

APPORTS D'UNE ONTOLOGIE A L'APPRENTISSAGE DES STATISTIQUES

Jean-Marc Meunier¹, Samuel Szoniecky², Myriam Lammolle³

¹Laboratoire Paragraphe, Université Paris 8. jmeunier@univ-paris8.fr

²Laboratoire Paragraphe, Université Paris 8. samuel.szoniecky@univ-paris8.fr

³Laboratoire LIASD, Université Paris 8. m.lammolle@iut.univ-paris8.fr

Résumé. Nous développons dans cet article l'idée que la structuration des connaissances est un enjeu majeur pour l'enseignement des méthodes statistiques. Parallèlement, l'usage croissant de ressources en ligne pose aux apprenants des problèmes de repérage et d'évaluation de la pertinence de celles-ci. La convergence de ces deux problématiques nous a conduits à proposer une ontologie informatique pour l'indexation des ressources pédagogiques. Nous en discutons les intérêts et les perspectives pour le partage des ressources pédagogiques, l'étude des difficultés d'acquisition et le développement de nouveaux dispositifs d'aide à l'apprentissage.

Mots-clés. Apprentissage des statistiques, ontologie, structure des connaissances, ressources pédagogiques

Abstract. Our claim in this article is that the structure of knowledge is a main subject for teaching statistical methods. At the same time, the increasing use of online resources poses problems for learners to identify and evaluate the relevance of those resources. The convergence of these two problems led us to propose a computer ontology for the indexation of teaching resources. We discuss interests and perspectives for sharing educational resources, study learning difficulties and development of new learning aids

Keywords. Teaching statistics, ontology, knowledge structure, pedagogical resources

1 Structurer ses connaissances, un challenge pour l'apprentissage.

Les statistiques constituent pour les étudiants, particulièrement en sciences humaines et sociales, un des cours les plus redoutés et certains développent une attitude négative, voir une appréhension pour ce qui touchent à cette matière. Plusieurs études sont consacrées à ce phénomène d'appréhension pour les statistiques (statistics anxiety). Il trouve son origine en partie dans des facteurs individuels. Cependant la complexité du domaine en constitue également un facteur déterminant (Crvalho, 2010; Onwuegbuzie, 2003). Apprendre les statistiques nécessite l'acquisition d'un savoir très abstrait et en même temps des procédures et les conditions d'application de celles-ci afin de pouvoir les utiliser dans des situations professionnelles. Dans le contexte des sciences humaines, il s'agit pas bien sûr de transmettre des connaissances, mais surtout de fournir aux étudiants des moyens d'apprendre quand et surtout pourquoi utiliser des procédures, ce qui suppose qu'on construise les classes de situations correspondant aux différents cas de figure (Nitko & Lane, 1990). Or la catégorisation pertinente est loin d'être spontanément construite par les étudiants.

Ainsi, Quilici & Mayer (2002) ont montré que l'organisation des connaissances en statistiques est fortement influencée par le contenu sémantique des exemples utilisés au cours de l'apprentissage. Cet effet est d'autant plus important que le nombre d'exemples utilisés au cours de l'apprentissage est faible. C'est également vrai dans d'autres domaines que les statistiques notamment la résolution de

problèmes (Clément, 2009). Ainsi, plusieurs études ont montré que les experts réussissent à dépasser ces difficultés en construisant des classes de situations et en leur associant des schémas de résolution (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Sweller, 1988). Cela demande cependant une expérience solide que les cours ne suffisent pas toujours à construire de manière stable. Il est possible d'aider l'apprenant en le faisant travailler sur la structure conceptuelle du domaine soit à l'aide d'un réseau sémantique (Meunier, 2008; Sander, Meunier, & Bosc-Miné, 2004), soit en le faisant travailler sur des cartes conceptuelles (Chiou, Wang, & Lee, 2014; Roberts, 1999). Les progrès dans les environnements informatiques d'apprentissage humains incitent à aller beaucoup plus loin.

