Langages documentaires et ontologies

Synthèse

Je travaille pour mon stage de master 2 à l'Inrp et à l'Ens sciences de Lyon à améliorer des taxonomies, que les concepteurs, des scientifiques de différentes disciplines appelaient thésaurus. J'ai à l'origine cherché des informations sur ces sujets pour réussir à définir et à donner un nom exact aux outils qui avaient été élaborés. Au fil de mes lectures, je me suis intéressée aux ontologies, je propose ici une synthèse de mes lectures sur ces sujets.

Certains auteurs cités par Brunon Menon, proposent de nommer tous ces outils « systèmes d'organisation des connaissances » ou « schémas de concept ». Ces expressions permettent de regrouper les listes d'autorités, glossaires, dictionnaires, classifications, plans de classement, taxonomies, thésaurus et ontologies. Je reprendrais cette appellation dans cette synthèse.

Taxonomie

Les taxonomies (de termes, de classes, de concepts) désignent la hiérarchie ou l'arborescence autour de laquelle sont construits différents types d'instruments, comme les thésaurus ou les ontologies.

D'un point de vue fonctionnel, une taxonomie est un cadre d'organisation pour des ressources numériques de toute nature (et pas seulement documentaires), destiné à en permettre une présentation ordonnée et y donnant accès par navigation hypertextuelle.

Thésaurus

Un thésaurus est un index de termes organisés selon une liste hiérarchique. Le terme étant défini comme une représentation linguistique

d'un concept unique. Ils sont organisés en thèmes et possèdent des liens sémantiques entre eux : synonymie, équivalence, terme spécifique (lien vers un concept de sens plus précis), terme général (lien vers un concept de sens plus large).

Un thésaurus a trois grandes fonctions : l'indexation des ressources, la recherche documentaire en réduisant par un vocabulaire contrôlé l'impact de la synonymie dans le langage naturel. La troisième grande fonction est de permettre un repérage exhaustif des ressources par des liens entre les concepts et mots proches.

Ontologies

Petit Robert : Ontologie : partie de la métaphysique qui étudie l'être en tant qu'être, indépendamment de ses déterminations particulières.

Depuis une dizaine d'année, c'est dans le domaine de l'ingénierie des connaissances que l'ontologie est apparue sous une acception informationnelle. Une ontologie est un ensemble de concepts, de propriétés, d'axiomes, de fonctions et de contraintes explicitement définies. C'est une conceptualisation qui est le résultat de la formulation exhaustive et rigoureuse d'un domaine. Le vocabulaire qui compose une ontologie est choisie de manière consensuelle et la conceptualisation de celui-ci doit pouvoir être comprise et interprétée par un ordinateur.

L'ontologie organise les concepts de manière hiérarchique, élabore des relations entre concepts limitées à un nombre déterminé à l'avance ainsi que des règles et axiomes qui les contraignent. Les objets sont définis d'une part et leurs relations d'une autre.

Les ontologies : concepts et représentations

Cette partie est une synthèse de l'article de Fabien Gandon, assez clair ce qui est rare sur la question des ontologies…

Pour exprimer un concept nous choisissons une représentation symbolique souvent linguistique et verbale ou iconique. Pour un même objet,

nous pouvons donner diverses représentations linguistiques. On pourrait parler de voiture, d'auto, de véhicule automobile, de tacot, de bagnole…etc.
Nous dissocions les concepts et leurs manifestations linguistiques. Un terme n'est pas un concept et vice-versa. Un terme peut être ambigu alors qu'un concept n'a qu'un seul sens, une seule définition. Il faut alors gérer les problèmes de synonymie (un concept dénoté par plusieurs termes) et d'homonymie (un terme dénotant plusieurs concepts).

Dans une ontologie, il faut distinguer les intentions et les extensions d'un concept. L'intention est un ensemble de propriétés communes aux individus auxquels le concept s'applique et permettant de définir ce concept.
Par exemple : le concept *C* est une sous-catégorie de véhicule de transports automobiles, aménagé pour le transport d'un petit nombre de personnes et comprenant au minimum 3 roues.
L'extension est un ensemble d'entité entrant dans cette catégorie.
Par exemple : la twingo de Rose, la kangoo d'Olivier

