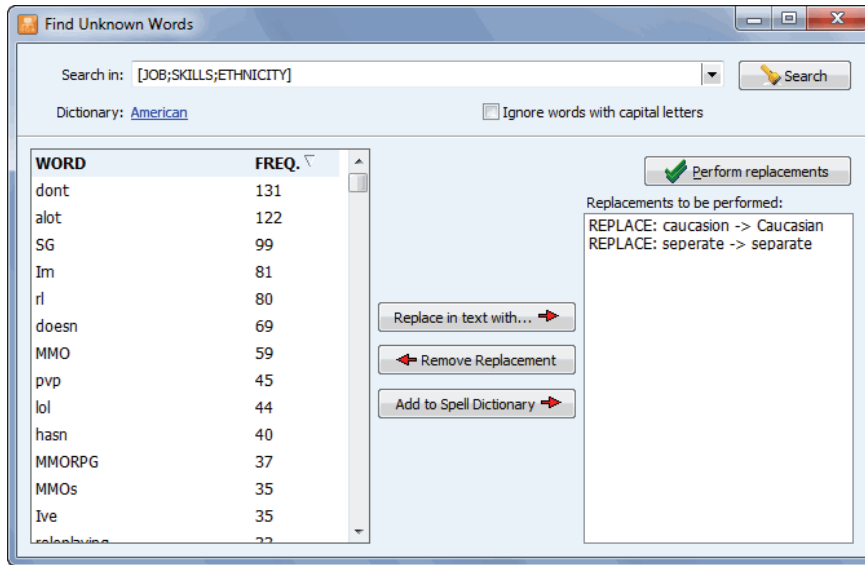
- Click the  button. A Save File dialog box will appear.
- In the Save As Type list box, select the file format under which to save the table. The following formats are supported: ASCII file (\*.TXT), Tab delimited file (\*.TAB), Comma delimited file (\*.CSV), MS Word (\*.DOC), HTML file (\*.HTM; \*.HTML), XML files (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).
- Type a valid file name with the proper file extension.
- Click the SAVE button.

### To print the table:

- Click the  button.

# Identifying and Substituting Unknown Words

The Unknown Words feature of QDA Miner provides a way to quickly identify misspelled words by comparing the list of word forms encountered in the entire text collection or in the currently selected cases against a list of common words. By default, the extraction is performed in reference to common English words. To identify unknown words in documents written in another language, click the **Dictionary** link.

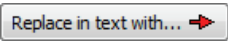


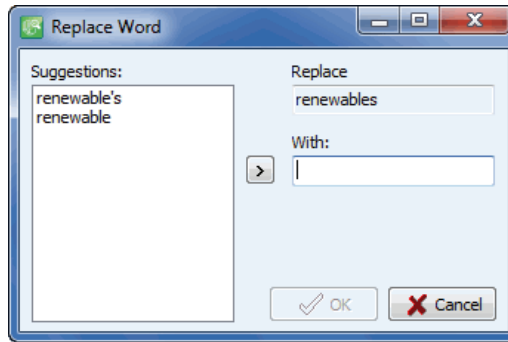
Once the various options have been set, click the  **Search** button to start searching for vocabulary words. The list of words retrieved are then listed in a frequency table on the left of the screen and presented in descending order of frequency. To sort this list in alphabetical order, click the top of the first column.

Two types of operations are allowed on these words: 1) You can replace all instances of a selected word in the original document by another word or phrase; or 2) you can add this word to a custom list of valid words causing the program to ignore those words the next time there is a search for vocabulary words

Replacing words in the original documents or adding items to the list of valid words that should, from now on, be ignored are not performed immediately. Instead, they are added to an action list allowing you to review, modify or cancel previously defined actions prior to the application of all the specified changes.

