

» Pourquoi faut-il utiliser à la fois QDA Miner et WordStat ?

Il existe de nombreuses façons de combiner les fonctions de codage qualitatif de QDA Miner et les outils d'analyse de contenu et d'extraction de texte disponibles dans WordStat. Même ceux qui privilégient un codage manuel trouveront les fonctionnalités d'analyse exploratoire de WordStat utiles pour se familiariser avec l'éventail des sujets abordés dans l'ensemble du corpus à analyser et pour identifier non seulement les codes que l'on pourrait utiliser, mais aussi une structure de livre de codes. L'analyse de contenu ou les fonctions de classification automatique de documents de WordStat peuvent également être utiles pour sélectionner un nombre limité de documents pertinents provenant d'une quantité énorme de données textuelles (trop volumineuses pour être codées manuellement) en vue d'un examen attentif par des codeurs experts.

Les chercheurs qui utilisent WordStat pour quantifier automatiquement les réponses textuelles désireront souvent effectuer des ajustements manuels afin d'obtenir une plus grande précision. S'ils trouvent certains sujets difficiles à identifier automatiquement, ils pourront avoir besoin de revenir à un codage manuel par un expert dans ce domaine. Enfin, la mise au point d'un modèle de classification automatique de documents requiert un ensemble de réponses déjà classées par des codeurs humains. Si un tel ensemble n'est pas disponible, le chercheur n'aura d'autre choix que de le créer. Les fonctionnalités de QDA Miner s'avéreront alors très utiles pour coder et valider rapidement un tel ensemble de données d'apprentissage.

Quelques ressources supplémentaires

La page Web suivante présente des ressources supplémentaires pour en apprendre un peu plus sur l'utilisation de QDA Miner et de WordStat pour analyser des questions ouvertes :

www.provalisresearch.com/questions-ouvertes

Sondages utilisant QDA Miner ou WordStat

La liste ci-dessous présente des études ayant utilisé QDA Miner, WordStat ou une combinaison des deux pour analyser des réponses aux questions ouvertes. La taille de ces enquêtes varie d'un petit échantillon de 104 chercheurs (Wright & Wyatt, 2008) à une enquête beaucoup plus vaste auprès de 41 500 employés du gouvernement américain.

- Ames, D. R. & Bianchi, E.C. (2008). The agreeableness asymmetry in first impressions: Perceivers' impulse to (mis)judge agreeableness and how it is moderated by power. *Personality and Social Psychology Bulletin*, 34, 1719-1736.
- Behruzi, R., Hatem, M., Goulet, L., Fraser, W. (2011). The facilitating factors and barriers encountered in the adoption of a humanized birth care approach in a highly specialized university affiliated hospital. *BMC Women's Health*, 11(1):53 (2011) PMID 22114870.
- Frank, B.A. & Walsh, R.J. (2011). Does reflective learning take place in online MBA introductory quantitative courses? *Journal of Instructional Pedagogies*.
- Harper, C.A., Shaw, C.E., Fly, J.M., & Beaver, J.T., (2012). Attitudes and motivations of Tennessee deer hunters toward quality deer management. *Wildlife Society Bulletin*, 36(2), 277-285.
- Kim, S., Cha, J., Cichy, R.F., Kim, R., & Tkach, J.L. (2016). Roles of private club volunteer leaders: An exploratory study of content analysis. *International Journal of Hospitality & Tourism Administration*, 17(1), 43-71.
- Latham, S. (2009). Contrasting strategic response to economic recession in start-up versus established software firms. *Journal of Small Business Management*, 47(2), 180-201.
- Leurer, M. D., & Misskey, E. (2015). "Be positive as well as realistic": A qualitative description analysis of information gaps experienced by breastfeeding mothers. *International Breastfeeding Journal*, 10(1), 1-11.
- McCandless, D.W, Bruce, R.D., Gebken, R.J. (2010). Retention of Construction teachers in secondary education. *International Journal of Construction Education and Research*, 6(2), 104-121.
- McComas, K.A., Besley, J.C., & Trumbo, C.W. (2007). Why citizens do and do not attend public meetings about local cancer cluster investigations. *Policy Studies Journal*, 34(4), 671-698.
- Pullman, M. McGuire, K, Cleveland, C. (2005). Let me count the words: Quantifying open ended interactions with guests. *Cornell Hotel and Restaurant Administration Quarterly*, 46(3), 323-343.
- Saunders, M.N.K. (October, 2011). Web versus mail: The influence of survey distribution mode on employees' response. *Field Methods*. Online. October 9, 2011.
- Spinks, N., Silburn, N., & Birchall, D. (2006). Educating engineers for the 21st century: The industry view. Henly Management College: Henley on Thames: UK.
- U.S. Merit Systems Protection Board (2011). Making the Right Connections: Targeting the Best Competencies for Training. Washington, D.C.
- Wright, T.S.A, & Wyatt, S.L (2008). Examining influences on environmental concern and career choice among a cohort of environmental scientists. *Applied Environmental Education and Communication*, 7, 30-39.