Je cite librement Fabien Gandon dans les paragraphes qui suivent : « Dans une ontologie les intentions sont organisées, structurées et contraintes de façon à représenter notre conception du monde et de ses contraintes. Par exemple, une voiture est forcément un véhicule. L'intention permet de savoir de façon non ambiguë comment concevoir de façon précise le sens. Ce qui permet sa manipulation par une machine. Les intentions sont organisées par taxonomie ou hiérarchie de types. Cette organisation a une grande importance : la classification et la catégorisation sont des inférences que nous faisons à longueur de journée.
Exemple : « - Tu connais un *restaurant* proche ?
- Il y a une *pizzeria* au coin de la rue.
- Merci. »
Dans une conversation aussi banale, la première personne a généralisé sa requête au concept de *restaurant*, qui représente la catégorie la plus abstraite recouvrant toutes les formes de réponses acceptables. La deuxième a, probablement sans même y prêter attention, utilisé sa taxinomie de concepts pour en déduire qu'une *pizzeria* est un *restaurant*, et que par conséquent sa

réponse est pertinente. Le fait que cette connaissance taxinomique soit partagée est implicite, puisque la deuxième personne suppose que sa réponse sera comprise sans préciser qu'une *pizzeria* est un *restaurant*, et que c'est effectivement le cas. Le recours à des conceptualisations partagées et aux inférences qu'elles permettent est donc au cœur d'activités aussi simples que cet échange d'information. Le fait de rendre explicites les connaissances ontologiques et de s'assurer de leur nature consensuelle est un des problèmes majeurs de l'ingénierie ontologique. Ainsi, dans un système d'information, le simple ajout de cette connaissance peut permettre d'améliorer considérablement les capacités des machines.

Les connaissances taxinomiques permettent de faire un certain nombre de déduction comme:

- Un roman est un livre, un livre est un document
- Un homme et une femme sont des sous-types d'humains qui sont eux-mêmes sous-type d'être vivants.

Il s'agit d'expliquer au système une partie de notre réalité, l'ontologie permet de factoriser des connaissances et de pouvoir les réutiliser ou les échanger.

Mais les connaissances ontologiques dépassent ces connaissances taxinomiques.

On peut introduire dans une ontologie :

- des définitions complètes « une personne est directeur seulement s'il dirige une organisation »,
- des contraintes d'intégrité « un parent ne peut pas être plus jeune que ses enfants »,
- des fonctions de calcul,
- des relations réciproques « si Thomas est marié à Stéphanie alors Stéphanie est mariée à Thomas »,
- des connaissances par défaut « une voiture a 4 roues »,
- Le calcul d'une fermeture transitive « si un coupé est une voiture et une voiture est un véhicule alors un coupé est un véhicule. » »

Web sémantique

Le développement des ontologies va de pair avec le concept du web sémantique. Le web sémantique est une extension du web dit syntaxique. Cette nouvelle conception du web, appelé aussi web 2.0, propose d'améliorer la recherche d'information par une représentation sémantique des ressources, qui devrait permettre des contenus exploitables et interprétables par des machines.

Le développement du web sémantique repose sur un certain nombre de composants comme des métadonnées, une méthode de raisonnement, un format d'organisation des connaissances et des systèmes d'organisation des connaissances. Ces trois derniers points étant la composante même des ontologies. « Pour être susceptible d'être exploitées automatiquement, les métadonnées doivent être entièrement explicites et exprimées selon un vocabulaire formellement défini. Les ontologies sont le réceptacle de ces définitions. » J. Chaumiez

Contrairement aux autres langages de formalisation, les ontologies n'ont pas comme seule application l'aide à la gestion de ressources documentaires. Les ontologies se retrouvent dans de nombreux domaines d'application comme : l'intégration d'informations géographiques, la gestion de ressources humaines, l'aide à l'analyse en biologie, le commerce électronique, l'enseignement assisté par ordinateur, les bibliothèques numériques, les échanges commerciaux entre partenaires industriels, le suivi médical informatisé...

Ontologies et thésaurus

Les ontologies, de la même façon que les thésaurus, les taxonomies ou les classifications sont des systèmes de représentation des connaissances

d’un domaine et utilisés dans des manipulations sémantiques afin de rechercher de l’information. Le tableau ci-dessous recense les ressemblances et différences des thésaurus et ontologies.

	Thésaurus	Ontologies
Origine	Sciences documentaires	Philosophie puis le concept a été repris en intelligence artificielle
Fonctions		
Indexation de ressources	manuelle	automatique
Recherche documentaire	• Permet une recherche pertinente mais se pose un problème d'incompatibilité entre les thésaurus • Possibilité de trouver des concepts signalés comme proches dans le thésaurus	• Permet une recherche intelligente et raisonnée par les liens sémantiques entre concepts contrôlés par des règles. • Permet le rappel de notions liées par une généralisation de la requête s'il y a trop peu de résultats ou à l'inverse un regroupement thématique des résultats s'ils sont en grand nombre • Recherche sémantique sur le web par des recherches dans des sources hétérogènes, • Annotations de ressources web,
Utilisation	Production manuelle de description documentaire	Description des ressources automatiques
Autres utilisations		• Gestion de mémoire d'entreprise, • Permettre l'interopérabilité de ressources en réseau • Piloter des traitements de traitement automatique des langues, • Vérifier la cohérence de modèle • Permettre des raisonnements temporels et spatiaux