2. Apprendre à l'heure d'Internet

S'agissant des statistiques, comme dans bien d'autres domaines, la quantité de ressources pédagogiques libres ou éditoriales disponibles sur le web est colossale. Sous réserve de pouvoir s'y repérer, il devient possible de développer ses connaissances dans des domaines extrêmement variés. Cette opportunité nécessite cependant, tant pour l'apprenant que pour l'enseignant, qu'ils possèdent un certain nombre de compétences afin de naviguer dans ces ressources (Terras, Ramsay, & Boyle, 2013).

Ces compétences sont la capacité à chercher l'information, à faire le tri parmi les résultats de la recherche puis à extraire de ces documents les éléments utiles et à les traiter pour développer ses connaissances ou, lorsqu'on est enseignant, en faciliter l'acquisition à ses étudiants. Plusieurs auteurs ont montré que les stratégies de recherches varient d'un individu à un autre. Les uns fondent leurs recherches plutôt sur des critères verbaux comme des mots clés ou des phrases, d'autres plutôt sur des images. Certains appréhendent les informations dans leur ensemble (stratégie globale) ou focaliser sur les différents aspects de la situation (stratégie analytique). Ces caractéristiques sont très liées à l'expérience de l'internaute pour la recherche sur le web (Kinley, Tjondronegoro, Partridge, & Edwards, 2014) et à ses connaissances sur le domaine, indispensables pour formuler des requêtes pertinentes. Celles-ci peuvent parfois être en décalage avec les structures de mots clés utilisées pour l'indexation des ressources dans les systèmes documentaires (Holman, 2011).

3. Vers une ontologie des concepts statistiques

La convergence de ces deux problématiques, faciliter l'organisation des connaissances chez l'apprenant et structurer un ensemble de ressources pédagogiques, nous a orienté vers les technologies du Web sémantique pour proposer une ontologie des concepts statistiques. Ce projet, baptisé Ontostats (<http://gapai.univ-paris8.fr/ontostats/>) a reçu le soutien du ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche dans le cadre de l'appel à manifestation d'intérêt 2016 de la MIPNES (Mission de la pédagogie et du numérique pour l'enseignement supérieur).

Le projet n'est pour l'instant que dans sa phase 1, l'élaboration de l'ontologie. Il s'agit ici de réaliser un démonstrateur permettant de valider l'utilisabilité d'une ontologie pour l'indexation des ressources éducatives libres et leur réutilisation par les enseignants et les étudiants. Ce travail s'appuie sur des exemples antérieurs d'indexation de ressources pédagogiques à l'aide d'une ontologie comme SemUNT (Isaac, Bourda, & Grandbastien, 2012).

Pour construire l'ontologie, nous avons adopté une démarche déjà éprouvée dans d'autres domaines et connue sous le nom de « méthontology » (Corcho, Fernández-López, Gómez-Pérez, & López-Cima, 2005). Nous ne rentrerons pas dans les détails techniques de ce développement mais en

présenterons succinctement les principales étapes.

La première étape consiste à dresser la liste aussi complète que possible des termes qui seront utilisés. Afin d'adopter un vocabulaire largement partagé, nous avons retenu le glossaire de l'International Statistical Institute (<http://isi.cbs.nl/glossary/>). Celui-ci comporte 3942 entrées traduites en 29 langues. Pour chacune d'elles, les synonymes sont indiqués. Nous avons retenu 453 entrées correspondant au vocabulaire utilisé dans nos cours. Ce choix de disposer d'un vocabulaire multilingue nous a conduits à ne pas réutiliser une ontologie existante sur les concepts statistiques : STATO (Gonzalez-Beltran, Rocca-Serra, Burke, & Sansone, 2014). Cette dernière, développée à l'université d'Oxford, est cependant suffisamment proche dans sa conception de notre approche pour que nous puissions envisager une connexion entre les deux ontologies. La seconde étape consiste à définir les relations entre les classes et leurs attributs

Enfin la validation de l'ontologie, tant dans ses aspects structuraux que sémantiques sera réalisée en nous inspirant de la méthodologie développée par Richard, Aimé, Krebs, & Charlet (2015). Elle prévoit l'utilisation d'outils dédiés tels que des raisonneurs pour vérifier la cohérence structurelle et des focus groupes avec des experts du domaine pour les aspects sémantiques.