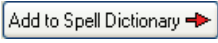
## To replace words in the original documents:

- Select the word to be replaced.
- Click the  **Replace in text with...** button. A dialog box similar to this one will appear:



- Enter the new replacement word or phrase or choose from the **Suggestions** list box on the left side of the dialog box.
- Click the **OK** button to confirm this replacement and add this operation to the list of actions to perform.

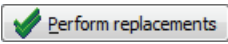
### To add a word to the custom list of words to ignore:

- Select the word you would like to add to the custom dictionary.
- Click the  button.

### To remove operations previously defined:

- Select the operations that you would like to remove.
- Click the  button. All words associated with the removed actions are moved back to the list of unknown words and positioned at the bottom of the list.

### To perform all the defined word replacements:

- Click the  button.

# Managing Outputs with the Report Manager

The Report Manager is a separate application that has been designed to store, edit and organize documents, notes, quotes, tables of results, graphics and images created by QDA Miner or imported from other applications. Items can be added to the Report Manager directly from QDA Miner without needing to run the Report Manager.

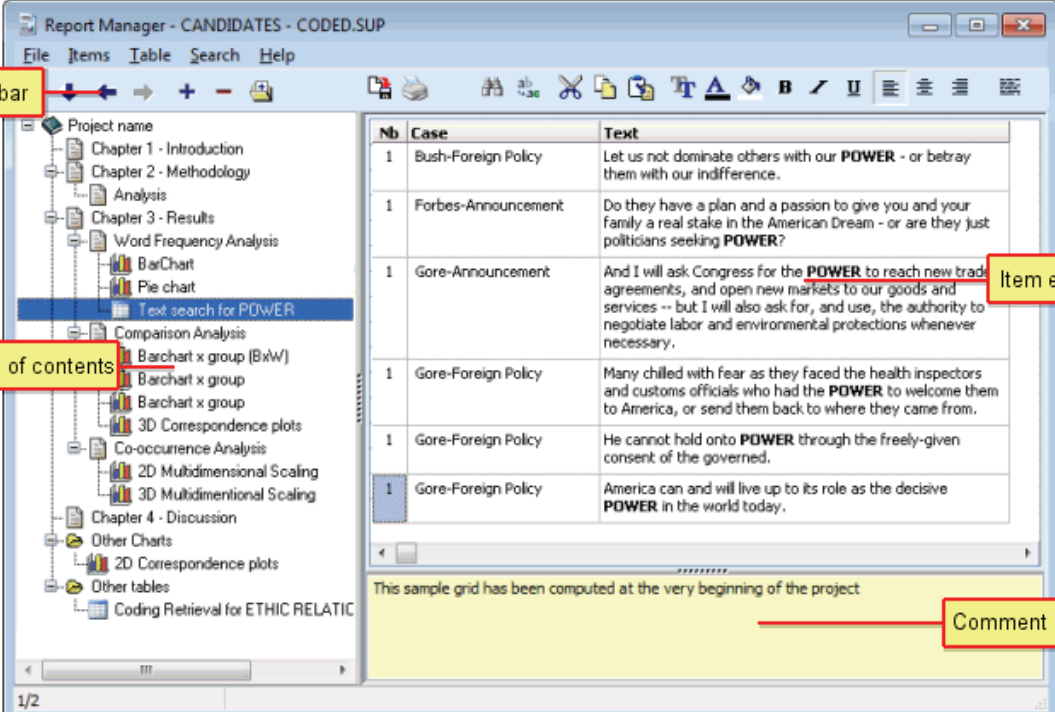
The  button, found in many locations in QDA Miner, may be used to copy entire documents, tables and charts to the Report Manager.

Selected text segments or image areas may also be appended by clicking the  button.

To access the Report Manager from QDA Miner, run the REPORT MANAGER command from the PROJECT menu.

The program presents its information as an outline, allowing a hierarchical organization of miscellaneous pieces of information that is ideal for project management, organizing ideas, structuring information, or designing and writing a research report.

