

Sommaire

Inhoudstafel

60^{ème} année - 2006 - n° 1

60ste jaargang - 2006 - nr 1

- Éditorial - Woord vooraf 3
- Les nouvelles tendances des langages documentaires
Ron Davies 4
- Les bases de données spatiales
Jean-Paul Donnay 11
- Een kort overzicht van data warehousing en OLAP
Jef Wijsen 19
- Eurovoc, thésaurus multilingue : Maintenance et
aspects 25
Exemple à la bibliothèque du Parlement européen
Isabelle Gautier
- Cash for knowledge ? Ethical implications of Patenting
Academic Research 35
Quatrième Forum Éthique de la Fondation Universitaire
Compte-rendu de Simone Jérôme
- Nouvelles parutions – Nieuwe publicaties 42
- Regards sur la presse – Een blik op de pers 45
- Index 2005 51

par
door

Simone JÉRÔME

traduit par
vertaald door

Pascale VAN DINTER

Une nouvelle année des *Cahiers* s'ouvre avec ce numéro et le calendrier veut que sa sortie suive de près notre Assemblée Générale. À l'heure où j'écris ces lignes, aucun des deux événements ne s'est produit et je ne puis qu'exprimer souhaits ou prédictions. Quand vous lirez ce texte, il vous suffira de changer le présent en imparfait et de vérifier si ces dernières se sont réalisées.

Plusieurs administrateurs sont à la fin de leur mandat. Je pense néanmoins que peu d'entre eux vont quitter le navire et je prévois même une ou deux, qui sait peut-être trois, nouvelles arrivées. Que ces nouveaux membres soient les bienvenus. Le travail ne manque pas et, croyez-m'en, l'ambiance est excellente dans l'équipe. Je puis l'affirmer sans craindre la contradiction.

Notre équipe des *Cahiers* va continuer son petit bonhomme de chemin et si j'en crois les échos, son action est appréciée par nos lecteurs. Nous n'avons qu'une ambition : continuer sur la même voie pour vous tenir informés de l'actualité des sciences et de la pratique documentaires, à travers des articles de qualité agréablement présentés.

Mais savez-vous qu'il n'est pas nécessaire d'être membre du conseil d'administration pour faire partie du comité de rédaction des *Cahiers* ? Il n'est pas non plus nécessaire d'avoir une longue expérience en information - documentation. Au contraire, nous souhaitons rajeunir le comité pour assurer l'avenir. Si vous pouvez nous consacrer un peu de temps, venez nous rejoindre. Nous avons des projets et nous avons besoin de vous pour les réaliser. Je vous le promets : vous ne le regretterez pas, vous apprendrez beaucoup et élargirez votre réseau professionnel. N'hésitez pas à me contacter à l'adresse <cahiers-bladen@abd-bvd.net>, que ce soit pour nous rejoindre, pour nous soumettre un article ou pour toute question relative à cet éditorial.

Merci d'avance.

Een nieuw jaar voor de *Bladen* begint met dit nummer en de planning wil dat deze uitgave zeer kort volgt op onze Algemene Vergadering. Op het moment dat ik deze lijnen neerschrijf, heeft nog geen enkel van de twee evenementen plaats gevonden en kan ik noch wensen of voorspellingen uitdrukken. Wanneer u deze tekst zal lezen, volstaat het om het heden te veranderen in het verleden en te verifiëren of deze laatsten ondertussen zijn gerealiseerd.

Meerdere bestuursleden zijn aan het einde van hun mandaat. Ik denk nochtans dat slechts enkelen van hen het schip zullen verlaten en ik voorzie zelfs dat er een of twee, wie weet zelfs misschien drie, nieuwe zullen bijkomen. Dat deze nieuwe leden welgekomen zijn. Er is zeker geen gebrek aan werk en geloof mij, de sfeer is uitstekend in het team. Ik kan dit bevestigen zonder het tegendeel te vrezen

Ons team van de *Bladen* zal dezelfde weg blijven volgen en als ik de echo's mag geloven, worden zijn acties gewaardeerd door de lezers. Wij hebben slechts één ambitie: op dezelfde manier verder gaan om u geïnformeerd te houden over de actualiteiten ivm theorie en praktijk van de documentaire wetenschappen, dit met aange-naam gepresenteerde kwaliteitsartikels.

Maar weet u dat het niet nodig is om lid te zijn van de Raad van Beheer om deel uit te maken van het redactieteam van de *Bladeren*? Het is ook niet nodig om een lange ervaring in informatie en documentatie te hebben. Integendeel, wij willen het team verjongen om zodoende onze toekomst te verzekeren. Als u ons een beetje van uw tijd kan aanbieden, kom u dan bij ons vervoegen. Wij hebben projecten en wij hebben u nodig om ze te realiseren. Ik beloof het u: u zal er geen spijt van hebben, u zal veel leren en uw professioneel netwerk vergroten. Aarzel niet om mij te contacteren op het volgende adres <cahiers-bladen@abd-bvd.net>, hetzij om u bij ons te vervoegen, hetzij om ons een artikel aan te bieden of voor alle vragen in verband met dit woord vooraf.

Dank bij voorbaat.

LES NOUVELLES TENDANCES DES LANGAGES DOCUMENTAIRES

Ron DAVIES

Expert conseil indépendant

Article rédigé suite à la conférence donnée par l'auteur dans le cadre de l'*Inforum 2005, Systèmes et langages documentaires : fictions et réalités*, organisé par l'Association Belge de Documentation, le 28 avril 2005 à Bruxelles.

Het artikel is opgesteld naar aanleiding van een conferentie gegeven door de auteur in het kader van *Inforum 2005, Informatiesystemen: fictie en werkelijkheid*, georganiseerd door de Belgische Vereniging voor Documentatie op 28 april 2005 te Brussel.

▪ Les nouvelles applications de gestion d'information, la mondialisation de la société de l'information, la vaste gamme de ressources disponibles sur Internet et le rêve de créer un Web sémantique : toutes ces évolutions ont suscité de nouveaux besoins et ont fortement changé la donne pour les langages documentaires. De nouveaux types d'outils ont donc été développés, telles les taxonomies et les ontologies. L'auteur explique quelle est l'influence de ces tendances sur les activités de développement ou de révision des normes nationales et internationales et nous détaille quatre efforts de mise à jour en cours de réalisation.

▪ Nieuwe toepassingen voor het beheren van informatie, de mondialisering van de informatiemaatschappij, het vaste gamma van bronnen beschikbaar op het Internet en de droom een semantisch Web te creëren: al deze evoluties hebben nieuwe noden doen ontstaan en hebben sterk de kaart van de documentaire talen door mekaar gehaald. Nieuwe soorten instrumenten zijn daardoor tot stand gekomen, zoals de taxonomie en de ontologie. De auteur verduidelijkt welke de invloed van deze tendensen is op de ontwikkelingsactiviteiten of op de herziening van nationale en internationale normen en geeft ons een gedetailleerd overzicht ivm de update van vier nog in uitvoering zijnde inspanningen

I l y a dix ans, on se demandait si les langages documentaires n'étaient pas dépassés : la recherche en langage naturel était de plus en plus performante et de plus en plus utilisée. La quantité d'information disponible en format électronique, soit sur Internet, soit au sein d'une organisation, ne cessait de croître. Les coûts, inévitablement élevés, de l'indexation manuelle suscitaient naturellement des questions concernant la faisabilité économique de cette procédure. Tout portait à croire, à l'époque, que les classifications, les vedettes-matières et les thésaurus étaient voués à un avenir sombre et une vie brève.

Cependant, depuis quelques années, les langages documentaires connaissent un regain d'intérêt aussi surprenant que bienvenu. Il est beaucoup question de nouveaux types de langages documentaires comme les taxonomies. Les projets de création de liens entre différents langages documentaires se multiplient afin de faciliter la recherche dans toute une gamme de bases de données différentes. Et, de plus en plus, de nouvelles applications informatiques intègrent dans leurs fonctionnalités le recours à un langage documentaire.

En peu de temps, la perception et les attentes en matière de langages documentaires ont donc fondamentalement changé. Pourquoi ? Qu'est-ce qui a évolué entre les années 90 et aujourd'hui

pour que notre approche ait tellement changé ? Comment devons-nous réagir à ces changements ? Passons d'abord en revue les tendances dont l'influence est très importante sur les langages documentaires pour ensuite décrire quelques initiatives de développement de nouvelles normes ou de révision de normes existantes dans ce domaine.

Les tendances actuelles

En ce qui concerne les tendances actuelles, on peut identifier au moins cinq facteurs importants :

Le problème de la gestion d'un volume croissant d'informations

La plupart des sites Web se sont développés sans plan préétabli. On crée un site, on ajoute des pages, et un beau jour, ce sont des milliers de pages d'information qui se retrouvent à disposition sans qu'il y ait de véritable cohérence entre elles. Dans le même ordre d'idées, on commence à stocker des documents en format électronique et quelques années plus tard, on se rend compte que l'on a créé un entrepôt d'informations incohérentes, sans structure logique adéquate. Par conséquent, le plus souvent, les utilisateurs se perdent dans les méandres de

vastes sites mal organisés, demeurent perplexes face à l'énorme quantité de résultats fournis par un moteur de recherche sous la forme d'une liste de références dont la plupart ne semblent pas du tout pertinentes. Ces difficultés de navigation ou de recherche ne sont pas sans conséquences; par exemple :

- les employés d'une société perdent leur temps;
- les utilisateurs externes doivent demander l'intervention d'un spécialiste au lieu de trouver la solution à leur problème eux-mêmes;
- les clients de sites commerciaux s'orientent vers d'autres sites plus simples.

Vient donc le moment charnière, la crise - c'est presque inévitable - qui pousse à agir. On fait appel aux spécialistes comme les bibliothécaires, les documentalistes ou les architectes d'information qui par leur formation et leur expérience professionnelle se sont familiarisés avec les systèmes de classification et les thésaurus. Mieux structurer l'information, employer les métadonnées pour décrire des documents, utiliser un plan global et une approche cohérente dans l'ensemble du site, accorder un soin tout particulier aux termes employés (surtout dans les menus et les barres de navigation) : autant de moyens permettant de réduire les délais de recherche d'information et d'éviter la frustration de l'utilisateur. Outils parmi d'autres, les classifications simples appelées "taxonomies" sont utilisées pour mieux structurer les sites, tandis que les thésaurus sont utiles pour améliorer l'efficacité de la recherche. Les langages documentaires utilisés dans ce type d'environnement ne sont peut-être pas exactement pareils à ceux qui ont cours dans un centre de documentation traditionnel, mais ils y ressemblent.

Nouveaux types d'applications

Parallèlement à ce besoin d'une meilleure structuration des sites, plusieurs nouvelles applications informatiques de gestion de l'information au niveau de l'entreprise ont vu le jour. Elles concernent la gestion de contenu, les portails et la gestion électronique de documents. Les langages documentaires constituent une partie importante de ces nouveaux services.

Par exemple, le fait d'exploiter une taxonomie ou une classification est devenu un point fort de ces nouveaux logiciels. Certains systèmes comme Documentum¹ sont livrés avec une taxonomie de base consacrée à une application d'entreprise. D'autres, comme OpenText², permettent de restreindre une recherche à un certain domaine ou sous-domaine, défini par une structure logique. D'autres encore, comme Verity³, offrent à l'utilisa-

teur la possibilité de choisir entre différents concepts permettant au logiciel de lancer une recherche prédéfinie sur texte intégral. Ces termes doivent être présentés aux utilisateurs d'une manière logique, c'est-à-dire via une classification simple.

Les langages documentaires sont également une composante essentielle du commerce électronique. Une société qui vend toute une gamme de produits sur Internet doit créer une classification desdits produits qui soit simple, claire et permette de multiples approches différentes, en fonction du point de départ de l'utilisateur. Dans leurs efforts visant à réduire les coûts et à améliorer les services en mettant en œuvre ce que l'on appelle le gouvernement électronique, les administrations des États sont confrontées aux mêmes nécessités. En Grande Bretagne, par exemple, le développement d'une simple taxonomie, la Government Categories List (la liste des catégories gouvernementales)⁴ a été considéré comme essentiel pour la promotion de services gouvernementaux par la voie du Web.

Mondialisation et multilinguisme

D'autres changements ont également eu un impact sur la manière dont on traite les langages documentaires, par exemple, la mondialisation de notre société de l'information et par conséquent, l'importance grandissante du multilinguisme. Bien sûr, les thésaurus ou les classifications multilingues existent depuis au moins une bonne trentaine ou quarantaine d'années, mais ils n'étaient employés que dans certaines situations pointues et spécialisées. Auparavant, les thésaurus multilingues ne concernaient que les institutions de pays présentant plus d'une langue officielle comme la Belgique ou le Canada, ou encore certaines institutions internationales comme les organisations du système des Nations Unies ou les institutions européennes. De nos jours la disponibilité d'Internet et du Web partout dans le monde, fait de cette fonctionnalité multilingue une condition sine qua non pour toute organisation désireuse de diffuser son information au-delà de ses frontières linguistiques.

Par exemple, une société commerciale ayant des antennes et du personnel dans différents pays, devra impérativement disposer d'un système de gestion de documents prévoyant l'interrogation en plusieurs langues. Les sites Web des organisations à but non lucratif sont visités par quiconque dans le monde, chaque visiteur correspondant à un donateur ou à un parrain potentiel. Dans le domaine de l'information proprement dit aussi, les bibliothèques

nationales et les grandes universités collaborent pour desservir un public qui est de plus en plus international. Le recours aux langages documentaires multilingues permet aux utilisateurs de s'exprimer dans différentes langues et d'utiliser un site d'une manière cohérente, efficace et conviviale.

Correspondances entre différents langages documentaires

Le multilinguisme évoqué plus haut est un multilinguisme qui s'applique aux langues naturelles. Il permet d'interroger une base de données en ayant recours, par exemple, à des termes anglais, français ou néerlandais selon la langue maternelle de l'utilisateur. Mais les questions liées au multilinguisme ne s'arrêtent pas aux langues naturelles. Les problèmes du multilinguisme concernent également l'utilisation des modèles conceptuels que sont les langages documentaires. Un utilisateur qui maîtrise un thésaurus ou un système de vedettes-matières voudra naturellement utiliser les termes repris dans ce langage documentaire pour interroger une brochette de bases de données disponibles par le biais d'Internet, y compris celles qui sont indexées dans d'autres langages. Pour répondre à ces attentes, il faut donc prévoir la possibilité de traduire un concept rencontré dans un langage documentaire vers un autre provenant d'un langage différent.

Par exemple, un chercheur en agriculture peut vouloir interroger la base AGRIS de la FAO à Rome, les bases du CABI (ancien Commonwealth Agricultural Bureau) en Angleterre et la base AGRICOLA de la National Agricultural Library aux États-Unis. Celles-ci sont disponibles en anglais mais chaque base est indexée avec un langage documentaire différent. Un chercheur, même anglophone, doit reformuler sa recherche trois fois et ré-exprimer ses idées en trois langages documentaires différents pour arriver aux résultats souhaités.

Pour remédier à cette incohérence, le multilinguisme doit exister au niveau des langages documentaires eux-mêmes. Il faut établir des liens ou des correspondances (ce qu'on appelle des "mappings" en anglais) entre un concept exprimé dans un langage documentaire et un concept similaire exprimé dans un autre langage documentaire, pour qu'une recherche puisse être traduite automatiquement.