4. Intérêts et perspectives

Une ontologie est avant tout un système de gestion des connaissances permettant d'explicitier des relations. Grâce aux liens entre les concepts, l'ontologie permet de faire un certain nombre d'inférences. Ce sont ces inférences qui donnent à l'ontologie toute sa valeur. Elle est bien plus qu'un thésaurus organisant un ensemble de termes. Elle permet de décrire dans un langage qui soit traitable par une machine un ensemble de connaissances. Pour l'enseignement des statistiques, comme dans bien d'autres domaines, une ontologie offre plusieurs avantages.

Partager des connaissances

L'objectif premier d'une ontologie est le partage des connaissances. Elle s'appuie bien sûr sur un vocabulaire, qui sert à étiqueter des concepts, mais surtout sur des relations entre ces concepts. Les concepts sont les classes considérées dans l'ontologie et non les étiquettes elles-mêmes ce qui autorise une description multilingue du domaine de connaissances. Les relations sont des relations d'inclusion de classes, mais également des propriétés ou attributs pouvant prendre plusieurs valeurs. On peut ainsi dire d'un individu qu'il est un homme ou une femme ce qui revient à l'affecter à une classe ou dire de lui qu'il a un sexe pouvant prendre deux valeurs disjointes (homme ou femme) ce qui revient à lui donner un attribut. Ces relations permettent de traiter ce qui est habituellement implicite dans le discours. Prenons l'exemple de la notion de variable. Nous trouvons dans de nombreux de manuels une opposition entre variables qualitatives et variables quantitatives, les variables ordinales étant traitées comme un cas particulier des premières en raison de la relation d'ordre existant entre les échelons. Or les modalités d'une variable numérique sont aussi organisées par une relation d'ordre. D'un certain point de vue, une variable numérique est donc aussi un cas particulier d'une variable ordinale qui en plus de l'ordre entre les modalités respecte un intervalle constant entre celles-ci. La transitivité de l'inclusion de classes nous conduit donc naturellement à considérer que les variables quantitatives constituent une sous-classe des variables qualitatives. Pour dépasser cette difficulté, il faut expliciter des relations qui ne le sont pas en posant par exemple l'existence d'une classe des variables ordonnées. Cet exemple limité peut être généralisé à bien d'autres notions ou situation d'analyse. L'ontologie permet ainsi de partager une compréhension commune de la structure des

connaissances, mais aussi d'explorer les différentes facettes d'un concept en le présentant sous différents points de vue.

Construire des systèmes d'aide à la décision.

Une des applications des ontologies, en particulier dans le domaine médical, est l'élaboration de système d'aide à la décision. On peut citer par exemple Raby & Raveau (2011) qui ont utilisé une ontologie pour la réalisation d'un outil de formation au diagnostic chez des infirmières. Dans le domaine de l'enseignement des statistiques, l'équivalent du diagnostic pour un apprenant est l'identification de la procédure adéquate. Il est possible de leur donner un arbre de décision pour sélectionner une procédure (Johnson & Karunakaran, 2014), mais l'ontologie offre l'opportunité d'une approche plus conceptuelle qui permettent en même temps de donner à voir à l'apprenant les prérequis, les limites et les justifications qui conduisent au choix d'une procédure. Gonzalez-Beltran et al. (2014) en ont développé quelques exemples à l'aide de STATO (<http://stato-ontology.org/>). Il est ainsi de demander à l'ontologie quels tests sont applicables en fonction d'un objectif d'analyse, d'en récupérer la liste et de vérifier les conditions d'application de chacun d'eux. Une telle aide à la décision est utile à l'apprenant, mais aussi au praticien, en particulier aux chercheurs non-statisticien qui pourraient ainsi vérifier que leurs choix méthodologiques sont corrects.