The workspace emulates the appearance of Windows Explorer or of a standard Help file with the Table of Contents (TOC) on the left and the Editor on the right.



The screenshot shows the 'Report Manager - CANDIDATES - CODED.SUP' window. The left pane displays a hierarchical 'Table of contents' with items like 'Chapter 1 - Introduction', 'Chapter 2 - Methodology', 'Chapter 3 - Results', and 'Chapter 4 - Discussion'. The right pane shows an 'Item editor' with a table of results. The table has columns 'Nb', 'Case', and 'Text'. The bottom of the right pane features a 'Comment editor' with a text area containing a sample grid comment.

| Nb | Case                | Text                                                                                                                                                                                                                                               |
|----|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Bush-Foreign Policy | Let us not dominate others with our <b>POWER</b> - or betray them with our indifference.                                                                                                                                                           |
| 1  | Forbes-Announcement | Do they have a plan and a passion to give you and your family a real stake in the American Dream - or are they just politicians seeking <b>POWER</b> ?                                                                                             |
| 1  | Gore-Announcement   | And I will ask Congress for the <b>POWER</b> to reach new trade agreements, and open new markets to our goods and services -- but I will also ask for, and use, the authority to negotiate labor and environmental protections whenever necessary. |
| 1  | Gore-Foreign Policy | Many chilled with fear as they faced the health inspectors and customs officials who had the <b>POWER</b> to welcome them to America, or send them back to where they came from.                                                                   |
| 1  | Gore-Foreign Policy | He cannot hold onto <b>POWER</b> through the freely-given consent of the governed.                                                                                                                                                                 |
| 1  | Gore-Foreign Policy | America can and will live up to its role as the decisive <b>POWER</b> in the world today.                                                                                                                                                          |

This sample grid has been computed at the very beginning of the project

## Table of Contents panel

Report Manager files are made up of items or topics that are like chapters in a book. Each item can be thought of as a separate word processor, a table or a graphic file editor or viewer, all of which are stored together in the QDA Miner project file. This panel provides powerful functions for organizing items and structuring the information in a hierarchical manner.

## Item Editor

The largest panel on the right of the program window is the Item Editor, which is like a built-in word processor. This is where the item selected in the Table of Contents can be edited. Clicking a Table of Contents item displays its contents for editing.

## Toolbar

The Toolbar provides quick access to the most frequently-used functions. Just position the mouse over a tool button and wait for the display of a brief text describing its function.

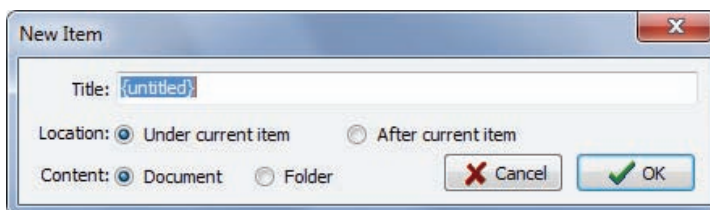
## Comment panel

The Comment panel below the Item Editor allows the insertion and editing of comments related to the selected topic. When new items are added to the Report Manager from QDA Miner, a default comment is often already present, providing useful information about the origin of this item.

## Working with the Table of Contents

### To create a new item:

- Select the Table of Contents entry that will be the "parent" or "sibling" of the new item.
- Select the NEW command from the ITEMS menu or click the  button. A dialog box like this one will appear:



- Enter the title for the new item.
- If the new item should be a "child" of the selected item, click **Under Current Item**; if the item should be positioned after the current item, then set it to **After Current Item**.
- Select whether the new item will be a **Document** or a **Folder**. Folders are empty items that are used as containers for other items.
- Click **OK**. The new item will become the current one.