Toile sémantique

La dernière tendance en matière d'évolution des langages documentaires est la Toile sémantique ou Semantic Web. Qu'est-ce là donc ? Et quel est le lien entre ce projet et les langages documentaires ?

Le Web, qui nous est familier depuis le début des années 90, dépend des capacités de l'individu de lire et de comprendre des documents trouvés et d'agir en conséquence. En revanche, le Web sémantique tel qu'expliqué par Tim Berners-Lee⁵ repose plutôt sur l'idée d'informations traitées non pas par des personnes mais par des applications informatisées. Ces applications (appelées "agents"), pourraient se substituer à l'homme pour accomplir de nombreuses tâches onéreuses.

Pour illustrer le fonctionnement de la Toile sémantique, Berners-Lee se sert d'un exemple tout à fait banal. Il prend le cas de quelqu'un qui veut régler avec sa sœur tous les détails liés à l'accompagnement de leur mère lors d'une série de consultations médicales :

- choisir un prestataire de services de qualité dans un endroit spécifique;
- prendre rendez-vous pour une série de traitements en tenant compte d'autres obligations, et
- établir les meilleurs itinéraires pour se rendre à ces rendez-vous en tenant compte de la circulation routière à différents moments de la journée.

Tout ce travail de planification serait exécuté par le Web sémantique avec un minimum d'intervention de la part de la personne.

Ceci ne sera possible que dans la mesure où existeront des langages normalisés qui définissent d'une manière claire et nette les concepts de différents domaines. Par exemple, il faut une seule et unique façon d'exprimer des idées comme *traitement, histoire médicale, prestataire de services, adresse, heures d'ouverture, rendez-vous, disponibilité, routes, et heures de pointe*. Ces concepts devraient être exprimés d'une manière explicite pour permettre leur traitement par ordinateur. Et ces concepts devraient être liés entre eux en fonction de leur signification, c'est-à-dire avec des liens sémantiques. Un langage qui exprime des concepts dans un certain domaine et leurs relations sémantiques avec d'autres concepts s'appelle une ontologie.

Évidemment la création de toutes ces ontologies, ces modèles conceptuels normalisés, représente une tâche titanesque. Mais il n'a pas échappé à

Berners-Lee et à ses collègues du World Wide Web Consortium qu'existent déjà des langages se composant de concepts et établissant certains liens sémantiques entre ceux-ci : il s'agit des langages documentaires, en particulier les thésaurus. Toutefois, l'idée serait de créer des liens non seulement entre les termes à l'intérieur d'un seul thésaurus, mais aussi entre termes qui se trouvent dans différents thésaurus spécialisés dans des domaines différents. On imagine aisément la vaste toile sémantique qui en résulterait et les immenses possibilités de navigation et de recherche encore insoupçonnées qu'elle offrirait.

Ce concept de Toile sémantique suscite énormément d'intérêt et des sommes importantes ont été allouées à la recherche dans ce domaine en Europe et aux États-Unis. Entre 2002 et 2004, le groupe de recherche SWAD-Europe (Semantic Web Advanced Development in Europe)⁶, subventionné par l'Union Européenne, a travaillé sur plusieurs dossiers dont l'un est justement consacré aux thésaurus. L'idée de base de cette dernière activité est d'établir les méthodes nécessaires à la présentation du contenu d'un thésaurus en utilisant une structure normalisée et des méthodes compatibles avec la Toile sémantique.

Même si un certain scepticisme est de mise quant à la matérialisation du rêve de la Toile sémantique, ce travail s'avère fort intéressant. Une présentation normalisée des langages documentaires sur la Toile faciliterait l'interopérabilité entre eux, non seulement pour les tâches sophistiquées imaginées par Berners-Lee mais aussi pour la "simple" recherche d'information. Prenez par exemple un moteur de méta-recherche tel qu'en offrent les produits de portail documentaire actuellement sur le marché. En appliquant un protocole normalisé avec un format standard pour la description d'un concept, on peut lier un concept exprimé dans un langage documentaire avec un concept identique (ou presque identique) exprimé dans un autre langage. Cet effort facilitera les recherches simultanées, car des applications informatiques traduiront une recherche exprimée dans le langage d'une base de données vers un autre langage employé par un autre système. Des normes pour le transfert d'informations stockées dans un thésaurus faciliteraient aussi leur échange entre l'institution ayant développé le thésaurus et les organisations qui l'utilisent déjà ou qui veulent le traduire ou l'adapter.

Ces cinq tendances ont eu un effet directeur sur le développement des nouvelles normes pour les langages documentaires. Vu leur importance dans la création, la gestion et l'application des langages documentaires, plusieurs équipes travaillent actuellement à la mise à jour des normes

les plus importantes dans ce domaine. Ce sont des normes internationales et nationales concernant la conception et la gestion des thésaurus. Quatre projets en cours de réalisation pourraient s'avérer très importants.

Activités du groupe SWAD-Europe

Les activités en termes de langages documentaires de l'équipe de recherche du World Wide Web Consortium SWAD-Europe ont été regroupées sous le sigle SKOS. SKOS est l'acronyme anglais de *Simple Knowledge Organization Systems, knowledge organization system* étant l'expression courante en anglais pour un langage documentaire, un concept qui regroupe les systèmes de classification, les taxonomies, les listes de vedettes-matières et les thésaurus. Mais pourquoi ajoute-t-on le mot "simple"? Justement parce que ces langages documentaires sont plus simples qu'une ontologie, et que les ontologies sont au cœur du travail du World Wide Web Consortium pour la construction de la Toile sémantique.

Le produit le plus important de cette recherche est une méthode pour publier un thésaurus électronique sur la Toile qui s'appelle SKOS Core. SKOS Core⁷ est une façon systématique de décrire les concepts et les relations d'un thésaurus de manière à ce qu'ils soient compatibles avec les besoins de la Toile sémantique. Cette méthodologie peut s'appliquer également aux autres langages documentaires. SKOS Core utilise les technologies XML⁸ et RDF⁹ qui fournissent la structure de base de la Toile sémantique. Parce que SKOS Core répond aux besoins de la Toile sémantique, cette méthodologie pourrait jouer un rôle important dans la création d'une nouvelle génération d'outils Internet.

L'activité de recherche SWAD proprement dite est maintenant terminée mais les principaux promoteurs du projet continuent à oeuvrer pour sensibiliser le public au travail accompli, pour encourager les autres à adopter ces méthodes et pour faire en sorte que ce travail débouche sur une recommandation du World Wide Web Consortium et devienne une norme au même titre que les autres recommandations de ce consortium. Si le lecteur s'intéresse aux langages documentaires, c'est un dossier qu'il devra suivre de près¹⁰.

Lignes directrice d'IFLA

Le deuxième développement dont il est question ici n'est pas une norme proprement dite, mais plutôt un document faisant référence à une norme. Un groupe, au sein de l'association IFLA, a rédigé un document¹¹ qui définit certaines lignes directrices susceptibles de constituer une sorte

d'annexe à la norme ISO pour les thésaurus multilingues. En préparation depuis 2002, il a été publié de manière officielle au printemps 2005. Ses recommandations se fondent sur deux axes différents.

D'abord, on prévoit la possibilité d'un thésaurus multilingue non symétrique, c'est-à-dire un thésaurus ne comportant pas forcément un descripteur dans chaque langue pour chaque concept. Au lieu de cette correspondance stricte, on peut créer des équivalences inexactes ou partielles pour établir un lien entre différents concepts.

Le deuxième but de ce document est de concrétiser les expériences acquises lors du projet MACS¹². MACS, pour *Multilingual Access to Subjects*, visait à créer des correspondances entre plusieurs systèmes de vedettes-matières, en l'occurrence des listes de vedettes-matières en allemand, en français et en anglais. Le but de cette activité était de permettre à un usager unilingue d'interroger les bases bibliographiques des bibliothèques nationales en Suisse, en Allemagne, en France et en Grande-Bretagne et de retrouver dans toutes ces bases des notices pertinentes. MACS était donc un projet visant l'interopérabilité de listes de vedettes-matières qui n'étaient pas conçues comme une structure multilingue.

Révision de la norme NISO

Le troisième effort concerne un groupe de travail mis sur pied par le NISO, l'organisation de normalisation dans le domaine de l'information aux États-Unis. Ce groupe de travail a revu la norme américaine pour les thésaurus unilingues et a publié une mise à jour de cette norme en avril 2005¹³. Certaines parties de cette norme ont beaucoup évolué depuis la dernière version, et présentent :

- une explication théorique de certains concepts;
- l'interopérabilité entre des langages documentaires;
- les conseils en matière de test set d'évaluation des thésaurus, et
- des recommandations concernant des systèmes de gestion de thésaurus.

Ce document représente une contribution importante à la mise à jour générale de la compréhension des thésaurus.

Révision des normes britanniques

Enfin, au British Standards Institute en Grande-Bretagne, un groupe travaille depuis à peu près deux ans sur une révision en profondeur des

normes pour la gestion des thésaurus unilingues et multilingues. Le groupe compte dans ses rangs certaines des personnalités les plus reconnues et faisant autorité dans ce domaine : Stella Dextre Clarke, Leonard Will et Alan Gilchrist (l'auteur y participe également en tant qu'expert associé). Le but de cette révision est de clarifier et de mettre à jour les recommandations faites il y a quinze ans, mais aussi de les modifier afin de prendre en compte les nouvelles tendances que l'on vient d'évoquer. Les changements les plus importants sont :

- *Une révision de la terminologie*
On essaie de revoir et de clarifier un certain nombre de concepts, notamment ceux qui sont liés à l'application du principe des facettes au développement du thésaurus.
- *La prise en compte d'autres langages documentaires, p. ex. les classifications et les taxonomies.*
Depuis longtemps, on a reconnu que les thésaurus avaient beaucoup de points en commun avec les classifications et d'autres outils d'organisation des connaissances comme les listes de vedettes-matières et les taxonomies. Maintenant que l'interopérabilité de ces différents outils est envisagée, il faut identifier d'une manière plus précise, quels sont leurs différences et leurs points communs afin de pouvoir choisir l'outil approprié et aussi d'évaluer l'effet de ces différences sur la recherche d'information. On veut également expliquer comment ces langages documentaires pourraient être utilisés dans toute une gamme d'applications, que ce soit pour naviguer dans une base ou pour effectuer une recherche.
- *Le traitement des correspondances ("mappings") entre concepts dans différents systèmes, y compris les différentes versions linguistiques d'un thésaurus multilingue.*
Dans le passé, il existait une norme pour les thésaurus unilingues et une autre pour les thésaurus multilingues. Dorénavant, il n'y aura qu'une seule norme traitant d'une manière plus claire et plus logique la création de trois types de relations d'équivalence :
 - des équivalences entre des descripteurs et non-descripteurs dans une seule langue comme dans un thésaurus unilingue, c'est-à-dire les relations de type *Employer* et *Employé pour*;
 - des équivalences entre les termes dans un langage naturel avec les termes dans une autre langue comme dans un thésaurus multilingue, et
 - des équivalences entre les concepts d'un thésaurus et les concepts d'un autre lan-

gage documentaire dans un domaine connexe.

Ces correspondances comprendront des équivalences inexactes ou partielles où l'on ne peut rencontrer un terme ayant le même sens dans une langue et dans une autre.

- *La mise à jour de ces normes pour tenir compte des pratiques courantes dans la gestion, l'exploitation et l'échange des thésaurus électroniques*

Les bases de données documentaires et les systèmes de gestion de langages documentaires sont tous informatisés et presque tous disponibles sur le Web. Cela permet d'utiliser plusieurs applications en même temps ou de naviguer d'une application et sa base vers une autre base ou une autre application. Il est aussi possible de transférer les thésaurus d'un système de gestion vers un autre avec le même degré de fiabilité que lors du transfert de notices bibliographiques d'un progiciel vers un autre. Mais il faut des lignes directrices pour la gestion des thésaurus, et des normes bien définies pour l'interopérabilité de différents systèmes disponibles sur la Toile.

La nouvelle norme *BS 8723 Structured vocabularies for information retrieval* sera structurée en cinq parties. La première, très importante, donne une définition de tous les termes employés, et proposera une définition plus claire de certains concepts dans le domaine.

La deuxième partie traite des principes de construction et des caractéristiques particulières d'un thésaurus.

La troisième partie abordera les différences et les points communs entre différents types de langages documentaires, y compris les classifications, les taxonomies, les listes de vedettes-matières, les thésaurus conçus pour faciliter la recherche sur texte intégral et même éventuellement les ontologies. Le but de cette partie n'est pas de décrire en détail tout ce qu'il y a lieu de faire en créant un de ces autres langages documentaires, mais d'en faire comprendre les caractéristiques, car les particularités sont très importantes si l'on veut créer un lien entre, par exemple, un descripteur dans un thésaurus, une rubrique dans une liste de vedettes-matière et une classe dans un système de classification.

La quatrième partie traite des correspondances entre différents vocabulaires, y compris les différentes versions linguistiques d'un thésaurus multilingue.

Et finalement, la cinquième partie porte sur l'interopérabilité des langages documentaires et des

applications informatiques– les applications qui gèrent ces langages et les applications qui les exploitent. Dans cette dernière partie, on espère notamment préciser les formats et les protocoles qui permettront d'exploiter un thésaurus d'une manière efficace.

Les deux premières parties ont été publiées en novembre 2005¹⁴, la quatrième sera disponible pour commentaires au cours de l'année suivante. Quant à la troisième, elle est à l'état de brouillon et le groupe de travail est sur le point d'aborder, avec l'aide de plusieurs experts, la cinquième et dernière partie, qui porte sur des formats et protocoles électroniques. Même si le groupe est constitué au niveau national en Grande-Bretagne, la norme britannique vise les utilisateurs de thésaurus à une échelle internationale et le groupe attend des commentaires en provenance du monde entier. Si vous voulez participer à cette entreprise, n'hésitez pas à prendre contact avec l'auteur qui se fera un plaisir de vous expliquer comment procéder pour faire part de vos commentaires.

Conclusion

Les langages documentaires sont bien vivants. Le défi visant à mieux organiser l'information sur le Web a donné vie à de nouveaux outils comme les taxonomies et a insufflé une bouffée d'oxygène aux systèmes de classification et aux thésaurus. Une approche architecturale de la conception des sites de la Toile et l'intégration des taxonomies à plusieurs nouveaux types d'applications informatiques nous ont montré comment ces instruments peuvent être utiles en dehors d'une bibliothèque ou d'un centre de documentation traditionnels. La possibilité d'exploiter à l'avenir une interopérabilité sémantique a stimulé l'innovation et la recherche, qui ont retrouvé un niveau comparable à celui qu'elles ont connu dans les années 60. Il est peu probable que les langages documentaires d'aujourd'hui deviennent le latin ou le sanscrit de l'ère moderne, rarement utilisés et connus uniquement de certains initiés. Au contraire, le fait que les langages documentaires changent et évoluent en fonction de leur environnement et de leurs utilisateurs est la parfaite démonstration qu'ils sont aussi vivants et importants pour la communication que les langues vivantes que nous parlons tous les jours.