Coder et analyser les QUESTIONS OUVERTES

» Exploitez la pleine valeur des questions ouvertes

Un sondage contient généralement deux types de questions : ouvertes et fermées. Les questions fermées pour lesquelles la personne interrogée se voit proposer un choix parmi des réponses préétablies sont facilement quantifiables et analysables, mais nécessitent une connaissance préalable permettant de proposer des réponses adéquates, sous peine de retrouver maintes réponses dans la catégorie « autre ». Les questions ouvertes, quant à elles, donnent aux répondants la possibilité de répondre librement, sans être limités ou influencés par les choix de réponses. Ces réponses qualitatives sont cependant plus difficiles à analyser, en raison de leur caractère non structuré. Pourtant, elles peuvent s'avérer très avantageuses. En effet, elles fournissent des réponses souvent plus riches en informations que des questions fermées. Poser des questions ouvertes est une manière de s'ouvrir aux idées neuves, aux réponses non anticipées.

QDA Miner et **WordStat** simplifient grandement les procédures longues et complexes qu'implique le traitement manuel des questions ouvertes, offrant une assistance informatique unique pour coder et analyser très rapidement un grand volume de réponses non structurées. Nous examinons ici certaines fonctionnalités de ces deux outils et vous proposons quelques astuces pour vous permettre de choisir la meilleure stratégie pour vos analyses. Nous nous concentrons sur le codage et l'analyse des questions ouvertes d'un sondage, mais les mêmes techniques s'appliquent à d'autres données composées d'un grand volume de réponses courtes – commentaires de clients, tweets et autres flux de médias sociaux.

» QDA Miner ou WordStat ?

QDA Miner et WordStat offrent tous deux des solutions de codage et d'analyse des questions ouvertes. Le choix du logiciel dépend d'une variété de facteurs : quantité de réponses à coder, temps et ressources financières disponibles, récurrence, fréquence et processus de collecte des données.

Quand utiliser QDA Miner ?

QDA Miner est un outil de codage manuel qui donne le plein contrôle au codeur et où la décision finale d'appliquer ou non un code à une réponse spécifique lui revient. Bien que QDA Miner offre de nombreuses fonctionnalités pour rendre le codage plus rapide et plus fiable, le codeur a toujours la possibilité d'examiner les suggestions avant que tout codage soit appliqué. La fonction d'assistance au codage la plus élémentaire est l'outil de recherche de texte qui permet de récupérer et coder toutes les réponses contenant des mots ou phrases clés spécifiques. Cependant, le logiciel va bien au-delà en offrant des fonctions de recherche avancées reposant sur des techniques de recherche d'informations comme l'apprentissage par exemples avec rétroaction sur la pertinence (« relevance feedback »), l'appariement flou, etc. Voici quelques-uns des outils avancés les plus utiles pour analyser les questions ouvertes :

- L'outil **d'extraction de groupements d'items** est un dispositif de codage innovant. Il est basé sur une méthode d'apprentissage machine non supervisée. Il regroupe automatiquement des réponses similaires (même avec des fautes d'orthographe) et permet d'examiner et de coder rapidement ces groupements à l'aide d'un dispositif intuitif de codage glisser-déposer. Il permet d'augmenter énormément la vitesse du codage manuel.