## To import items from files:

- Select the item under which the imported items will be stored.
- Select the IMPORT FILES command from the ITEMS menu or click the  toolbar button. An Open dialog box will appear.
- Select the type of data you would like to import by selecting the appropriate Files of Type list box option. The Report Manager can import the following data types:
  - DOCUMENTS - Plain text (.TXT), MS Word (.DOC), WordPerfect (.WPD), Rich Text (.RTF) or HTML files (.HTM or .HTML)
  - GRAPHICS - Windows Bitmap (.BMP), Windows Metafile (.WMF), JPEG files (.JPG or .JPEG) and Portable Network Graphic files (.PNG)
  - CHARTS - QDA Miner or WordStat Charts (.WSX)
  - DELIMITED DATA - Tab delimited (.TAB) or Comma Separated Value (.CSV) data files.
- Select one or several files to be imported and click the OPEN button.

## To rename an item:

- Select the item to be renamed.
- Select the RENAME command from the ITEMS menu or click the  toolbar button.
- In the Item Title Dialog, change the title.
- Click **OK**.

## To delete an item:

- Select the item to delete in the Table on Contents.
- Select the DELETE command from the FILE menu or click the  toolbar button.
- You will be asked to confirm that you really want to delete the item. If you're sure, then click **Yes**.

NOTE: Be aware that you cannot undo this if you make a mistake.

## Moving Items

As more items are created and the Report Manager hierarchy grows, it is inevitable that you will want to move items around, either to place one item under another, or to promote one to a higher level.

The easiest way to move items in the Table of Contents is by using drag-and-drop operations. Using the mouse, you can move an item to a different location or move a group of items stored under a "parent" item by dragging this "parent" item to its new location.

- Select the item to move by clicking and holding down the left mouse button. (Keep the mouse pressed until the drag-and-drop operation is completed.)
- Drag the item to its new location and, only then, release the mouse button.
- The dragged item will now become a "child" of the destination item.
- To move the item to the same level as the item under the cursor, simply hold the ALT key while dropping the dragged item.

You can also use menu commands and toolbar buttons to move items. To promote an item is to move it to a higher level in the hierarchy, making the item a "sibling" to its former "parent". To demote an item is to move it to a lower level and make it a "child" of its previous "sibling". After selecting the item you want to move, use one of the following four commands:

- To promote the selected item, click the  button, or select the PROMOTE command from the ITEMS menu.
- To demote the selected item, click the  button, or select the DEMOTE command from the ITEMS menu.
- To move the selected item up relative to its siblings, click the  button or select the MOVE UP command from the ITEMS menu.
- To move an item down relative to its siblings, click the  button or select the MOVE DOWN command from the ITEMS menu.

## Adding or Editing Item Comments

The Comment panel below the Item Editor allows one to insert a new comment or edit an existing one related to a selected topic.

To type a new comment or to edit one, simply click in the yellow region of this panel and start to type. The comment is automatically saved as soon as you move to another item or leave the Comment panel. While there is no menu item or toolbar icon associated with this feature, standard clipboard operations are supported for copy and pasting text. A popup menu with standard editing features can also be obtained by clicking the right mouse button.

To search for text in comments, see the information on the **Global Search** command.

## Editing Documents

The Report Manager offers many editing features to create and edit both simple text documents and documents with complex formatting, as well as tables and graphics. When a document item is selected in the Table of Contents, a DOCUMENT menu appears, displaying all available formatting and editing options. A similar menu can also be obtained by right-clicking anywhere in this document. The toolbar portion directly over the editing area also displays buttons to access the most often-used editing and formatting functions.

Individual documents may also be printed or exported to disk in various file formats such as plain text, Rich Text or HTML format. An IMPORT command is also available to read a document file stored in plain text, Rich Text, MS Word, WordPerfect, HTML and a few additional formats. Executing such a command will replace the existing content with the content of the imported file.

## Editing Tables

The Report Manager offers many editing features to customize the appearance of a table, to change the text alignment or font setting, to set the cell background color, or to delete entire rows or columns. When a table item is selected in the Table of Contents, a TABLE menu become visible displaying all available formatting and editing options. A similar menu can also be obtained by right-clicking anywhere in this table. The toolbar portion above the editing area also displays buttons to access the most often-used table editing and formatting functions.