Ron DAVIES

Av. Baden-Powell, 1 Bte 2
1200 Bruxelles
ron@rondavies.be

31 octobre 2005

Notes

- 1 <<http://www.documentum.com>> (consulté le 10 novembre 2005).
- 2 <<http://www.opentext.com>> (consulté le 10 novembre 2005).
- 3 <<http://www.verity.com>> (consulté le 10 novembre 2005).
- 4 La CGL était fusionné avec d'autres vocabulaires dans le Integrated Public Sector Vocabulary (IPSV). Le rôle d'une simple taxonomie est maintenant assuré par une forme abrégée de l'IPSV. Dextre Clarke, Stella. *IPSV Maintenance Guide*. London: Porism, 2005. Disponible à : <<http://www.esd.org.uk/documents/IPSVMaintenanceGuide.pdf>> (consulté le 10 novembre 2005).
- 5 Berners-Lee, Tim; Hendler, James; Lassila, Ora. The semantic Web, *Scientific American*, 2001, vol. 284, no. 5, pp 34-43. <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>> (consulté le 10 novembre 2005).
- 6 SWAD-Europe <<http://www.w3.org/2001/sw/Europe>>
- 7 Mikes, Alistair; Brickley, Dan (eds). *SKOS Core Guide*. (W3C Working Draft 2 November 2005); <<http://www.w3.org/TR/swbp-skos-core-guide>> (consulté le 10 novembre 2005).
- 8 Extensible Markup Language (XML) <<http://www.w3.org/XML>> (consulté le 10 novembre 2005).
- 9 Resource Description Framework (RDF) <<http://www.w3.org/RDF>> (consulté le 10 novembre 2005).
- 10 <<http://www.w3.org/2004/02/skos/>> (consulté le 10 novembre 2005).
- 11 Working Group on Guidelines for Multilingual Thesauri. *Guidelines for Multilingual Thesauri*. [s.l.]: IFLA, 2005. <<http://www.ifla.org/VII/s29/pubs/Draft-multilingualthesauri.pdf>> (consulté le 10 novembre 2005).
- 12 Multilingual Access to Subjects (MACS) <<https://ilmacs.uvt.nl/pub>> (consulté le 10 novembre 2005).
- 13 National Information Standards Organization. Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies. Bethesda, Md. NISO, 2005. (ANSI Z 39-19 - 20005). <http://www.niso.org/standards/standard_detail.cfm?std_id=814> (consulté le 10 novembre 2005).
- 14 *British Standards Institute. Structured Vocabularies for Information Retrieval. Guide. Definitions, Symbols and Abbreviations*. London: BSI, 2005 (BS 8723-1:2005) ISBN 0 580 46798 8.

LES BASES DE DONNÉES SPATIALES

Jean-Paul DONNAY

Professeur ordinaire, Université de Liège – Unité de Géomatique

▪ Derrière la "carte", média le plus commun pour accéder à l'information spatiale, il y a des données, qu'il a fallu collecter, mémoriser et gérer, sous forme numérique depuis près d'un demi-siècle. Les données, spatiales ou géographiques, présentent plusieurs particularités dont la moindre n'est certes pas la présence, parmi ses caractéristiques, d'une géométrie censée traduire la portion de la surface terrestre portant une ou plusieurs propriétés thématiques. Indépendamment des techniques de collecte de l'information géographique (topographie, photogrammétrie, etc.), des techniques de visualisation (cartographie) et de traitement (analyse spatiale), la seule gestion de l'information a connu une évolution sensible au cours des dernières décennies. Partant de systèmes de fichiers, la gestion de l'information spatiale a successivement fait appel à des modèles de gestion de bases de données de plus en plus élaborés, au point d'exploiter aujourd'hui les ressources les plus pointues des bases de données réparties reposant sur un modèle objet-relationnel.

▪ Achter "de kaart", het meest gebruikte medium om de ruimte-informatie te bereiken, zijn er gegevens, die men heeft moeten verzamelen, opslaan en, onder digitale vorm beheeren sedert bijna een halve eeuw. Deze gegevens ivm aardrijkskunde en ruimte, presenteren verschillende bijzonderheden waarvan de minste zeker niet de aanwezigheid is van een meetkunde die verondersteld wordt de portie van de landoppervlakte weer te geven dat één of meer thematische eigenschappen in zich draagt. Los van de techniek van het verzamelen van geografische informatie (topografie, fotogrammetrie, enz), van de visualiseringstechnieken (cartografie) en aanpak (ruimtelijke analyse), heeft alleen al het beleid van de informatie een belangrijke evolutie in de loop van de laatste decennia gekend. Uitgaande van bestandensystemen, heeft het beheer van de ruimte-informatie achtereenvolgens op steeds meer uitgewerkte modellen van databasebeheer een beroep gedaan, om alzo vandaag de dag de meest betrouwbare bronnen van de zeer geavanceerde gegevensbanken te exploiteren berustende op het model "object-relationeel".

Pour la majorité des utilisateurs, l'accès à l'information géographique passe par la consultation d'une carte : la carte topographique du promeneur, la carte routière – imprimée ou numérique et affichée sur l'écran du système de navigation – du conducteur, ou le plan de ville du touriste. Cartes et plans ont en commun qu'ils sont une représentation conventionnelle, généralement plane, à une échelle donnée, des objets, concrets (matérialisés et visualisables), ou abstraits (non matériels, telles les limites administratives), liés à une partie de la surface terrestre. On voit déjà se dégager quelques caractéristiques essentielles de l'information géographique : elle concerne des "objets" identifiés et localisés. Mais la carte n'en fournit qu'une représentation "conventionnelle" : tous les objets ne sont pas représentés (la forêt mais pas les arbres) et ceux qui sont sur la carte sont, à la fois géométriquement simplifiés et figurés par des symboles graphiques variables en taille, couleur, forme, etc., selon une légende explicite. Les opérations de généralisation (en première analyse : la simplification) et de symbolisation qui président à la réalisation de la carte sont fonction de son échelle et de son objectif.

Puisque la carte ne consiste qu'en une représentation des données géographiques, il faut admettre qu'en amont de la carte existe une collection de données qui doivent faire l'objet d'une acquisition, d'une sauvegarde, d'une gestion et d'une mise à jour. Idéalement, ces données sont indépendantes de toute représentation, c'est-à-dire

de toute notion d'échelle et de toute symbolisation. C'est en cela que les données "géographiques" se distinguent des données "cartographiques". Les premières enregistrent le plus précisément possible, tant la géométrie (en première approximation, l'emprise au sol), que les qualités (nature, fonction, etc.) des objets sélectionnés dans la collection. La nature duale de l'information géographique, à la fois géométrique et sémantique ou thématique, confère à chaque opération un caractère original et, souvent, une complexité supplémentaire vis-à-vis des données alphanumériques traditionnelles. À titre d'exemple, la seule acquisition de la géométrie d'un objet a recours à des techniques spécifiques, telles que le levé topographique ou photogramétrique, afin de fixer les coordonnées des points – dans un référentiel adéquat – qui délimitent l'emprise au sol de l'objet.

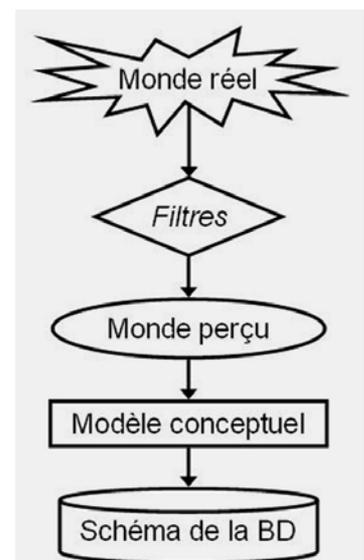


Fig. 1 : Du monde réel au schéma de la base de données géographiques, l'information subit une réduction et une formalisation croissantes.

L'information géographique

Encore faut-il savoir quels sont les objets du terrain susceptibles d'alimenter la collection de données géographiques. Outre le fait que nous ne percevons, naturellement ou techniquement, qu'une partie du monde réel qui nous entoure, il est indispensable de constituer un modèle de la réalité pour laquelle nous souhaitons des données (Fig. 1). Cette démarche, d'ordre concep-

- le type de géométrie (point, ligne ou polygone) correspondant à son emprise au sol;
- et le domaine de variation des attributs thématiques qui lui sont associés (intervalle de variation, unité, modalités permises, etc.).

Il faut aussi vérifier que l'organisation des données sélectionnées est cohérente et compatible avec l'usage qui en sera fait par les applications. Par organisation des données, il faut entendre les relations qu'elles entretiennent, du point de

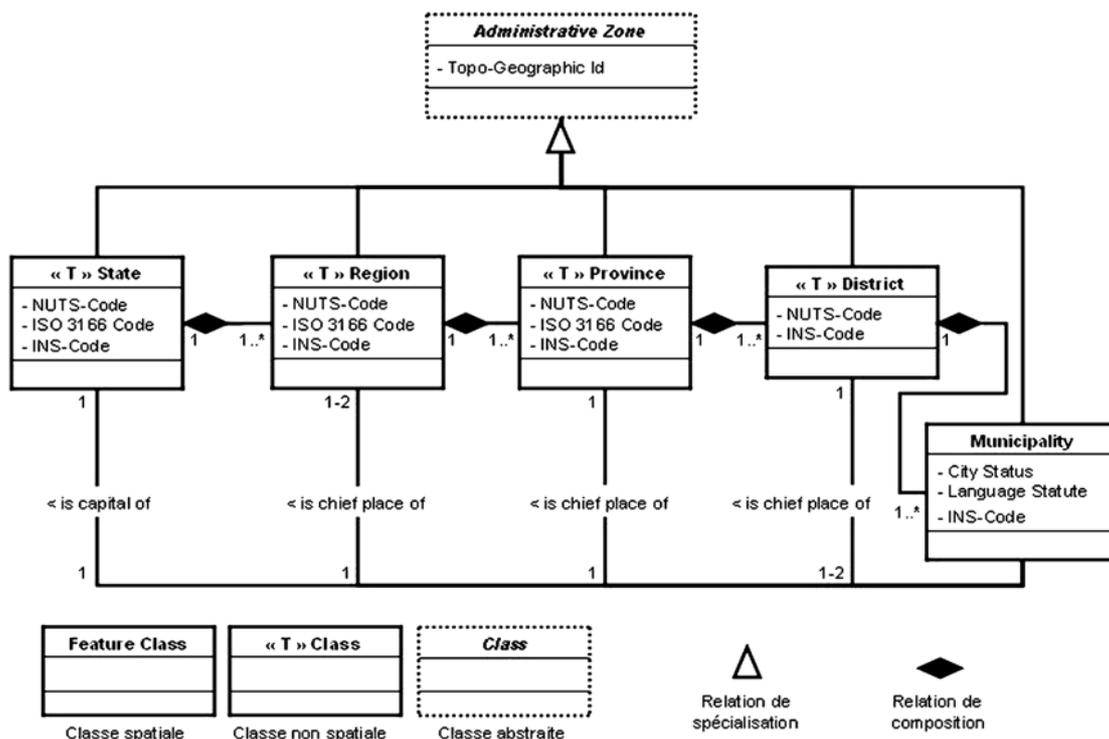


Fig. 2 : Modèle conceptuel de données formalisé en UML (Unified Modelling Language).
Thème des entités administratives, adapté du projet SGIS de l'IGN-B (3)

tuel, s'appuie sur l'usage pressenti des données, afin de déterminer les critères de sélection et la précision / résolution souhaitée. Étant donné la nature duale de l'information géographique, critères de sélection et précision sont à déterminer tant du point de vue géométrique que du point de vue sémantique. À titre d'exemple, faut-il s'attarder à la définition de chaque "arbre" pris isolément, fixé en position avec une précision décimétrique et caractérisé par son espèce, son âge, sa hauteur, etc. ; ou la notion de "forêt" est-elle suffisante, délimitée spatialement avec une précision métrique, et caractérisée par la seule espèce dominante ? La réalisation d'un tel catalogue de données géographiques, éventuellement accompagné d'un glossaire exhaustif, constitue une première étape au cours de laquelle sont définis, pour chaque type d'entités géographiques (chaque classe d'objets géographiques) :

vue logique et spatial. Les relations logiques sont celles que l'on peut trouver entre tous les types de données : l'association (ex. une capitale est associée à un pays), la composition (ex. un groupe de communes compose une province), la spécialisation (ex. une autoroute est une catégorie spécifique des voiries). Les relations spatiales sont caractéristiques des données géographiques. On a coutume de ne retenir que les relations topologiques, c'est-à-dire, en dernière analyse, les relations qui identifient le voisinage et la superposition entre les entités géographiques : à titre d'exemples, deux pays sont voisins à travers un élément de frontière commune ; une route peut se superposer à une rivière à l'endroit d'un pont. Le fait d'identifier et de conserver les relations logiques et topologiques entre les données géographiques, permet de mieux contrôler la qualité des données et autorise l'expression de requêtes plus élaborées sur celles-ci. Ainsi, les requêtes qualifiées de "spatiales" exploitent la

géométrie et / ou les relations topologiques des données géographiques : ex. trouver les communes traversées par une portion de route ; identifier les sites archéologiques localisés à moins de x kilomètres d'un site de fouille.

Tant les spécifications géométriques et sémantiques des données géographiques que les relations logiques et topologiques qu'elles entretiennent entre elles, peuvent être formalisées dans un langage graphique, afin de constituer un modèle conceptuel de données (Fig. 2). Ce modèle est indépendant de toute contrainte de logiciel et de matériel, mais il préfigure le schéma de la base de données géographiques qui doit être implémentée (1), (2).

La gestion des données géographiques

La difficulté de gestion des données géographiques provient de leur double nature, géométrique et thématique. Si l'on imagine aisément l'enregistrement d'attributs alphanumériques quelconques, caractérisant la thématique, dans une simple table (relation d'une base de données relationnelle, ou à la rigueur une simple feuille de calcul d'un tableur), il n'en va pas de même des coordonnées spatiales et des relations géométriques entre les entités. Les solutions envisagées pour résoudre cette difficulté ont évolué au cours du temps, suivant en cela les générations successives des modèles logiques des bases de données et les capacités croissantes des ressources informatiques. Les changements les plus importants étant apparus durant la dernière décennie, des solutions relevant de générations distinctes continuent à coexister aujourd'hui. Les logiciels informant ces solutions sont intitulés, un peu abusivement (cf. infra), "systèmes d'information géographique" (4).

Architectures hybrides

Pendant longtemps, le stockage et la gestion des données géographiques ont exploité des solutions dites "hybrides", c'est-à-dire traitant de façon distincte la géométrie et la sémantique. Le modèle le plus simple consiste à utiliser un système de gestion de base de données relationnel (SGBD-R) pour les attributs alphanumériques décrivant la thématique, et, en parallèle, un système de gestion de fichiers (SGF) pour les géométries des entités géographiques. Les attributs sont donc stockés dans des tables selon un schéma relationnel classique, et sont susceptibles d'être traités (mises à jour, requêtes) par le langage de transaction normalisé SQL (Structured Query Language), pourvu que ces transac-

tions ne portent que sur la thématique. Les coordonnées des points fixant les géométries des entités géographiques sont enregistrées dans des fichiers de formats distincts selon les géométries, de manière à faciliter leur accès et leur gestion. À cet égard, cependant, les différents producteurs de logiciels présentent des solutions informatiques différentes et les formats de fichiers sont qualifiés de "propriétaires" dans la mesure où l'accès aux données est strictement dépendant du logiciel utilisé. C'est également ce logiciel qui prend en charge l'association entre une géométrie particulière et les valeurs d'attributs correspondantes, reconstituant ainsi l'information complète relative à une entité géographique. Cette association utilise un "identifiant géographique" commun pour la géométrie en question et l'enregistrement correspondant dans la table d'attributs (Fig. 3). Le logiciel étant capable de reconstituer l'information géographique complète, il peut la soumettre à différents types de traitements relevant de l'analyse spatiale et/ou de la cartographie, mais on constate que les traitements sont dépendants des formats des données.