- La fonction de **similarité de codage** permet d'identifier des segments de texte similaires à des éléments déjà codés dans le projet actuel ou un autre projet. On peut l'utiliser pour accélérer le codage de projets partiellement codés. Dans des projets entièrement codés, elle détecte les éléments qui peuvent avoir été omis. Elle permet ainsi d'augmenter la vitesse et la fiabilité du codage.

- La fonction de **recherche-par-exemples** permet au codeur de se concentrer sur un segment de texte spécifique

ou sur toutes les réponses associées à un code spécifique et de trouver d'autres réponses similaires. Elle inclut une fonctionnalité de « pertinence » permettant au codeur d'indiquer les segments pertinents et non pertinents, ce qui amène le logiciel à apprendre de cette rétroaction, à affiner la recherche et à extraire des exemples plus pertinents.

- Une fois les réponses codées, QDA Miner offre de nombreux outils pour récupérer les réponses associées à un code spécifique, pour calculer des statistiques descriptives et créer des graphes sur ces codes (diagrammes à barres ou circulaires, nuages de codes), examiner la co-occurrence entre les codes en utilisant l'analyse par grappes (« clustering »), le positionnement multidimensionnel ou les graphes de proximité. QDA Miner permet également de mettre en relation les codes et les réponses aux questions fermées en utilisant des outils tels que le tableau croisé, l'analyse factorielle de correspondance, les cartes thermiques, les diagrammes à bulles, etc.

Cluster Codes	Score	Clustered Items	Item Codes
1.50	MANUFACTURING		
8.00	TRANSLATOR		
8.00	PHYSICAL THERAPIST		
8.00	PHYSICAL THERAPIST		
8.00	PHYSICAL THERAPIST		
2.50	PHYSICAL THERAPIST Assistant		
4.54	MASSAGE THERAPIST		
2.80	MASSAGE THERAPY Teacher	Teacher	
2.77	MASSAGE THERAPY		
2.77	MASSAGE THERAPY Student	Student	
3.94	REAL ESTATE		
3.94	REAL ESTATE		
4.54	REAL ESTATE AGENT		
4.54	REAL ESTATE AGENT		
4.54	REAL ESTATE AGENT		
8.00	FIREFIGHTER		

L'outil d'extraction par grappe permet de grouper et coder rapidement les phrases et paragraphes similaires.

Quand utiliser WordStat ?

WordStat se base sur une approche très différente pour le codage et l'analyse des réponses aux questions ouvertes. Au lieu d'interagir avec le codage manuel, il permet d'automatiser en partie ou en totalité l'analyse de grandes quantités de réponses ou l'extraction de thèmes. WordStat est particulièrement adapté à l'analyse de grands volumes de réponses aux questions ouvertes (jusqu'à plusieurs millions de réponses), ou lorsqu'on souhaite développer une solution de codage pouvant s'appliquer aux questions similaires de futurs sondages ou provenant d'une collecte continue. Bien qu'une révision manuelle soit toujours possible, celle-ci survient non pas avant les opérations de codage ou de classification, comme dans QDA Miner, mais plutôt après ces opérations. WordStat offre trois stratégies d'analyse pour les réponses aux questions ouvertes :