Indexer des ressources documentaires

Nous l'avons déjà évoqué plus haut, les ontologies présentent pour l'indexation des ressources pédagogiques de nombreux avantages par rapport à une norme (Bouzeghoub, Defude, Duitama, & Lecocq, 2005). Elles facilitent l'indexation des ressources. Si l'auteur n'est pas exhaustif, le système peut déduire les équivalences terminologiques et d'autres concepts pertinents pour indexer la ressource. Ainsi une ressource indexée avec le mot-clé « Test de Student » pourra être mis en relation avec les mots-clé « moyenne », « variable numérique » ou « test paramétrique » grâce à aux relations entre les concepts. Le langage de description et d'interrogations des ontologies sont également conçus pour assurer l'interopérabilité entre les dispositifs informatiques l'utilisant (Bouzeghoub & Elbyed, 2006). Il permet surtout l'agrégation de données multi-sources. Ainsi, dans notre projet nous avons fait converger les données du glossaire de l'International Statistical Institute, mais aussi les données disponibles sur Wikipedia et celles provenant de l'open data de l'enseignement supérieur et de la recherche pour la description des ressources numériques disponibles. Enfin les ontologies peuvent aussi être utilisées pour la fouille de données ou de documents et donc l'une indexation automatique (Sy, 2012) et/ou l'enrichissement automatique de l'ontologie.

Construire des dispositifs d'aide à l'apprentissage

L'ontologie, en tant que modélisation d'un domaine de connaissance, constitue une importante base pour l'étude des difficultés d'apprentissage des statistiques en même temps qu'un support pour la construction de système d'aide à l'apprentissage. Avec une telle ontologie, nous disposons en effet d'un modèle conceptuel offrant l'opportunité de mieux étudier les difficultés d'apprentissage mais aussi les erreurs raisonnement que peuvent faire les étudiants en testant par exemple des inférences produites à partir du modèle. Une ontologie peut aussi être utilisée pour concevoir des parcours de formation en suggérant par exemple la structuration des séquences pédagogiques (Guo & Chen, 2007), concevoir des systèmes de recommandation de ressources s'appuyant sur les relations entre les concepts (Shen & Shen, 2005), mais aussi, lorsque l'ontologie est associée à des données

d'apprentissage, la personnalisation de parcours (Castro & Alonso, 2011) et l'évaluation de l'apprenant (Romero, North, Gutiérrez, & Caliusco, 2015).

5. Conclusion

Avec le projet Ontostats, nous ambitionnons d'articuler des problématiques cognitives et des problématiques documentaires. Le démonstrateur en cours de développement offrira dans un premier temps une plateforme de partage de ressources documentaires, la possibilité pour les enseignants de les scénariser à leur guise et pour les étudiants d'y rechercher des compléments de cours. L'utilisation d'une ontologie pour modéliser le domaine des statistiques ouvre plusieurs perspectives pour l'étude des difficultés d'apprentissage, le partage des ressources documentaires et pour l'élaboration de dispositifs d'aide à l'apprentissage. Au-delà du dispositif technique, c'est la valeur heuristique de la démarche qui constitue le principal apport de ce projet que nous souhaitons partager le plus largement possible.

Bibliographie

- [1] Bouzeghoub, A., & Elbyed, A. (2006). *Ontology Mapping for Learning Objects Repositories Interoperability* (English). In *Lect. notes comput. sci.* (p. 794-797). Springer, Berlin, New York.
- [2] Bouzeghoub, A., Defude, B., Duitama, J.-F., & Lecocq, C. (2005). Un modèle de description sémantique de ressources pédagogiques basé sur une ontologie de domaine. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF)*, 12, 17 pages.
- [3] Castro, F., & Alonso, M. A. (2011). *Learning Objects and Ontologies to Perform Educational Data Mining* (English) (p. 532-536). CSREA Press.
- [4] Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations : How students study and use examples in learning to solve problems. *Cognitive science*, 13, 145-182.
- [5] Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121-152.
- [6] Chiou, C.-C. (2009). Effects of concept mapping strategy on learning performance in business and economics statistics. *Teaching in Higher Education*, 14(1), 55-69.
- [7] Chiou, C.-C., Wang, Y.-M., & Lee, L.-T. (2014). Reducing statistics anxiety and enhancing statistics learning achievement: Effectiveness of a one-minute strategy. *Psychological Reports*, 115(1), 297-310.
- [8] Corcho, O., Fernández-López, M., Gómez-Pérez, A., & López-Cima, A. (2005). Building Legal Ontologies with METHONTOLOGY and WebODE. In V. R. Benjamins, P. Casanovas, J. Breuker, & A. Gangemi (Éd.), *Law and the Semantic Web* (p. 142-157). Springer Berlin Heidelberg.
- [9] Cravalho, P. (2010). *Learning Statistics using Concept Maps: Effects on Anxiety and Performance* (master thesis). San Jose State University.
- [10] Ford, N., Miller, D., & Moss, N. (2001). The role of individual differences in Internet searching: An empirical study. *Journal of the American Society for Information Science and Technology*, 52(12), 1049-1066.
- [11] Gonzalez-Beltran, A., Rocca-Serra, P., Burke, O., & Sansone, S.-A. (2014). *stato-ontology.org*. Consulté 19 avril 2017, à l'adresse <http://stato-ontology.org/>