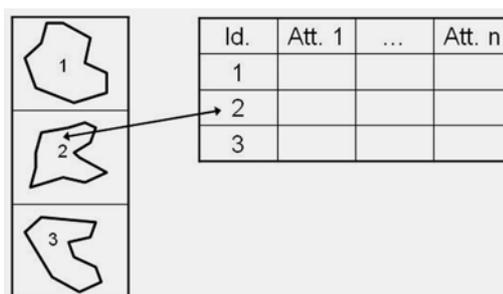


Fig. 3 : Architecture hybride. À gauche : le fichier des géométries. À droite : la table des attributs. La liaison passe par l'identifiant commun (ici, un numéro) des géométries et des enregistrements.

Les systèmes hybrides les plus simples ne conservent pas, et par conséquent ne peuvent exploiter, les relations entre les entités géographiques, en particulier les relations spatiales. Par contre, les systèmes dits "géo-relationnels" constituent une avancée dans ce domaine, en ce qu'ils identifient, conservent et permettent à l'utilisateur d'exploiter les principales relations topologiques entre entités géographiques. Les géométries et les attributs sont toujours gérés par des sous-systèmes distincts, respectivement un SGF et un SGBD-R, mais en plus, sur base des coordonnées, le logiciel identifie les relations spatiales entre les géométries. Ces relations sont stockées dans son système de gestion de fichiers, toujours dans un format propriétaire, mais le logiciel en fournit une vue à l'utilisateur, sous la forme de tables reprenant les principaux indicateurs de connexité et de connectivité pour chaque entité spatiale (Fig. 4). Ces tables sont

engendrées automatiquement dans le SGBD-R et peuvent bien sûr être associées, par exemple par jointure, aux tables d'attributs définissant la thématique des entités géographiques. L'utilisateur a donc la possibilité d'interroger la base de données au moyen du langage de re-

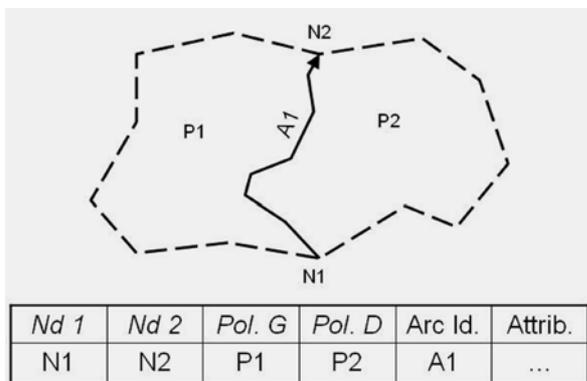


Fig. 4 : Vue engendrée par un système géo-relationnel, des relations topologiques (champs en italique) d'un arc délimitant une frontière commune entre deux polygones Nd 1 : noeud de départ - Nd 2 : noeud d'arrivée - Pol. G : polygone à gauche
Pol. D : polygone à droite - Arc Id. : identifiant de l'arc - Attrib. : attribut(s) de l'utilisateur

de l'information géographique, géométrie comprise, au sein d'un seul SGBD. On parle alors de solution ou d'architecture intégrée. L'intention, et même les solutions théoriques, existent depuis longtemps, mais l'implémentation était freinée par des limites du modèle relationnel de gestion de base de données. Le modèle "relationnel étendu", s'inspirant de l'approche orientée-objet, a permis de lever la plupart de ces obstacles, tandis que l'apparition de standards et de normes a largement contribué à la diffusion de ce nouveau modèle d'architecture (5).

Sans rentrer dans les détails, ni s'attarder sur l'historique (6), on retiendra que la principale avancée du modèle intégré consiste à utiliser, dans le schéma de la table des entités spatiales, un champ BLOB (Binary Long Object) pour y stocker les géométries. Chaque enregistrement (ligne de la table) est donc capable de conserver toute l'information relative à une entité spatiale : à la fois l'identifiant géographique et les attributs thématiques dans des champs alphanumériques traditionnels, mais aussi les coordonnées fixant la géométrie de l'entité, stockées "en vrac" dans le champ BLOB (Fig. 5). Si on avait laissé la liberté à chaque concepteur de bases de données d'organiser à sa guise le champ BLOB, on aurait permis autant de solutions propriétaires que de concepteurs. Par conséquent, l'efficacité de la méthode n'est atteinte que grâce à une standardisation. Un consortium d'organisations publiques et privées - l'Open GIS/Geospatial Consortium (7) - s'est révélé particulièrement actif dans l'élaboration de standards, généralement relayés par des normes au niveau du comité technique de l'ISO responsable de l'information géographique (8). En pratique, les types de géométries, leurs propriétés, leurs relations et leurs référen-

quête SQL, tant sur la thématique que sur les relations spatiales.

Architectures intégrées

Cette forme d'architecture hybride, qui a prévalu durant deux décennies, n'assure pourtant pas la transparence et la sécurité qu'est en droit d'attendre un administrateur de données. Les formats propriétaires des fichiers conservant la géométrie des entités spatiales sont à la fois un obstacle à leur diffusion et une contrainte sur les traitements auxquels les données pourront être soumises. En outre, un SGF ne garantit pas les fonctionnalités basiques de sécurité, archivage, etc., de façon aussi efficace qu'un SGBD. D'autre part, le SGBD associé au SGF et qui est en charge de la gestion des attributs thématiques des entités spatiales, n'est pas nécessairement compatible (modèle, version, etc.) avec le ou les autres systèmes gérant les autres bases de données de l'organisation. Ces raisons expliquent pourquoi les architectures les plus récentes se sont efforcées de gérer la totalité

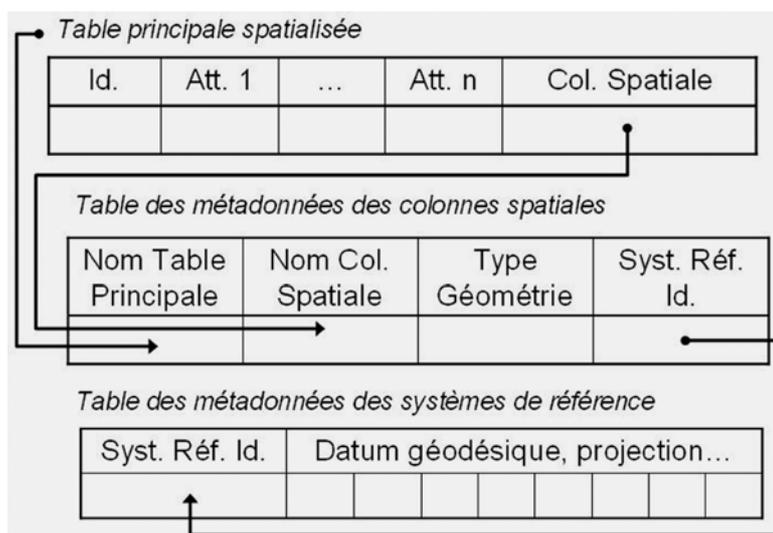


Fig. 5 : Exemple de spatialisation d'une table principale, et tables de métadonnées associées.
Id. : identifiant de l'entité géographique - Att. : attribut - Col. Spatiale : colonne spatiale (BLOB) - Syst. Réf. Id. : identifiant du système de référence géographique.

tiels spatiaux, susceptibles d'être conservés dans la base de données, sont décrits par des métadonnées, elles-mêmes stockées dans la base sous la forme de tables normalisées spécifiques. En outre, le langage transactionnel permettant d'agir sur la base de données, est enrichi de fonctionnalités propres aux informations géographiques. Cela permet d'assurer l'alimentation de la base, mais aussi son indexation (spatiale) et l'expression de requêtes spatiales. La spatialisation d'une base de données exploite ainsi une "extension spatiale" (parfois intitulée "cartouche spatiale") correspondant à :

- La définition d'une colonne spatiale dans la table principale des entités géographiques, pour y stocker les géométries ;
- La présence de tables de métadonnées spatiales décrivant la manière dont les données spatiales sont gérées ;
- L'ajout de fonctionnalités garantissant la gestion et l'exploitation des données spatiales.

De telles extensions spatiales sont aujourd'hui proposées par la plupart des SGBD, tant dans les logiciels commerciaux que dans les logiciels libres. Cependant, la migration d'une architecture hybride vers une architecture intégrée se heurte à une opération de *re-engineering* des bases de données qui n'est pas toujours simple et qui est susceptible de remettre en cause la conception du système d'information géographique de l'organisation hôte. D'autre part, quelles que soient les fonctionnalités spécifiques ajoutées au langage de requête, un SGBD spatial n'est pas en mesure d'effectuer tous les traitements attendus sur des données géographiques (fonctions avancées de cartographie et d'analyse spatiale, en particulier). Par conséquent, la présence d'une couche de logiciel supplémentaire, coiffant en quelque sorte le SGBD spatial, reste indispensable pour répondre aux besoins des organisations traitant de l'information géographique.

Systèmes d'information géographique

L'expression "système d'information géographique" (SIG) est souvent utilisée pour désigner le logiciel permettant la numérisation des données géographiques et leur exploitation par des traitements informatisés. Plus fondamentalement, le SIG n'est qu'un système d'information (SI) parmi d'autres au sein de l'organisation, au même titre que le SI relatif à la gestion du personnel de l'entreprise, ou son SI comptable. Cette notion de SI renvoie à l'analyse systémique des organisations (Fig. 6), dont le SI (ou les sous-SI dédiés) constitue(nt) l'un de ses composants essentiels, à côté du système opérant (SO) et du système de décision (SD). Ne gérant que de l'information, le SI est en théorie aisément informatisable : les données qui l'alimentent forment une base de données, et les traitements sur ces données sont confiés à un système de gestion de base de données (1).

Le SIG n'est donc qu'un SI spécialisé, dans la mesure où les informations concernées sont géographiques (géométries + attributs). Mais son informatisation est plus délicate, soit qu'elle passe par une solution hybride non assimilable aux solutions utilisées par les autres SI de l'organisation, soit qu'elle utilise l'extension spatiale d'un SGBD classique, tout en réclamant une couche supplémentaire de logiciel pour les traitements géographiques et/ou cartographiques.

Une autre différence fondamentale fait la spécificité du SIG.

Les informations alimentant un SI classique (non géographique) dans une organisation sont essentiellement engendrées par l'organisation elle-même (finances, personnel, stock, clientèle, etc.). Or, dans le cas d'une organisation exploitant un SIG (à titre d'exemples : une société de transport, un gestionnaire de réseau enterré, une société commerciale faisant du "géo-marketing", etc.), la grande majorité des informations géographiques ne sont pas créées par l'organisation, mais four-

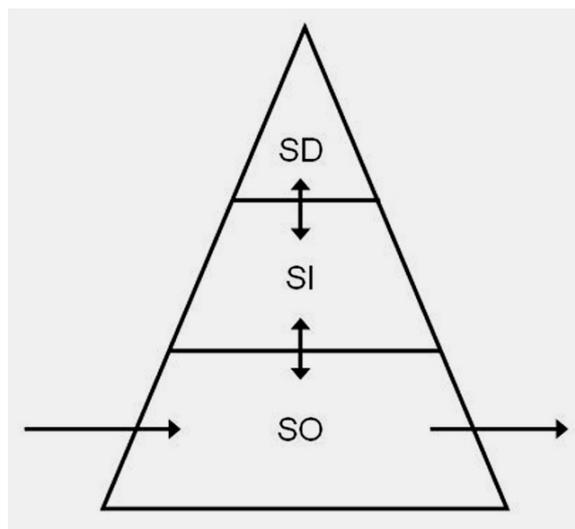


Fig. 6 : Les 3 systèmes modélisant une organisation :
 SO : système opérant, réalisant les "transformations" entre un flux entrant (ex. matières premières) et un flux sortant (ex. produits finis) – SI : système d'information : enregistrant qui fait quoi, où, quand et comment dans l'organisation – SD : système décisionnel, décidant de qui fait quoi, où, quand et comment dans l'organisation.

nies par un "producteur de données géographiques" externe. Bien entendu, chaque organisation peut être amenée à construire ses propres données spatiales, partielles et locales. Mais elle se basera toujours sur une information géographique de référence (voirie, parcellaire, adresses, etc.). Il s'agit donc d'acquérir ces informations de référence auprès d'un producteur spécialisé, capable de mettre en œuvre, selon des critères de qualité sévères, leur levé et leur mise à jour. Se pose ainsi le problème de la diffusion de l'information géographique, d'une part en termes d'adéquation entre les spécifications de producteurs et les besoins des utilisateurs et, d'autre part, en termes d'interopérabilité entre systèmes.

La carte imprimée a longtemps servi de support de référence aux utilisateurs. Avec la banalisation des appareils de numérisation, bon nombre d'entre eux ont été tentés d'enregistrer par leurs propres moyens l'information disponible sur les cartes. Il s'en est suivi une multitude de fichiers numériques de données "cartographiques" (et non pas géographiques), collectés de façon incohérente et redondante, sans validation ni mise à jour. La distribution, par les agences responsables de la cartographie, des fichiers numériques de première génération n'a pas significativement amélioré la situation. L'information disponible restait souvent sous forme cartographique, non adaptée – ni en qualité ni en quantité – aux besoins des utilisateurs, et sous une forme largement dictée par les formats propriétaires exploités par les logiciels relevant du modèle hybride. Ce n'est que récemment, à la suite de l'adoption par les agences productrices des architectures intégrées et de la standardisation qu'elles impliquent, que la problématique de la diffusion des données géographiques de référence a été correctement appréhendée.

Infrastructures de données géographiques

Les producteurs de données géographiques sont aujourd'hui en mesure de concevoir leurs bases de données autour de la notion d'objets géographiques, sous une forme non déterministe (traitements non a priori déterminés) et largement indépendante des supports physiques. De plus, selon les principes édictés par les instances internationales – telle l'U.E. dans le cadre du projet INSPIRE (9) – chaque producteur est invité à limiter l'acquisition, la tenue à jour, la gestion et la diffusion aux seules données pour lesquelles il a la responsabilité légale ou contractuelle. Il reste à informer les utilisateurs de l'offre existante en matière de données géographiques, afin que ces derniers sélectionnent, acquièrent et / ou exploi-

tent les données dont ils ont besoin. Cette "mutualisation" de l'information géographique peut profiter de la très large pénétration des réseaux informatiques auprès de tous les acteurs, tant producteurs qu'utilisateurs. Ensemble, les producteurs de données peuvent ainsi constituer, sur un territoire délimité, et éventuellement autour d'une thématique donnée, une infrastructure de données géographiques (10).