1. **TEXT MINING** - L'approche exploratoire des données textuelles combine les techniques TAL (traitement automatique des langues) et les méthodes statistiques. Les thèmes qui ressortent des questions ouvertes peuvent être identifiés automatiquement au moyen de la fonction d'extraction de thèmes (« topic modeling »), ou par les techniques de regroupement hiérarchique pour analyser la co-occurrence des expressions et des mots les plus courants. Le graphe de proximité peut également isoler les mots ou expressions le plus souvent associés à un sujet spécifique, une marque, un individu ou un nom d'entreprise et comparer rapidement ceux-ci avec des mots et expressions associés à un autre sujet (marque, individu ou entreprise). L'analyse factorielle de correspondance peut identifier des mots, expressions ou thèmes spécifiques à différentes classes de répondants. Ces techniques exploratoires peuvent produire des résultats utiles à partir d'une grande collection de données textuelles, souvent en quelques secondes.

2. **ANALYSE DE CONTENU** - WordStat propose des outils d'analyse de contenu à la fine pointe de la technologie pouvant être utilisés pour construire et appliquer des dictionnaires de catégorisation (ou « taxonomies »). Cette approche vise à mesurer la référence à des concepts ou thèmes spécifiques de façon plus précise et plus systématique que l'approche précédente. Un dictionnaire d'analyse de contenu peut inclure un grand nombre de catégories, chacune pouvant contenir plusieurs centaines de mots, de modèles de mots, d'expressions et de règles. On peut appliquer des dictionnaires d'analyse de contenu existants ou créer son propre dictionnaire. Bien que l'élaboration et la validation d'un dictionnaire nécessitent du temps, l'analyse d'un grand volume de réponses ouvertes à l'aide d'un tel dictionnaire s'avère souvent plus rapide que le codage manuel de chaque réponse. Une fois élaboré, le

dictionnaire peut être réappliqué à de nouvelles réponses à des questions similaires et permet ainsi de produire des résultats en quelques secondes. Une telle approche est idéale si l'on a besoin d'automatiser le codage des réponses textuelles à partir d'enquêtes récurrentes ou d'analyser des données textuelles collectées en continu, telles que les flux de commentaires dans les médias sociaux ou sur des sites web.

3. **CLASSIFICATION AUTOMATIQUE** - Une troisième approche consiste à coder des réponses ouvertes en appliquant des techniques d'apprentissage supervisé pour classer automatiquement les réponses en une ou plusieurs catégories. Ceci nécessite un certain volume de réponses déjà catégorisées. Les codes associés à ces réponses sont alors utilisés comme exemples au moyen desquels le logiciel tente d'identifier automatiquement les mots ou expressions caractérisant ces codes afin de les reconnaître dans les réponses non codées. WordStat offre deux techniques d'apprentissage machine : le « Naive Bayes » et la méthode des plus proches voisins (« k-nearest neighbors »). Ces techniques permettent d'effectuer des classements dichotomiques (ex. : présent ou absent, positif ou négatif) ou une répartition dans de multiples classes. La classification automatique de documents peut également être utilisée pour générer un score sur une échelle ordinale (telle qu'une échelle de Likert).

Même si chacune des techniques ci-dessus peut être utilisée séparément, on gagnera souvent à les combiner. Par exemple, on peut employer un regroupement hiérarchique des mots les plus fréquents pour identifier non seulement les thèmes les plus courants, mais également l'agencement de ces thèmes. Une telle structure peut être répliquée dans un dictionnaire d'analyse de contenu par la création de catégories spécifiques pour mesurer ces thèmes et en groupant ces catégories spécifiques sous des thématiques plus larges. Une analyse factorielle de correspondance sur les mots peut également identifier des hypothèses pouvant être testées ultérieurement par l'élaboration de dictionnaires ciblés. Il est possible d'utiliser un dictionnaire d'analyse de contenu pour restreindre la classification automatique à des mots ou expressions spécifiques, ou pour regrouper les formes fléchies, mal orthographiées et les synonymes et s'assurer ainsi qu'ils soient traités comme autant de références à une même idée.