Individual tables may also be printed or exported to disk in various file formats such as ASCII (\*.TXT), tab delimited file (\*.TAB), comma delimited (\*.CSV), MS Word (\*.DOC), HTML (\*.HTM; \*.HTML), XML (\*.XML), Excel spreadsheet file (\*.XLS; \*.XLSX), and SPSS data file (\*.SAV).

## Editing Charts

Many charts saved in the Report Manager may be edited using many of the same options as those available in QDA Miner, such as the multidimensional scaling plot obtained through the CODE CO-OCCURRENCE analysis command, the correspondence analysis plots, and the bar charts and line charts created by the CODING BY VARIABLES command, as well as the bar charts and pie charts produced by the CODING FREQUENCY command. To obtain information on the display options available for those charts, see their corresponding page in this manual. Other charts - such as dendrograms and heatmaps - are stored as image files, so cannot be modified. However, just like other charts, they may be exported to disk in various file formats such BMP, JPG or PNG graphic files.

## Searching and Replacing Text

Two broad types of text search are available in the Report Manager. A local, item-based search-and-replace feature allows one to perform text searches and replacements on individual documents or tables, and a global search engine for searching text patterns in several or all documents, tables and comments in the Report Manager.

### To perform a search in a single document or a table:

- Select the document or table you would like to search, by selecting its entry in the Table of Contents.
- Position the editing cursor in the document or select the cell in the table where you want the search to begin.
- Select the FIND command from the SEARCH menu, or click the  button.

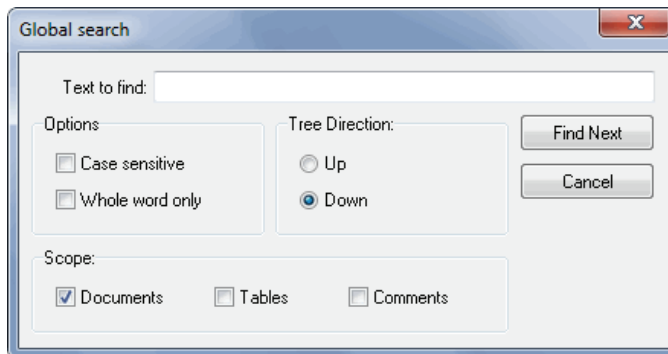
- Enter a search expression, set the desired search options and then click **Find Next**.
- To find additional instances of the same text, continue to click **Find Next**.

### To replace text in a single document or table:

- Select the document or table in which you would like to perform the text replacement by selecting its entry in the Table of Contents.
- Position the editing cursor in the document or select the cell in the table where you want the search to begin.
- Select the REPLACE command from the SEARCH menu, or click the  button.
- In **Find What**, type the characters or words you want to find. In **Replace With**, type the text you want to replace it with. Set the desired search options and then click **Find Next**. Click **Replace** to change the selected text. To replace all instances of the text, click **Replace All**.

### To perform a global search:

- Select the GLOBAL FIND method from the SEARCH menu. A dialog box similar to this one will appear:



The **Text to Find** edit box allows you to specify the text you want to find. The **Case Sensitive** and **Whole Word Only** options function in the same way as in a standard word processor.

The search starts at the current topic item. Select **Forward** to continue searching items below the current one, or **Backward** to move up and search items above the current document or table in reverse order. To search all items, select the top item in the Table of Contents before using the Global Search dialog box.