Dans un premier temps, il s'agit d'identifier les données géographiques disponibles et de décrire leurs spécifications. C'est aux producteurs de documenter les données qu'ils produisent, au moyen de métadonnées standardisées (norme 19115), qu'ils publieront sur le réseau dans un ou plusieurs dictionnaires accessibles aux moteurs de recherche (11). Dans la pratique, une recherche efficace réclame encore des utilisateurs qu'ils partagent avec les producteurs une "connaissance" du domaine documenté. Une mutualisation de ces connaissances demanderait la définition d'une ontologie géographique, toujours à faire, permettant d'envisager une recherche totalement exploratoire (12).

L'accès aux données proprement dites constitue une seconde étape. Une fois les sources identifiées et la qualité jugée adéquate, l'utilisateur peut souhaiter acquérir les données géographiques auprès du producteur. Cela peut se faire de plusieurs manières, y compris par téléchargement sur le réseau. Cette duplication des données, depuis le serveur du producteur, vers le poste client de l'utilisateur, présente cependant plusieurs contraintes :

- L'utilisateur doit disposer des capacités de stockage, gestion et traitement de l'information géographique, tant au niveau physique (matériel / logiciel), qu'au niveau des compétences et ressources humaines.
- L'utilisateur doit être en mesure d'alimenter sa propre base de données, ce qui réclame le respect de standards et de normes dans la structuration des données et les formats d'échange, tant au niveau du producteur qu'à celui de l'utilisateur.
- L'approche risque d'engendrer à court ou moyen terme, une désynchronisation de la tenue à jour des données entre le producteur et l'utilisateur, voire une incohérence entre les données si l'utilisateur exploite les données de plusieurs producteurs.

Une approche alternative consiste pour l'utilisateur à accéder aux données géographiques sur le site du producteur, et à traiter ces données "en ligne", c'est-à-dire sans les télécharger, le poste client ne visualisant que le résultat du traitement. Si le traitement demandé se ré-

sume à une visualisation cartographique d'une sélection de données, des solutions diverses sont disponibles, du simple "plug-in" téléchargeable sur le client, à l'exécution de procédures stockées via un port CGI (Common Gateway Interface) du serveur (13).

La généralisation d'une telle approche conduit à l'élaboration d'un système d'information géographique distribué, où les données et les traitements sont mutualisés par plusieurs producteurs distincts (14). Les producteurs de données géographiques participant à un tel SIG distribué, doivent toujours publier les métadonnées normalisées, mais en outre rendre leurs données accessibles à des "services Web" (modules de traitement de données exécutables en ligne) invoqués par les utilisateurs. D'un autre côté, les services offerts (généralement par des producteurs indépendants des producteurs de données) doivent également être documentés, afin d'être identifiés et sélectionnés par les utilisateurs. Au catalogue de métadonnées décrivant les données, s'ajoute donc un catalogue ou registre des services Web, l'ensemble devant être normalisé pour assurer l'interopérabilité des traitements sur les données. Plusieurs types de services Web traitant les données géographiques sont déjà standardisés ou en voie de standardisation par l'ISO (15) ; à titre d'exemples : Web Map Service – WMS – pour le traitement relativement passif de données cartographiques ; Web Feature Service – WFS – pour un traitement dynamique de données géographiques ; Web Coverage Service – WCS – pour des traitements portant sur des données en format maillé ; etc. La solution, aujourd'hui opérationnelle pour des traitements courants sur des données cartographiques (16), est amenée à se développer rapidement pour faciliter l'accès aux données géographiques de référence, à l'avantage de toutes les catégories d'utilisateurs (Fig. 7).

Conclusion

Depuis les premiers logiciels de cartographie assistée par ordinateur jusqu'aux infrastructures de données spatiales, la gestion numérique de

l'information géographique est allée dans le sens d'une structuration et d'une normalisation croissantes. Profitant de l'évolution des modèles de gestion de bases de données, en particulier des avancées apportées par le modèle objet-relationnel et les bases de données réparties,

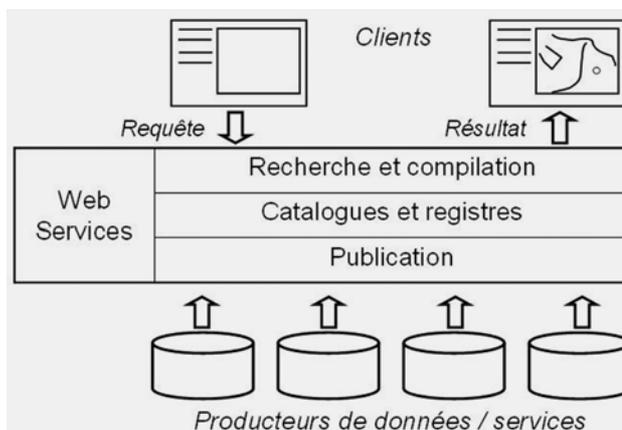


Fig. 7 : L'interface des services Web, entre producteurs et clients, dans une infrastructure de données spatiales.

l'information géographique apparaît désormais comme une ressource partageable, au bénéfice d'une communauté d'utilisateurs de plus en plus large. L'incorporation des fonctionnalités de spatialisation des bases de données dans les versions courantes des SGBD actuels en est témoin. Le fait que ces facilités apparaissent au moment où les systèmes de positionnement géographique et de navigation par satellite (GNSS, tels GPS et Galiléo) sont en passe de révolutionner le rapport de l'individu à l'espace, n'est évidemment pas une simple coïncidence.

Jean-Paul Donnay
 Université de Liège
 Allée du 6 Août, 17
 4000 Liège
 jp.donnay@ulg.ac.be

21 décembre 2005

Bibliographie

1. Pantazis, D. ; Donnay J.-P. *La conception de SIG. Méthode et formalisme*. Série Géomatique. Paris : Hermès, 1996, 343 p.
2. Priou, D. ; Le Gallic, J.-M. ; Pelle, S. ; Richard, D. *UML, Java et les données géographiques*. Collection ENSG-IGN. Paris : Hermès-Lavoisier, 2004, 2 volumes, 470 et 397 p.

3. Chaidron, C. ; Fourneau, D. ; Donnay, J.-P. *Assistance scientifique au projet "Seamless Geographic Information System – Reference (SGISR) de l'Institut Géographique National – Belgique (IGN-B)*. Rapport de recherche. Université de Liège, non publié, 2005.
4. Longley, P.A. ; Goodchild, M.F. ; Maguire, D.J. ; Rhind, D.W. *Geographic Information Systems and Science*. 2nd ed.. Chichester : Wiley, 2005, 517 p.
5. Shekhar, S. ; Chawla, S., *Spatial Databases: A Tour*. Upper Saddle River (N.J., USA) : Prentice Hall, 2003, 262 p.
6. Adam, N.A. ; Gangpadhyay, A. *Database issues in geographic information systems*. Norwell (MA, USA) : Kluwer , 1997, 136 p.
7. Open Geospatial Consortium (OGC). *OGC Reference Model*. OGC 03-040, version 0.1.3, 2003. <<http://www.opengeospatial.org/specs/?page=orm>> (consulté le 21 décembre 2005).
8. International Organization for Standardization. *ISO 19101: Geographic Information – Reference Model*. ISO/TC 211. Geneva : ISO, 2002. <<http://www.isotc211.org>> (consulté le 21 décembre 2005).
9. European Commission's Joint Research Centre. *INSPIRE: Proposal for a directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community*. Ispra : JRC, 2004. <<http://inspire.jrc.it>> (consulté le 21 décembre 2005).
10. Gittings B. (ed.). *Integrating Information Infrastructures with GI Technology*. Innovations in GIS series, vol. 6. London : Taylor & Francis, 1999, 280 p.
11. International Organization for Standardization. *ISO 19115: Geographic Information – Metadata*. ISO/TC 211. Geneva : ISO, 2003. <<http://www.isotc211.org>> (consulté le 21 décembre 2005).
12. Nogueras, J ; Zarazaga, F.J. ; Muro-Medrano, P.R. *Geographic Information Metadata for Spatial Data Infrastructures*. Berlin : Springer, 2005, 263 p.
13. Kraak, M.-J. ; Brown, A. *Web Cartography. Developments and prospects*. London : Taylor & Francis, 2001, 213 p.
14. Peng Z.R. & Tsou M.H. *Internet GIS*. Hoboken (N.J., USA) : Wiley, 2003, 679 p.
15. International Organization for Standardization. *ISO 19119: Geographic Information – Services*. ISO/TC 211. Geneva : ISO, 2005. <<http://www.isotc211.org>> (consulté le 21 décembre 2005).
16. International Organization for Standardization. *ISO 19128: Geographic Information – Web map server interface*. ISO/TC 211. Geneva : ISO, 2005. <<http://www.isotc211.org>> (consulté le 21 décembre 2005).

EEN KORT OVERZICHT VAN DATA WAREHOUSING EN OLAP

Jef WIJSEN

Professeur, Université de Mons-Hainaut (UMH) - Institut d'Informatique

▪ Veel organisaties hebben in de loop der jaren een overstelpende hoeveelheid gegevens verzameld in diverse database-systemen. De term "data warehousing" omvat de technologie om deze operationele en historische gegevens in een geïntegreerde en samengevatte vorm beschikbaar te maken voor de bedrijfsvoering van een organisatie. Het analyseren en verwerken van deze geaggregeerde informatie staat bekend onder termen als "OLAP (Online Analytical Processing)" en "data mining". Dit artikel geeft een beknopt overzicht van deze nieuwe technologieën, die de laatste jaren in een stroomversnelling zijn geraakt.

▪ Au fil des années, de nombreuses entreprises ont accumulé d'énormes quantités de données dans différents systèmes de bases de données. La notion de "data warehousing" désigne la technologie conçue en vue de rendre ces données opérationnelles et historiques disponibles pour la gestion d'une organisation dans une forme intégrée et condensée. L'analyse et l'exploitation de cette information agrégée est connue sous les appellations "OLAP (Online Analytical Processing)" et "data mining". Cet article donne un bref aperçu de ces nouvelles technologies, très en vogue ces dernières années.

Sinds de jaren zestig hebben onderzoek en ontwikkelingen in databases zich toegeleid op de automatisatie van repetitieve taken (of "transacties"), hetgeen thans aangeduid wordt met de term *online transaction processing* (OLTP). Dit onderzoek heeft onder andere geleid tot de zeer succesvolle relationele database-systemen. Veel bedrijven hebben in de loop der jaren een overstelpende hoeveelheid gegevens verzameld in dergelijke databases. Deze historische en operationele gegevens verhullen vaak een schat aan kennis over het bedrijf en de business, in de vorm van verborgen regels, trends, patronen... Een nieuwe uitdaging is om deze onbekende kennis te onthullen en bruikbaar te maken voor de bedrijfsvoering. Dit heeft geleid tot drie nieuwe technologische ontwikkelingen op het vlak van beslissingsondersteunende systemen (*decision support systems*):

	OLTP	OLAP
gebruiker	klerk.	manager.
werklust	frequente transacties:	regelmatige analyses:
toegang	lezen en schrijven, een tiental records.	vooral lezen, scan van miljoenen records.
gegevens	actueel.	actueel en historisch.
database grootte	100 MB tot GB.	100 GB tot TB.

Data warehousing

De OLTP gegevens zijn vaak verspreid over verschillende systemen, sterk gedetailleerd en/of van slechte kwaliteit. Data warehousing omvat het integreren, samenvatten en zuiveren van deze gegevens in een nieuwe opslagplaats, teneinde ze bruikbaar te maken voor analyse.

OLAP

Dit is een acroniem voor *online analytical processing*: het interactief analyseren van de gege-

vens in het data warehouse, waarbij de gegevens doorgaans voorgesteld en gemanipuleerd worden in de vorm van multidimensionale matrices of spreadheets, aangeduid met de term gegevenskubus (*data cube*).

Data mining.

Het exploreren van de gegevens op zoek naar interessante, nieuwe kennis (regels, trends, patronen...).

Er bestaat geen scherpe grens tussen OLAP en data mining. In OLAP zal de analist gewoonlijk precieze instructies geven, bijvoorbeeld over het deel van de gegevens waarop gefocust moet

Fig. 1: Verschillende karakteristieken van transactionele en beleidsondersteunende systemen.

worden. In data mining zal het systeem vaak zelf die focus bepalen. Zo is de vraag "Geef per maand het aantal abonnees dat ons verlaten heeft in het voorbije jaar", een typische OLAP query!, terwijl een data mining query eerder zou vragen "Welke factoren beïnvloeden het verlies van abonnees?".

Figuur 1 toont dat OLAP andere technologische eisen stelt dan het traditionele OLTP. Transactionele databases zijn afgestemd op een typische, gekende OLTP werklust. Het toevoegen van complexe OLAP queries aan deze werklust zou kun-

nen leiden tot een onacceptabele performantie van het gehele systeem. Dit is een bijkomende reden om een data warehouse te bouwen los van de bestaande transactionele databases.

Dit artikel is gestructureerd als volgt. Sectie 2 definieert het concept "data warehouse" en legt uit welke de verschillende stappen zijn bij de constructie van een data warehouse. Het concept "data mart" komt hier eveneens ter sprake. Sectie 3 gaat dieper in op het gegevensmodel en de queries in een OLAP omgeving. De termen ROLAP en MOLAP worden verduidelijkt. Sectie 4 duidt aan wat het verschil is tussen data mining en OLAP. Het was niet ons opzet om diep in te gaan op het onderwerp data mining. Sectie 5 geeft een visie op toekomstige ontwikkelingen op het vlak van data warehousing. Sectie 6 beschrijft een praktijkervaring. Sectie 7 eindigt met een persoonlijke keuze uit de literatuur.

Data warehouse

Wat is een data warehouse?

Referentie (2) definieert een data warehouse als een subject-georiënteerde, geïntegreerde, niet-vluchtige, historische gegevensopslagplaats ter ondersteuning van de besluitvorming op het niveau van het management van een bedrijf. We lichten deze termen kort toe.

Subject-georiënteerd en geïntegreerd

OLTP gegevens zijn vaak verspreid over verschillende databases die elk een bepaalde applicatie (facturering, levering, productie, ...) ondersteunen. Deze gegevens worden in een data warehouse geïntegreerd en georganiseerd rond bepaalde onderwerpen (klant, product, leverancier, ...).

Niet-vluchtig en historisch

De term "nietvluchtig" betekent dat de gegevens, eenmaal ingevoerd in het data warehouse, niet gewijzigd worden, hoewel verwijderen eventueel wel mogelijk is. De term "historisch" duidt erop dat de gegevens een zekere tijdspanne bestrijken, nodig voor het analyseren van trends.

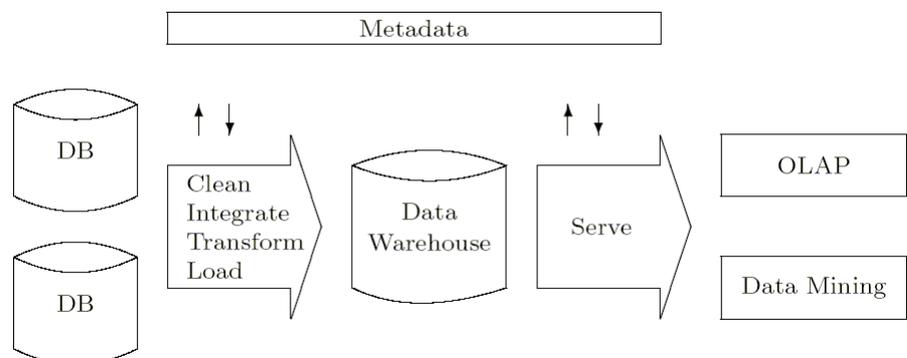


Fig. 2: Schematisch overzicht van data warehousing

Wat is een data mart?