		• Faire des approximations logiques
Conception		
Aides à l'élaboration	Normes	Une norme: l'OWL: Web Ontology Language (langage informatique utilisé pour modéliser des ontologies source: Wikipédia)
Méthodologie		• Construction d'un corpus de documents • Analyse linguistique du corpus • Normalisation sémantique des termes du corpus (ontologie différentielle • Définition des relations « ontologie formelle » • Traduction des relations dans un langage compréhensible par une machine « ontologie computationnelle » • Elaboration de l'ontologie opérationnelle
Liens sémantiques	Liens simples • hiérarchiques, • associatif • d'équivalence par la gestion des synonymies et antonymie (descripteurs et non-descripteurs)	Liens complexes • hiérarchiques, • règles ou axiomes définissant et contrôlant les liens sémantiques complexes en fonction de l'objectif de l'ontologie
Langage de formalisation	Langage naturel	Langage opérationnel compréhensible par une machine

Pour conclure on peut dire qu'il existe des ressemblances structurelles entre ontologies et langages documentaires. Mais malgré des analogies entre ces modes de représentation des connaissances, il est inexact d'assimiler ontologies et langages documentaires. Une ontologie a une vocation plus large. « Une ontologie n'est pas un catalogue du monde, une taxonomie, une terminologie ou une liste d'objets, de choses ou de quoi que ce soit d'autre. Une ontologie serait plutôt le cadre général – la structure- au sein duquel catalogues, taxonomies, terminologies peuvent recevoir une organisation appropriée » (Roberto Poli cité et traduit par Bruno Menon).

Un thesaurus peut être transformé en ontologie par un travail de conceptualisation et de formalisation des connaissances. On peut à ce propos citer Bruno Bonchimont « Les ontologies n'étant pas sans rapport avec les terminologies, on peut trouver dans les thésaurus des ressources pour amorcer une ontologie. Mais il faut prendre garde au fait que ce ne sont que des «ressources pour» et pas des «embryons de». Le même auteur précise que « les ontologies sont des artefacts construits en fonction d'une tache précise et ne peuvent être réutilisées, en tant qu'objets formels, pour une autre tâche.» (citations tirées de l'article de Jacques Chaumier)

Bibliographie :

Jean Aitchison, Alan Gilchrist, *Construire un thesaurus, Manuel pratique*, (traduction Dominique Hervieu, révision scientifique Jacques Maniez,) ADBS, Paris, 1992. ISBN 2 901 046 43 6

Jacques Chaumier, *Les ontologies, antécédents, aspects techniques et limites*, Langages documentaires et outils linguistiques, numéro spécial Documentalistes-Sciences de l'information, février 2007, vol. 44 , n°1.

Renaud Eppstein, *Ontologies* http://artist.inist.fr/article.php3?id_article=243 [consulté le 11 avril 2007]

Fabien Gandon , *Ontologies informatiques* http://interstices.info/display.jsp?id=c_17672 [consulté le 11 avril 2007] (une partie de cet article est largement citée, la plupart des exemples de la partie portant sur les ontologies en sont tirés)

Anissa Ghouas Dziri, *Ontologies et thésaurus : des systèmes d'organisation des connaissances différents et complémentaires*, Indice, index, indexation, actes du colloque organisé les 3 et 4 novembre 2005 à l'université Lille-3 par les laboratoires CERSATES et GERICO, ADBS, Paris, 2006.ISBN 2 84365 088 7

Michèle Hudon, *Le thésaurus: conception, élaboration, gestion*, Coll.Clé en main, Asted, Montreal, 1994. ISBN 2 921 548 14 3

Bruno Menon, *Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan*, l'œil de l'ADBS http://www.adbs.fr/site/publications/oeil_adbs/17/oeil_adbs_no17.html#metiers_mot_cle

Bruno Menon, *Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan*, Langages documentaires et outils linguistiques, numéro spécial Documentalistes-Sciences de l'information, février 2007, vol. 44 , n °1.

Chantal Reynaud, Brigitte Safar, *Ontologies, fondements du Web sémantique*, http://master.lri.fr/recherche/intranet/current/cours/affiche_cours.php?wsl_dir_project=/users/master/master/WWW/recherche/intranet/current&cours=53 [consulté le 11 avril 2007]

A voir aussi : un cours qui peut aussi être intéressant, par ses explications illustrées de schémas.

Robert Laurini, Ontologies et Interopérabilité, janvier 2001, http://lisi.insa-lyon.fr/~laurini/disic/dea3/sld001.htm [consulté le 11 avril 2007]