- [12] Guo, W.-Y., & Chen, D.-R. (2007). An Ontology Infrastructure for an E-Learning Scenario. *International Journal of Distance Education Technologies*, 5(1), 70-78.
- [13] Holman, L. (2011). Millennial Students' Mental Models of Search: Implications for Academic Librarians and Database Developers. *The Journal of Academic Librarianship*, 37(1), 19-27.
- [14] Isaac, Y., Bourda, Y., & Grandbastien, M. (2012). SemUNIT - French UNT and Linked Data (Vol. 840, p. 6 pages). Présenté à LiLe-2012 at WWW-2012, CEUR workshop proceedings.
- [15] Johnson, L. R., & Karunakaran, U. D. (2014). How to Choose the Appropriate Statistical Test Using the Free Program « Statistics Open For All » (SOFA). *Annals of Community Health*, 2(2), 54-62.
- [16] Kinley, K., Tjondronegoro, D., Partridge, H., & Edwards, S. (2014). Modeling users' web search behavior and their cognitive styles. *Journal of the Association for Information Science and Technology*, 65(6), 1107-1123.
- [17] Meunier, J.-M. (2008). Des préconceptions en statistiques des étudiants de psychologie à la conception d'un EIAH. Présenté à Journée d'étude « Recherche sur la cognition pour les EIAH », Université Paris 8, Saint-Denis.
- [18] Nitko, A. J., & Lane, S. (1990). Solving Problems Is Not Enough: Assessing and Diagnosing the Ways in Which Students Organize Statistical Concepts.
- [19] Onwuegbuzie, A. J. (2003). Modeling statistics achievement among graduate students. *Educ. psychol. meas.*, 63(6), 1020-1038.
- [20] Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, 88(1), 144-161.
- [21] Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, 16(3), 325-341.
- [22] Richard, M., Aimé, X., Krebs, M.-O., & Charlet, J. (2015). LOVMI : vers une méthode interactive pour la validation d'ontologies. Présenté à Plate-forme Intelligence artificielle, Rennes, France.
- [23] Roberts, L. (1999). Using concept maps to measure statistical understanding. *International Journal of Mathematical Education in Science & Technology*, 30(5), 707-717.
- [24] Romero, L., North, M., Gutiérrez, M., & Caliusco, L. (2015). Pedagogically-Driven Ontology Network for Conceptualizing the e-Learning Assessment Domain. *Journal of Educational Technology & Society*, 18(4), 312-330.
- [25] Sander, E., Meunier, J. M., & Bosc-Miné, C. (2004). Approche ontologique et navigation dans un E.I.A.H: Le cas de l'enseignement des statistiques. *Revue STICEF*, 11.
- [26] Schau, C., & Mattern, N. (1997). Use of Map Techniques in Teaching Applied Statistics Courses. *The American Statistician*, 51(2), 171-175.
- [27] Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12(2), 257-285.
- [28] Sy, M.-F. (2012). Utilisation d'ontologies comme support à la recherche et à la navigation dans une collection de documents. Thèse de doctorat. Université Montpellier II - Sciences et Techniques du Languedoc.
- [29] Terras, M. M., Ramsay, J., & Boyle, E. (2013). Learning and Open Educational Resources: A Psychological Perspective. *E-Learning and Digital Media*, 10(2), 161-173.