Veel bedrijven wensen één enkel data warehouse te bouwen voor de ondersteuning van het geheel van beslissingsondersteunende activiteiten. De constructie van een data warehouse dat het ganse bedrijf omspant, is evenwel een complex proces dat een uitgebreide modellering van de business vereist en verschillende jaren in beslag kan nemen. Sommige bedrijven kiezen daarom voor *data marts*, die kunnen omschreven worden als departementale data warehouses gefocust op een specifiek onderdeel van de business. Een voorbeeld is een marketing data mart met informatie rond de onderwerpen klant, product en verkoop. Deze data marts kunnen sneller gerealiseerd worden omdat ze geen consensus op het bedrijfsniveau vereisen, maar kunnen op lange termijn tot complexe integratieproblemen leiden indien geen compleet business model voorhanden is.

Zelfs als er een geïntegreerd data warehouse bestaat, kan het om redenen van flexibiliteit en performantie interessant zijn om data marts te extraheren uit het data warehouse.

Constructie van een data warehouse

De constructie en het onderhoud van een data warehouse is een complexe bezigheid, die verschillende taken omvat. Voor elk van deze taken zijn commerciële hulpmiddelen beschikbaar. Figuur 2 geeft een schematisch overzicht van de verschillende fasen en componenten betrokken bij data warehousing.

Extractie

Extractie is het proces dat gegevens onttrekt aan de transactionele databases en andere gegevensbronnen. Het betreft typisch een (nachtelijk) batch proces, vaak verschillende subprocessen in parallel om de doorlooptijd te verkorten.

Zuiveren

Vermits een data warehouse wordt gebruikt in de besluitvorming, is het belangrijk dat de gegevens in het data warehouse correct zijn. Aangezien in een data warehouse grote hoeveelheden gegevens uit verspreide, heterogene gegevensbronnen samengevat worden, bestaat er een reëel gevaar voor fouten en anomalieën. Typische zuiveringsoperaties zijn het invullen van ontbrekende waarden, het corrigeren van typografische fouten en het uniformiseren van synoniemen. Gegevens die duidelijk foutief zijn maar niet kunnen worden gecorrigeerd, worden verwijderd.

Integratie en transformatie

De gezuiverde gegevens bevinden zich doorgaans nog niet in de vorm vereist door het data warehouse. Integratie omvat het samensmelten van verschillende gegevensbronnen. Een typisch probleem hierbij is het met elkaar in overeenstemming brengen van identificaties van entiteiten: hoe kan men uitmaken dat klant-id in de ene databases naar dezelfde entiteit verwijst als kl_nr in de andere? Een ander probleem betreft het doen overeenstemmen van gegevens uitgedrukt in verschillende eenheden: de ene database bevat verkoopcijfers per maand en de andere per week, waarbij een week kan gesplitst zijn over twee maanden. Integratie kan leiden tot nieuwe ongerijmdheden in de gegevens, hetgeen een bijkomende fase van zuivering kan vereisen.

Voorbeelden van gegevenstransformaties zijn aggregatie en normalisatie: individuele verkopen kunnen geaggregeerd worden tot dagelijkse verkoopcijfers; de waarde van een variabele kan genormaliseerd worden zodat hij tussen 0 en 1 komt te liggen.

Laden en actualiseren (*refresh*)

De gezuiverde, geïntegreerde en getransformeerde gegevens worden vervolgens in het data warehouse geladen. Op dit moment worden indexen gecreëerd nodig om zoekopdrachten te versnellen. Vermits de informatie in het data warehouse een (sterk bewerkte) kopie is van de gegevens in de OLTP databases, is het nodig om het data warehouse op geregelde tijdstippen te actualiseren. Het is doorgaans te duur om het data warehouse daarbij volledig te ledigen en vervolgens opnieuw te laden.

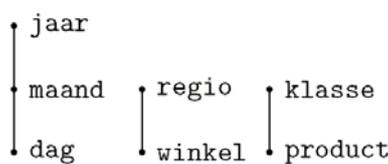


Fig. 4: Concept hiërarchieën.

Daarom worden incrementele *update* technieken gebruikt, waarbij wijzigingen in de OLTP databases op regelmatige tijdstippen worden ge-propageerd naar het data warehouse.

OLAP

Gegevenskubus

Zoals reeds aangehaald in Sectie 1 slaat OLAP (*online analytical processing*) op het interactief analyseren van de gegevens in het data warehouse. Dergelijke analyses zijn doorgaans gebaseerd op allerlei overzichtsrapporten, bijvoorbeeld de maandelijkse verkoopcijfers per regio en per product. Men kan zich zo een overzichtsrapport gemakkelijk voorstellen als een kubus: de drie dimensies van de kubus indiceren de maanden, de regio's en de producten; een cel van de kubus met coördinaat $\langle \text{jan 2001, België, Lego} \rangle$ bevat het aantal verkochte stuks Lego in België gedurende januari 2001. Ter ondersteuning van OLAP worden de gegevens in het data warehouse dus doorgaans voorgesteld in de vorm van een multidimensionale matrix of "gegevenskubus". Het zou correcter zijn om te spreken over hyperkubus in plaats van kubus, omdat het aantal dimensies verschillend van drie kan zijn. De dimensies van de kubus corresponderen met onafhankelijke variabelen en de cellen van de kubus bevatten de overeenkomstige waarden voor de afhankelijke variabele(n), ook wel aangeduid als meting(en) (*measure(s)*). Figuur 3 toont een kubus met drie dimensies dag,

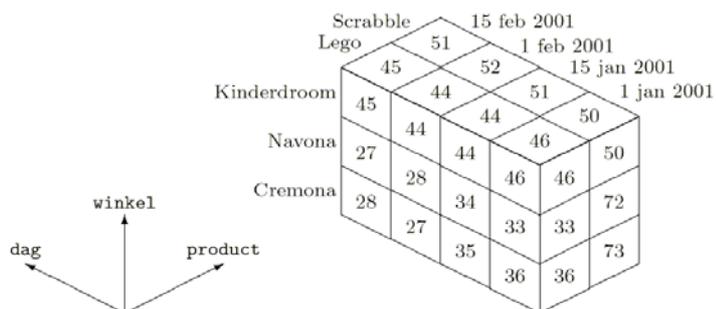


Fig. 3: Een gegevenskubus met drie dimensies.

winkel en product; de cellen van de kubus bevatten het aantal verkochte stuks. OLAP software biedt doorgaans verschillende mogelijkheden voor de grafische visualisatie van de gegevenskubus. Elke cel kan bijvoorbeeld een blokje bevatten waarvan de grootte of de kleurintensiteit evenredig is met de numerieke celwaarde. De dimensies zijn doorgaans georganiseerd in hiërarchieën, die de wijzen vastleggen waar-

op de gegevens logisch kunnen worden gegroepeerd. Figuur 4 toont dat dagelijkse cijfers aanleiding geven tot maandelijkse en jaarlijkse tota- len; winkels zijn gegroepeerd in regio's, produc- ten in klassen.

Rollup: een typische OLAP query

Een typische OLAP query geeft voor elke dimen- sie het niveau aan waarop de gegevens moeten gepresenteerd worden. Figuur 5 toont het ant- woord op de query "Geef totale verkoopcij- fers per product, regio en maand". Een der- gelijke geaggregeerde kubus, berekend uit de basiskubus, wordt ook wel *kuboïde* genoemd.

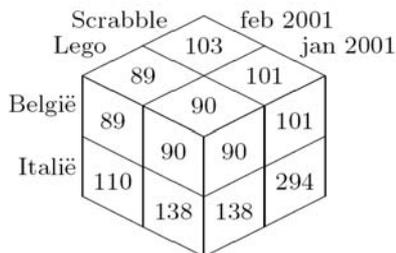


Fig. 5 : De kuboïde maand region product

Het overgaan van een gedetailleerde naar een geaggregeerde kubus wordt aangeduid met de term *rollup*.

OLAP queries kunnen ook een reduc- tie van het aantal dimensies inhouden. De tweedimensionale kuboïde maand product in Figuur 6 geeft maandelijkse verkoopcijfers per product, over alle winkels heen.

Uiteraard kan elke kuboïde desge- wenst berekend worden uit de basis- kubus. Nochtans zullen OLAP syste- men op bepaalde vragen anticiperen en kuboïdes reeds op voorhand bere- kenen en "materialiseren", teneinde antwoordtijden te verminderen. De wijze waarop dit best gebeurt, is een belangrijk onderzoeksthema.

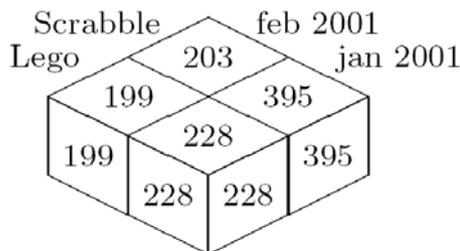


Fig. 6 : De kuboïde maand product.

Technologische keuze: ROLAP of MO- LAP

De technologische uitdaging in OLAP kan ruw- weg worden omschreven als het efficiënt onder- steunen van spreadsheet operaties op databa- ses van meerdere gigabytes. Al naargelang de gebruikte technologie kan OLAP software wor- den bestempeld als ROLAP of MOLAP: ROLAP (*relational OLAP*) maakt gebruik van bestaan- de relationele database-technologie; het uit- gangspunt van MOLAP (*multidimensional OLAP*) daarentegen is dat SQL databases wei- nig geschikt zijn om OLAP te ondersteunen.

ROLAP

Het multidimensionale gegevensmodel dat ho- ger beschreven werd, kan gemakkelijk geïm- plementeerd worden in een conventionele SQL database. De basiskubus wordt gestockeerd in een zogeheten "feitentabel", en de hiërarchieën in "dimensietabellen". Het resulterende schema wordt doorgaans een "sterschema" genoemd. Figuur 7 toont dit voor het lopende voorbeeld. Merk op dat de feitentabel reeds vooraf bere-

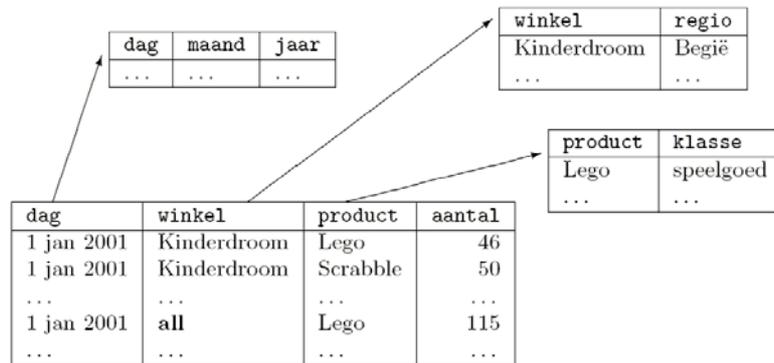


Fig. 7: Een sterschema.

kende subtotalen kan bevatten (symbool **all**). ROLAP servers breiden relationele servers door- gaans uit met gespecialiseerde middleware om de uitvoering van OLAP queries te versnellen. Er worden ook reeds voorstellen geformuleerd om de standaard relationele querytaal SQL uit te breiden met operators ter ondersteuning van OLAP.

MOLAP

Hoewel OLAP gegevens kunnen gestockeerd worden in relationele databases, geloven velen dat MOLAP een beter alternatief biedt. MOLAP impliceert een multidimensionale database waarbij gegevenskubussen worden gestockeerd in (*sparse*) matrices. Deze opslagwijze zou effi- ciënter zijn dan in ROLAP; het nadeel is dat de integratie met bestaande SQL databases moei- lijker is.

Data mining

In OLAP stuurt de eindgebruiker zelf de analyse: hij kiest de dimensies en afhankelijke variabelen en specificereert de queries. De analist vormt een hypothese en verifieert deze vervolgens met een reeks queries. Een probleem met deze werkwijze is dat de gegevens in het data warehouse vaak onvoldoende begrepen zijn (dit is precies de motivatie om analyses uit te voeren!), zodat het quasi onmogelijk wordt om interessante hypothesen te maken, de juiste gegevenskubus te kiezen en de goede queries te stellen. Het uitgangspunt van data mining is om de kracht van de computer te gebruiken om interessante patronen in de gegevens te onthullen, eerder dan hypothetische patronen te verifiëren.

We nemen een probleem van classificatie als voorbeeld. Veronderstel dat we het risico willen inschatten verbonden aan een nieuwe kredietaanvraag. Gebruik makend van historische data warehouse gegevens over achterstallige kredieten, zou een data mining programma de volgende regel kunnen afleiden: *"Indien inkomen \leq 20.000 Euro en anciënniteit \leq 5 jaar dan risico = hoog, anders risico = laag"*. Het programma ontdekt zelf dat inkomen en anciënniteit het risicobedrag bepalen, eerder dan andere kenmerken zoals leeftijd, opleiding,...

Toekomst van data warehousing

Data warehousing, OLAP en data mining zijn snel groeiende technologieën. Toekomstige ontwikkelingen kunnen in verschillende richtingen gaan. We wagen ons aan een drietal voorspellingen.

Beslissingsondersteunende systemen zullen in de toekomst meer pro-actief worden. Ze zullen niet wachten tot de analist een OLAP of data mining query stelt, maar zullen zelf aan continue zelfanalyse doen. Wanneer het systeem een interessante nieuwe evolutie ontdekt, zal het een waarschuwing zenden naar de gebruiker.

De hedendaagse systemen voor data warehousing en data mining zijn vaak ontworpen om te werken voor gelijk welke business. Toekomstige systemen zullen meer gericht zijn op een welbepaalde business, bijvoorbeeld de petroleumsector. Dergelijke business-specifieke systemen zullen meer domeinafhankelijke logica bevatten die de kwaliteit van de analyses zal verhogen.

Het data warehouse zal in toenemende mate aangevuld worden met achtergrondinformatie afkomstig uit externe gegevensbronnen, beschikbaar via bijvoorbeeld het Web. Een onderneming

heeft geen directe controle over de toegang tot deze externe gegevensbronnen, dit in tegenstelling tot de eigen databases. Standaarden voor gegevensuitwisseling, zoals XML, spelen hierbij een belangrijke rol.

Getuigenis

De eindverhandeling (3) is illustratief voor de problemen die kunnen gepaard gaan met de invoering van data warehousing en data mining in een onderneming. De onderneming in kwestie beschikt over een tiental belangrijke applicaties, die gebruik maken van database management systemen van een viertal verschillende leveranciers. Het oorspronkelijke opzet van het onderzoek was om door middel van data mining technieken een antwoord te vinden op vragen als: *"Wie zijn onze klanten?"* en *"Met welke service zijn de klanten meest gebaat?"*. Al spoedig bleek dat data mining onmogelijk zou zijn zonder een grondige "preparatie" (zie sectie 2.3) van de gegevens. Enkele netelige kwaliteitsproblemen waren de volgende:

- Dubbele registratie van eenzelfde entiteit. Bijvoorbeeld, <RAYTEC, Rue du Commerce 2,...> en <S.A. RAYTEC, 2 Rue du Commerce,...>.
- Veelvuldig gebruik van de "fourre-tout" code "andere" voor attributen zoals opleiding of beroep. Een ander probleem is dat de codes in verschillende toepassingen niet op elkaar afgestemd zijn.
- Ontbrekende, onmogelijke of achterhaalde attribuutwaarden.