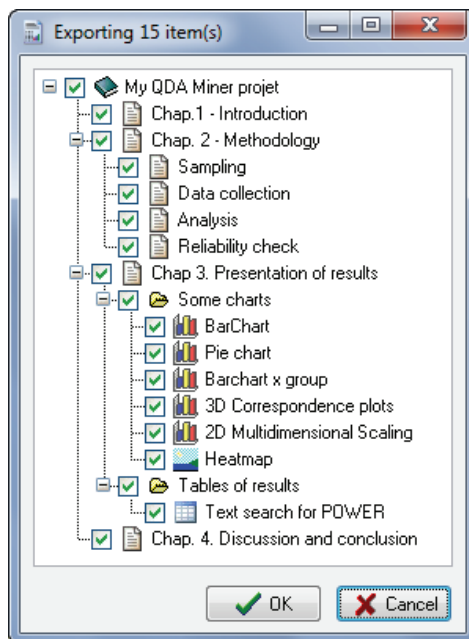
The **Scope** option box is used to specify what is to be searched. You can restrict the search to **Documents**, **Tables**, or **Comments** attached to items, or any combination of these three.

Once the search options have been set, click the **Find** button to start the search as well as to continue searching for additional instances of this text.

## Exporting items to HTML or Word

Individual documents, tables, graphics and images may be exported to disk in numerous formats. Such exportation can be achieved by clicking the  toolbar button or by selecting the EXPORT command from the associated menu.

The Report Manager also offers the possibility of exporting the entire content or selected items into a single HTML or MS Word document. The exportation is achieved by selecting the proper command from the FILE | EXPORT menu. For example, to export items to HTML, select the HTML command. A dialog box similar to this one will appear.



By default, all items are marked for export. To prevent some items from being included in the exported file, simply remove the check marks beside them. Clicking a "parent" item affects all "children" items in the same way. To unselect all items, uncheck the project item located at the very top of the tree.

Once the selection process is completed, click the **OK** button. A Save File dialog box will be displayed, allowing you to enter a file name and select the location where the file should be saved. After the file is created, you will be asked if you want to view this file. Clicking **Yes** will open a web browser if the exported file is an HTML document or either MS Word or Wordpad if the exported file is a Word document.

# References

## Selected References on Qualitative Data Analysis

- DEY, I (1993). *Qualitative Data Analysis: A User-Friendly Guide for Social Scientists*. Routledge.
- KRIPPENDORFF, K. (2004). *Content Analysis: An Introduction to Its Methodology*. 2nd ed. Thousand Oaks, CA: Sage.
- MILES, M. B. & HUBERMAN, M. (1994). *Qualitative Data Analysis : An Expanded Sourcebook*. Thousand Oaks, CA: Sage.