Verdere informatie

Mijn persoonlijke ervaring i.v.m. de literatuur over OLAP en data warehousing is dat sommige documenten op het Web interessanter en actueler zijn dan heel wat boeken. De Web site <<http://www.daniel-lemire.com/OLAP/>> bevat een interessant overzicht van wetenschappelijk onderzoek in data warehouses en OLAP; de site bevat ook links naar *white papers* geschreven door commerciële software leveranciers. Referentie (1) geeft een degelijk overzicht van data warehousing en data mining. Referentie (4) is een goed boek over data mining.

In tegenstelling tot de database-industrie, is de OLAP-markt vandaag de dag sterk gefragmenteerd, zonder dominante spelers. De Web site <<http://www.olapreport.com>> bevat interes-

sante informatie over de OLAP-markt. De Website van KDnuggets <<http://www.kdnuggets.com>> geeft een overzicht van data mining software.

Jef WIJSEN
Université de Mons-Hainaut
Bâtiment «Le Pentagone»
Avenue du Champ de Mars, 6
7000 Mons
Jef.Wijisen@umh.ac.be

17 januari 2006

Bibliografie

1. Han, J. ; Kamber, M. *Data Mining: Concepts and Techniques, 2nd ed.* San Francisco : Elsevier/Morgan Kaufmann, 2005, 800 p.
2. Inmon, W. *Building the Data Warehouse 2nd ed*, New-York : Wiley, 1996.
3. Van Puyvelde, H. *De l'information opérationnelle à l'intelligence décisionnelle par le data mining - Étude de faisabilité appliquée au cas d'un service public.* Master's thesis, Université de Mons-Hainaut, 2000.
4. Witten, I. ; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed.*, San Francisco : Elsevier/Morgan Kaufmann, 2005, 560 p.
Deze studie bevestigt wat verschillende auteurs vermelden: dat vaak 80% of meer van de inspanningen in een data mining project besteed wordt aan het prepareren van de gegevens.

Nota's

1. Een query is een vraag gesteld aan een databasesysteem in een bepaalde gegevensopvraagtaal.

EUROVOC, THÉSAURUS MULTILINGUE : MAINTENANCE ET ASPECTS PRATIQUES

Exemple à la bibliothèque du Parlement européen

Isabelle GAUTIER

Spécialiste de l'information

Représentante du Parlement européen au Comité de maintenance Eurovoc

Chargée du groupe de travail "Coordination de l'indexation" à la bibliothèque du Parlement européen

Article rédigé suite à la conférence donnée par l'auteur dans le cadre de l'*Inforum 2005, Systèmes et langages documentaires : fictions et réalités*, organisé par l'Association Belge de Documentation, le 28 avril 2005 à Bruxelles.

Het artikel is opgesteld naar aanleiding van een conferentie gegeven door de auteur in het kader van het *Inforum 2005, Informatiesystemen: fictie en werkelijkheid*, georganiseerd door de Belgische Vereniging voor Documentatie op 28 april 2005 te Brussel.

- Eurovoc est un thésaurus multilingue spécialement construit à l'origine pour le traitement de l'information documentaire des institutions de l'Union européenne. Cet article présente tout d'abord le thésaurus Eurovoc, sa gestion et sa maintenance puis explique son utilisation quotidienne à la bibliothèque du Parlement européen et dans les autres domaines d'activités du Parlement européen.
- Eurovoc is een meertalige thesaurus aanvankelijk speciaal opgesteld voor het behandelen van documentaire informatie van de Europese unie. Het artikel bespreekt eerst en vooral de Eurovoc thesaurus, zijn beheer en zijn onderhoud en nadien zijn dagelijks gebruik in de bibliotheek van het Europese Parlement en in andere activiteitendoelens van het Europese Parlement.

Présentation d'Eurovoc

Genèse et développement

Conçu en 1984, le thésaurus Eurovoc est le fruit d'une initiative conjointe du Parlement européen et de l'OPOCE (Office des publications officielles des Communautés européennes). Pour le Parlement européen, il s'agissait d'indexer les documents parlementaires et ceux de la bibliothèque, et pour l'OPOCE, d'indexer les textes publiés au Journal Officiel. Eurovoc répond au besoin constaté d'un outil multidisciplinaire qui, tout en étant généraliste, puisse en même temps couvrir tous les domaines traités par l'Union européenne. L'accent est également mis sur les activités parlementaires, ce qui permet son utilisation par des Parlements nationaux.

Cet outil documentaire est multilingue depuis sa création : il comprend au départ sept langues : l'allemand, l'anglais, le danois, le français, l'italien, le grec et le néerlandais. La langue utilisée pour sa conception est le français, les autres langues sont des traductions.

Les transformations de l'Union européenne impliquent de nouvelles éditions et de nouvelles versions linguistiques ; en 1987 paraît la

deuxième édition qui comprend deux langues supplémentaires : l'espagnol et le portugais. La troisième édition, la version 3.0, date de 1995. Les éditions paraissent alors en version papier.

À partir de 1999, la gestion du thésaurus est modifiée. Deux comités, l'un de pilotage, l'autre de maintenance sont mis en place. Une décision interinstitutionnelle décide une rapide mise à jour du thésaurus, ce qui mène à la version 3.1. Elle comprend deux langues supplémentaires : le suédois et le finlandais. Avec cette nouvelle version, Eurovoc dispose, pour la première fois, de son propre site Internet et est désormais accessible en ligne.

Entretemps, est décidée une révision totale du thésaurus qui aboutit à la version 4.0 parue en novembre 2002. L'édition suivante, la version 4.1, est mise en ligne début 2004.

De nombreux parlements nationaux se sont intéressés à Eurovoc via le réseau CERDP (Centre européen de recherche et de documentation parlementaires). Pour se préparer à leur intégration dans l'Union européenne, plusieurs parlements ont traduit le thésaurus, ainsi, la Lituanie, la Slovaquie, la Pologne, la Roumanie et la République tchèque. D'autres parlements ont décidé de traduire le thésaurus notamment dans le ca-

dre d'une coopération renforcée avec l'Union. Il s'agit de l'Albanie, de la Croatie et de la Russie.

Depuis 1995, le thésaurus n'avait plus été édité en version papier. De nombreux utilisateurs nous ont fait part de son utilité si bien que la version 4.1 a été publiée en deux volumes : un volume présentant la version alphabétique permutée et un volume présentant la version thématique. Cette édition est disponible en onze langues (anglais, allemand, danois, espagnol, finlandais, français, grec, italien, néerlandais, portugais, suédois). De nouvelles versions linguistiques dans les langues officielles de l'Union européenne seront publiées en même temps que de prochaines versions du thésaurus.

bien de modifications de descripteurs existants)

- au téléchargement du thésaurus au format pdf (version thématique ou bien alphabétique permutée)
- la transition entre deux versions du thésaurus (les modifications et les nouveautés intégrées)
- la liste multilingue (possibilité à partir d'une langue d'avoir les équivalences linguistiques dans trois autres langues).

Enfin, la colonne de gauche de la page d'accueil propose des liens pratiques (aide, contacts...).

La recherche par menu présente les 21 domai-

Address http://europa.eu.int/celex/eurovoc/sq/sqa_doc/eurovoc_dif/SERVEUR/menuprodi/MENU?langue=FR Go Links

Europa **eurovoc** THESAURUS français (fr)

information
écrivez-nous
helpdesk
aide

Publications Office

- » [Présentation du thésaurus](#)
- » [Naviguer dans le thésaurus](#)
- » [Chercher un terme dans le thésaurus](#)
- » [Fiche de maintenance](#)
- » [Télécharger le thésaurus](#)
- » [Transition entre la version 4.1 et la version 4.2](#)
- » [Liste multilingue](#)

Eurovoc est un thésaurus multilingue couvrant tous les domaines de l'activité de l'Union européenne, il permet d'indexer les documents dans les systèmes documentaires des institutions européennes et de leurs utilisateurs. Ce produit documentaire est actuellement utilisé par le Parlement européen, l'Office des publications des Communautés européennes, les parlements nationaux et régionaux en Europe, des administrations nationales et par certaines organisations européennes. Ce site vous présente la version 4.2 du thésaurus Eurovoc. Cette édition est le résultat de la collaboration entre les institutions européennes, les parlements nationaux et les utilisateurs d'Eurovoc.

Eurovoc 4.2 est actuellement disponible dans 17 langues officielles de l'Union européenne (espagnol, tchèque, danois, allemand, grec, anglais, français, italien, letton, lituanien, hongrois, néerlandais, polonais, portugais, slovène, finnois et suédois) et deux autres langues (croate et roumain). En plus de ces versions, Eurovoc a été traduit par les parlements nationaux d'autres pays: Albanie, Bulgarie et Russie.

Version actuelle 4.2

Depuis août 2005, la version 4.2 a été mise en ligne et les langues des nouveaux pays de l'Union européenne sont intégrées au fur et à mesure de l'achèvement de leur traduction.

Le thésaurus en ligne comporte plusieurs parties. Sur la page d'accueil, les principales rubriques permettent d'accéder à :

- la présentation du thésaurus
- la navigation (par domaine)
- la recherche d'un terme (terme entier ou avec troncature)
- la fiche de maintenance (pour envoyer des propositions de nouveaux descripteurs ou

nes couverts par Eurovoc. Il s'agit des principales activités communautaires et politiques de l'UE : institutions, vie politique, relations internationales, droit, questions sociales, emploi et travail, environnement, agriculture, transport....

La structure d'Eurovoc comporte une classification hiérarchique et peut comprendre jusqu'à quatre niveaux de termes spécifiques. Les domaines sont identifiés par deux chiffres et un énoncé (exemple : 10 Communautés européennes). Ces domaines comprennent des microthésaurus qui sont identifiés par quatre chiffres dont les deux premiers relèvent du domaine (exemple : 1011 Droit communautaire). Il y a en tout 127 microthésaurus.



Pour faciliter et simplifier la gestion du thésaurus, la polyhiérarchie n'est pas acceptée sauf pour les organisations internationales ou bien pour les pays. Les autres descripteurs relèvent d'un seul domaine, même si parfois le descripteur pourrait être rattaché à plusieurs.

La notice d'un descripteur comprend son environnement hiérarchique (le MT = le microthésaurus auquel il appartient), le UF (Use For = employé pour = le synonyme), le BT (Broad Term = terme générique), ses NT (Narrow Term = termes spécifiques) et ses RT éventuels (Related Term = termes associés). Si cela s'avère nécessaire, une SC (Scope Note) est ajoutée : elle est soit une note de définition (qui précise le sens du descripteur, à ne pas confondre avec un autre), soit une note d'application (qui précise le contexte d'utilisation du descripteur). Il est possible d'avoir directement la traduction du terme dans toutes les autres langues. De même, en fonction des besoins, un simple clic permet de facilement changer de langue au cours de l'interrogation en ligne.

Gestion du thésaurus Eurovoc

Comme indiqué dans l'historique, Eurovoc est géré depuis 1999 par deux comités : le comité de pilotage et le comité de maintenance. Ces deux

comités sont interinstitutionnels et composés de membres des institutions communautaires suivantes :

- Parlement européen
- Office des publications officielles des Communautés européennes
- Conseil de l'Union européenne
- Commission européenne
- Cour de justice des Communautés européennes
- Cour des comptes européenne



Les institutions les plus engagées dans le travail de mise à jour du thésaurus sont principalement l'OPOCE, le Parlement européen et la Commission européenne.

Rôle politique et stratégique du comité de pilotage

Le comité de pilotage définit les lignes directrices du thésaurus, notamment les objectifs et les priorités à atteindre. Il a, par exemple, décidé de revoir entièrement le thésaurus pour sa version 4.0. Pour ce travail, le Parlement européen a d'ailleurs détaché deux documentalistes occupées quasiment à temps plein durant dix-huit mois.

Le comité étudie les perspectives de développement du thésaurus et fixe également le calendrier des futures éditions. Il a ainsi décidé de publier une nouvelle édition du thésaurus tous les ans. C'est également le comité de pilotage qui avale une nouvelle version du thésaurus avant qu'elle ne soit mise en ligne et c'est encore lui qui prend la décision de produire une édition papier en sus de l'édition en ligne.

En outre, il attribue également le budget annuel prévisionnel, incluant les postes de travail, la mise à jour informatique et les licences.

La disponibilité de nouvelles versions linguistiques est suivie attentivement. Des contacts réguliers avec l'équipe de maintenance de l'OPOCE sont établis pour connaître le travail effectué, les besoins à prendre en compte, le nombre de licences contractées.

Une réflexion est en cours sur le thésaurus et son développement futur. En novembre 2005, un questionnaire a été envoyé à tous les utilisateurs recensés afin de mieux connaître leurs besoins et leurs attentes. Les conclusions de cette enquête seront présentées à l'occasion de la prochaine conférence Eurovoc qui se tiendra en mars 2006 au Parlement européen.

Par ailleurs, plusieurs organismes travaillent de concert pour développer un vocabulaire contrôlé dont Eurovoc serait l'élément de référence. Le comité de pilotage réfléchit à l'organisation d'un travail d'échange et de collaboration avec d'autres institutions communautaires dans ce domaine.

Rôle technique du comité de maintenance

Le comité de maintenance met en oeuvre les politiques définies par le comité de pilotage. Son rôle est principalement technique : il étudie de nouvelles propositions pour les mises à jour du thésaurus et coordonne les travaux entre les représentants des différentes institutions. Il peut aussi avoir un rôle de conseil et d'assistance.

Ainsi, il donne des conseils techniques au comité de pilotage lorsque cela s'avère nécessaire.

Au sein de ce comité, l'OPOCE est notamment chargé de la maintenance technique du thésaurus et doit gérer les bases du thésaurus, la version officielle en ligne et la version de maintenance permettant de travailler directement sur le thésaurus pour intégrer les modifications et les changements linguistiques. Il procède à la mise à jour des différentes versions informatiques et linguistiques. Actuellement, Eurovoc est disponible au format XML.

L'OPOCE veille également à la qualité des différentes versions linguistiques existantes. Parmi les nouveaux membres de l'Union, plusieurs pays ont traduit eux-mêmes le thésaurus.

Face à l'augmentation du nombre d'utilisateurs du thésaurus, il a été jugé opportun de mettre en place une licence qui établit un contrat d'utilisation d'Eurovoc. Plusieurs types de licences sont proposés aux organismes :

- gratuite pour les besoins propres de l'organisme ou payante, s'il y a une rediffusion commerciale ;
- gratuite avec adaptation linguistique par l'organisme qui, en contrepartie, rétrocède cette version à l'OPOCE ;
- gratuite et permettant l'enrichissement ou la transformation du thésaurus (adaptation régionale, nationale ou bien vocabulaire spécifique...); les différences doivent apparaître clairement par rapport à la version officielle et ces enrichissements sont fournis à l'OPOCE qui décide de les intégrer ou non dans la version officielle.