## Inter-coders Agreement Statistics

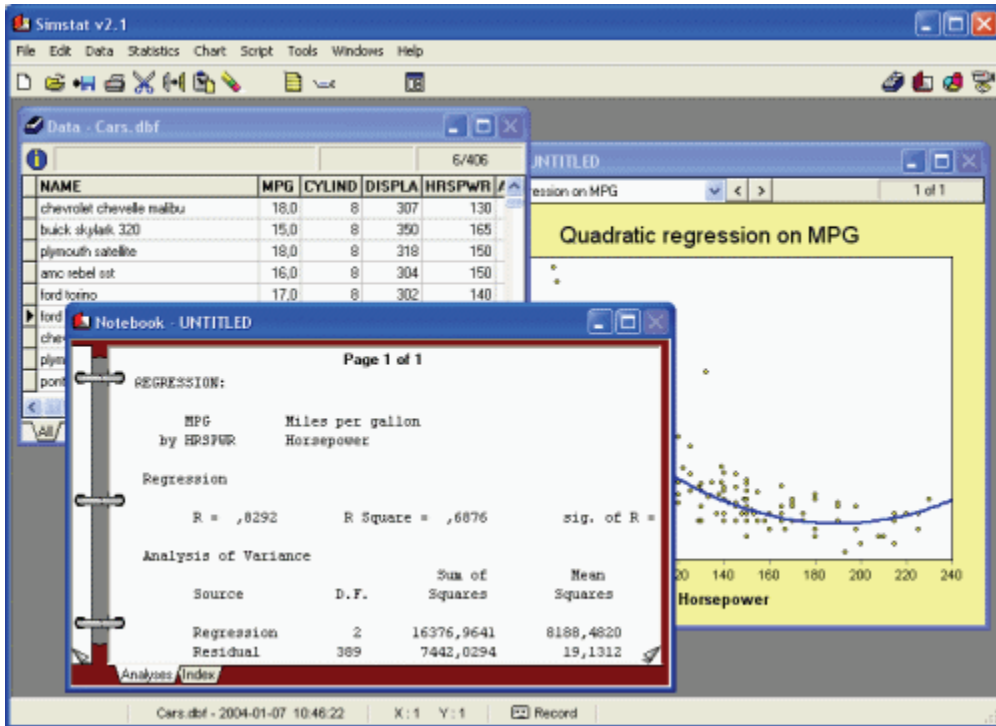
- BENNETT, E.M., ALPERT, R., & GOLDSTEIN, A.C. (1954). Communications through limited response questioning. *Public Opinion Quarterly*, 19, 303-308.
- BRENNAN, R.L., & PREDIGER, D. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41, 687-699.
- COHEN, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological measurement*, 20, 37-46.
- JASON, S., & VEGELIUS, J. (1979). On generalizations of the G index and the phi coefficient to nominal scales. *Multivariate Behavioral Research*, 14, 255-269.
- KRIPPENDORFF, K. (1970). Bivariate agreement coefficients for reliability of data. In E.F. Borgatta and G.W. Bohrnstedt (Eds.). *Sociological methodology: 1970*. San Francisco: Jossey-Bass.
- SCOTT, W.A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, 19, 321-325.
- ZWICK, R. (1988). Another look at interrater agreement. *Psychological Bulletin*, 103 (3), 374-378.

## Multivariate Analysis

- GREENACRE, M. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press. Orlando, Florida.

# Information on Simstat

Simstat for Windows is a general statistical program. It offers a wide range of statistical and graphical tools, intuitive output management features as well as its own scripting language to automate statistical analysis, write small applications, interactive tutorials, etc.



For more information, visit Provalis Research web site at:

<http://www.provalisresearch.com>

# Information on WordStat

WordStat is a text analysis module specifically designed to study textual information such as responses to open-ended questions, interviews, titles, journal articles, public speeches, electronic communications, etc. WordStat may be used for automatic categorization of text using a dictionary approach or various text mining procedures as well as for manual coding. WordStat can apply existing categorization dictionaries to a new text corpus. It also may be used in the development and validation of new categorization dictionaries. When used in conjunction with manual coding, this module can provide assistance for a more systematic application of coding rules, help uncover differences in word usage between subgroups of individuals, assist in the revision of existing coding using KWIC (Keyword-In-Context) tables, and assess the reliability of coding by the computation of inter-raters agreement statistics.

WordStat includes numerous exploratory data analysis and graphical tools that may be used to explore the relationship between the content of documents and information stored in categorical or numeric variables such as the gender or age of respondents, year of publication, etc. Relationships among words or categories as well as document similarity may be identified using hierarchical clustering and multidimensional scaling analysis. Correspondence analysis and heatmap plots may be used to explore relationships between keywords and different groups of individuals.

For more information, visit Provalis Research web site at:

<http://www.provalisresearch.com>

# Technical Support

If you encounter any problems or have any comments or suggestions for further improvement, please contact Provalis Research:

**By Phone:** 514-899-1672

**By FAX:** 514-899-1750

**By Email:** [support@provalisresearch.com](mailto:support@provalisresearch.com)

**Web site:** [www.provalisresearch.com](http://www.provalisresearch.com)

**Mail:** Provalis Research  
2997 Cedar Avenue  
Montreal, QC  
H3Y 1Y8  
Canada