Pour la mise à jour de l'édition officielle, le comité de maintenance fonctionne de la façon suivante : l'OPOCE reprend toutes les propositions reçues durant une période donnée et provenant soit directement des utilisateurs, soit des diverses institutions via les membres du comité de maintenance. Ces propositions sont rassemblées dans un document présentant les créations, les modifications et les suppressions de descripteurs et sont transmises aux membres du Comité de maintenance. Elles sont étudiées par chacun des membres du comité et sont votées. Les membres du comité de maintenance se réunissent régulièrement durant l'année. Au cours de ces réunions, les désaccords spécifiques sur certaines propositions sont débattus et une décision est prise.

Cas pratique : le vote sur la proposition "biométrie" n'ayant pas été unanime, la proposition a fait l'objet de discussions au sein du Comité de maintenance. Le désaccord concernait à la fois la définition de la biométrie et son terme générique (où le placer : sous le BT "données personnelles" en MT 3236 informatique et traitement des données ou bien sous le BT "biologie" ou encore sous le BT "mathématiques" en MT 3606 sciences naturelles et appliquées ?). Les membres du Comité ont consulté la place et la définition de "biométrie" dans diverses sources officielles : dictionnaires électroniques, bases de données terminologiques, autres thésaurus... Finalement, la proposition de ce nouveau descripteur a été intégrée dans la version 4.1 mais avec un BT et une définition différents de la proposition initiale :

- MT 3606 sciences naturelles et appliquées
- BT : mathématique
- RT : biologie ; données personnelles
- SN : analyse mathématique des caractéristiques biologiques d'une personne, destinée à déterminer son identité de manière irréfutable.

Si les membres n'arrivent pas à se mettre d'accord sur une proposition, l'OPOCE est chargé de trancher. Il arrive également qu'une proposition soit mise de côté pour l'édition suivante (concept trop innovant, manque de définition officielle...). Il est toujours possible de procéder à des modifications ultérieures, le thésaurus étant un outil en constante évolution.

Le comité de maintenance suit attentivement l'actualité pour intégrer de nouveaux concepts ; il veille également à garder un niveau de spécificité suffisamment pointu mais sans l'être trop, afin que le thésaurus puisse répondre aux besoins du plus grand nombre d'utilisateurs possible.

Eurovoc à la bibliothèque du Parlement européen : son application

Présentation de la bibliothèque

La bibliothèque du Parlement européen relève de la Direction de la Documentation parlementaire qui fait partie de la Direction Générale Information. La bibliothèque est située à Bruxelles et comporte également des bureaux à Luxembourg et à Strasbourg lors des sessions parlementaires.

La bibliothèque est un service interne qui a pour utilisateurs les députés et leurs assistants, les groupes politiques, les commissions parlementaires, les unités thématiques, les fonctionnaires et autres membres du personnel du Parlement européen. Les chercheurs externes, journalistes et étudiants sont réorientés vers les bureaux d'information existant dans chaque pays.

Près de 90 personnes travaillent à la bibliothèque et se répartissent en différentes équipes (direction, secrétariat, spécialistes de l'information et assistants, informatique, etc.). L'équipe des spécialistes de l'information comporte une trentaine de personnes qui sont à la fois généralistes et spécialistes : généralistes car répartis par pays et par langues (quasiment tous les pays de l'Union européenne sont représentés) et spécialistes car organisés par secteurs thématiques relatifs aux commissions parlementaires. Ils ont ainsi développé sur le site intranet de la bibliothèque des pages spécifiques d'information par secteurs. Divers produits documentaires sont proposés tels que des dossiers électroniques ou bien une veille documentaire sur des sujets spécifiques du travail parlementaire. La bibliothèque a développé toute une politique afin de mieux se faire connaître : notamment, en

organisant des sessions d'information (Info-Sessions) présentant ses services ou en donnant des formations à l'interrogation du catalogue bibliographique ou à des bases de données mises à la disposition des utilisateurs sur son site.

Eurovoc, vocabulaire commun

Il est utile de souligner l'importance d'Eurovoc pour la bibliothèque du Parlement européen. Ce thésaurus est le vocabulaire commun à tous les indexeurs qui proviennent de toute l'Europe, et qui ont des langues, des formations et des cultures fort diverses, des méthodes de travail différentes...d'où une plus grande difficulté pour définir une politique commune d'indexation. Grâce à Eurovoc, ces problèmes sont atténués et le thésaurus permet une certaine homogénéisation et harmonisation de l'indexation. Eurovoc s'avère être le seul thésaurus pertinent pour décrire notre fonds actuel.

Le fonds de la bibliothèque comporte environ 100 000 ouvrages dans les diverses langues de l'Union européenne. Parmi les langues officielles de l'Union européenne, un accent est plus particulièrement mis sur les ouvrages en français, en anglais et en allemand. Les types de documents mis à la disposition des utilisateurs consistent en des référentiels (annuaires, répertoires, dictionnaires, encyclopédies...), des livres, une sélection des publications des organisations internationales et de celles de l'Union européenne - dont les études du Parlement européen -, plus ou moins 1000 périodiques ainsi que les principaux journaux de chaque État membre. La bibliothèque comprend également un fonds spécifique sur l'histoire de l'Europe ainsi qu'un fonds consacré à la formation professionnelle.

Parmi les documents catalogués, une sélection est effectuée en fonction du thème traité et de la langue. Le document doit présenter un intérêt national, européen ou international.

Le programme de travail des institutions européennes (propositions de la Commission, procédures législatives...) est suivi de près et les acquisitions d'ouvrages ainsi que la sélection d'articles sont effectuées en fonction de l'actualité et des thèmes de travail qui seront abordés au cours de l'année parlementaire.

Selon les statistiques établies sur les notices catalographiques, les domaines d'Eurovoc les plus fréquemment exploités pour décrire notre fonds sont les suivants : la vie politique (04), les relations internationales (08), les Communautés européennes (10), la géographie (72), le droit (12), la vie économique (16), la vie sociale (2826), la documentation (3221), les sciences

humaines (3611), l'emploi (4406) et la politique de l'environnement (5206). Ce qui s'avère tout à fait en adéquation avec le travail de l'institution.

En 1999, le gestionnaire informatique de la base de données a changé. Ce qui a impliqué des méthodes informatiques et une organisation du travail différentes. L'ancienne base a été remplacée et l'incorporation des articles dans le catalogue également décidée. Jusqu'alors, existait un système de dossiers papiers comprenant une classification numérique ; les documentalistes et les bibliothécaires travaillaient séparément, ce qui n'est plus le cas. Actuellement, les spécialistes de l'information répondent indifféremment aux questions et participent au catalogue. Face à ces changements, une nouvelle politique d'indexation paraissait nécessaire et un groupe de travail chargé de la "Coordination de l'indexation" a alors été créé afin de réfléchir à cette nouvelle politique d'indexation et à sa mise en oeuvre.

Groupe de travail "Coordination de l'indexation"

Ce groupe de travail est composé d'indexeurs spécialistes de l'information, de nationalités et langues diverses. Il est chargé de :

- rédiger un guide interne pour l'application pratique des règles d'indexation au sein du service ;
- organiser régulièrement des réunions sur la politique d'indexation et sa mise en oeuvre ;
- réunir les propositions de nouveaux descripteurs afin de les transmettre au comité de maintenance Eurovoc ;
- former les nouveaux collègues.

Ce groupe existe depuis 1999. Il comprend actuellement 6 membres. Sa composition a fluctué suite aux changements du service. Une seule personne fait encore partie du groupe d'origine et assure ainsi la continuité au sein du groupe de travail.

Guide d'indexation

Ce guide a été rédigé pour permettre une meilleure cohérence de l'indexation dans le catalogue et sa bonne maîtrise technique dans le nouveau système informatique. Les techniques professionnelles ainsi mises par écrit permettent l'amélioration et l'harmonisation des pratiques d'indexation au sein de l'équipe.

Ce guide est un outil précieux tant pour assister les collègues actuels dans des tâches professionnelles parfois nouvelles que pour former les nouveaux collègues qui intègrent le service.

Il comporte trois parties : la définition et les règles de base de l'indexation ; la politique d'indexation du service ; l'application dans notre catalogue. Il est complété par des fiches-conseils d'indexation rédigées en fonction des besoins exprimés. Il a été rédigé en français et traduit en anglais.

La première partie reprend une information générale sur l'indexation et s'inspire d'un document déjà existant sur l'indexation des documents de la base EPOQUE du Parlement (base fermée actuellement). Sont rappelées les définitions de base de l'indexation avec des exemples, des conseils pour obtenir une bonne performance lors de la recherche, à travers différents concepts commentés : la pertinence, la profondeur, la spécificité, l'exhaustivité, le rappel et la précision, la pondération, la recherche inclusive (autopostage), les échanges d'informations.

La seconde partie présente la politique d'indexation adoptée par le service : savoir repérer et sélectionner les concepts, réfléchir à l'exhaustivité souhaitée, traduire des concepts en descripteurs Eurovoc ce qui implique la consultation obligatoire du thésaurus, la prise en considération du contexte, l'assistance via les termes génériques, spécifiques et reliés, les relations hiérarchiques, la recherche de la cohérence, la consultation de l'indexation antérieure, la solution des problèmes de traduction, et la consultation des sources.

La troisième partie présente l'application pratique et technique de la politique d'indexation adaptée au catalogue.

Particularités développées dans la pratique de l'indexation

Langues utilisées

Les langues de travail dans le catalogue sont pour l'instant le français et l'anglais. Les indexeurs se trouvent confrontés aux limites du logiciel documentaire. L'indexation se fait soit en français soit en anglais (durant le catalogage, les descripteurs à sélectionner défilent dans les deux langues). La langue française a été retenue pour l'intégration des descripteurs dans le catalogue mais avec renvoi aux termes anglais comme équivalent linguistique. Enfin, la recherche se fait en français.

Descripteurs principaux et secondaires

Comme dans l'ancienne base du catalogue, le choix a été fait de garder la délimitation entre descripteurs principaux et secondaires. Cette pondération permet au lecteur de la notice de déterminer rapidement si l'information est essentielle ou bien simplement complémentaire.

Par exemple : un article sur les femmes au travail comprenant un encadré sur le salaire féminin sera indexé avec "travail féminin" en descripteur principal et "salaire" en descripteur secondaire. Cette différence est immédiatement visible puisqu'elle apparaît dans deux champs différents.

Descripteur-candidat

La création d'un champ spécifique "descripteur-candidat" permet d'intégrer dans le catalogue des descripteurs spécifiques qui ne font pas encore partie de la version officielle Eurovoc ou qui ne le seront peut-être jamais s'ils sont refusés par le comité de maintenance. Dans ce cas, ils sont conservés dans une liste d'autorité interne pour les besoins propres du service et s'il s'avère au bout d'un certain temps que ces descripteurs-candidats ne sont pas utilisés, ils sont supprimés. Exemple : "contrôle juridictionnel" a été repris dans l'édition 4.2 mais "coopération énergétique" a été refusé et est resté dans la liste de la bibliothèque.

Même si certains descripteurs-candidats sont retenus dans la liste d'autorité interne du catalogue, il est impératif de continuer à indexer avec des descripteurs existant dans la version actuelle d'Eurovoc. L'indexeur devra donc saisir des combinaisons de descripteurs, en plus du descripteur candidat. Les principales raisons tiennent au fait que :

- tout utilisateur d'Eurovoc, qui n'a pas accès à la liste des descripteurs-candidats et n'en a donc pas connaissance, continue à faire des recherches avec les descripteurs existants ; c'est le cas de ceux qui consultent notre catalogue ;
- il faut prévoir la réindexation future nécessaire si le descripteur-candidat n'est pas retenu.

Indexation liée

L'indexation liée est l'indexation commune d'un ouvrage avec des chapitres spécifiques catalogués. Il peut arriver qu'un ouvrage comporte un chapitre spécifique qu'il paraît intéressant de traiter en tant que notice séparée. Dans ce cas, la règle instituée est que les descripteurs principaux de ce chapitre soient traités comme descripteurs secondaires de l'ouvrage.

Ordre des descripteurs

L'ordre d'apparition des descripteurs est également significatif : pour un ouvrage sur la sécurité des transports au Royaume-Uni et sur l'infrastructure des transports en Allemagne, on indexera pour chacun le descripteur sujet puis le descripteur pays, l'un derrière l'autre, ce qui permet de refléter au mieux le contenu de l'ouvrage.

Alors que si nous indexons par "sécurité des transports", "infrastructure des transports" puis ensuite "Royaume-Uni", "Allemagne", on pourrait croire à tort que ces sujets concernent les deux pays.

Réindexation

Les indexeurs doivent garder une copie des notices bibliographiques concernées par les descripteurs-candidats proposés, et les réindexer manuellement lorsque ceux-ci ont été intégrés dans la liste d'autorité interne de la bibliothèque.

A chaque nouvelle version du thésaurus, le catalogue est réindexé automatiquement par une routine informatique. Cependant, il reste toujours quelques corrections à effectuer manuellement.

Réunions d'indexation

Le groupe "Coordination de l'indexation" se réunit régulièrement pour analyser les propositions de nouveaux descripteurs ou de modifications envoyées par les indexeurs du service. Elles sont rédigées par eux en français ou en anglais, avec l'équivalence linguistique de la langue d'origine s'il s'agit d'une autre langue. Il n'y a pas de règles particulières pour proposer une création ou bien une modification de descripteur. Il est seulement demandé d'illustrer la demande en faisant appel à une ou plusieurs sources.

Voici quelques exemples de propositions et les résultats après analyse du groupe :

- "droit communautaire-droit international" / "Community law-international law" : descripteur-candidat créé ; ce descripteur qui n'a pas été accepté dans la version officielle d'Eurovoc est très utile pour notre service. Il fait donc partie de la liste interne de descripteurs-candidats ;
- "principe du dommage environnemental" / "environmental damage principle" : descripteur-candidat rejeté ; lorsqu'une proposition est rejetée, une explication est toujours donnée : soit l'indexeur n'a pas vu qu'il existait déjà un descripteur correspondant à son besoin, soit une combinaison de descripteurs s'avère suffisante pour rendre le concept. Pour cet exemple, le conseil d'indexation donné est de combiner "impact sur l'environnement" avec "prévention des risques" ;
- "principe pollueur payeur" / "polluter pays principle" : modification d'un descripteur ; ici, la modification porte sur la version anglaise qui n'était pas correcte (remplace "polluter pays policy").

Il s'agit également de répondre aux questions spécifiques posées.

Exemple : quelle est la différence d'utilisation entre "étude comparative" et "droit comparé" ?

- utiliser "étude comparative" quand le document traite de la situation dans divers pays. Ex : politique de la concurrence, étude comparative, France, Espagne.
- utiliser "droit comparé" quand le document traite de la législation et/ou de la jurisprudence dans divers pays. Ex : droit de la concurrence, droit comparé, France, Espagne.

Dans le cas où ces deux aspects sont décrits dans le document, il est évidemment nécessaire d'utiliser les deux descripteurs.

Selon les demandes exprimées par nos collègues indexeurs, le groupe rédige parfois des fiches-conseils. Ainsi, une fiche-conseil a-t-elle été créée sur l'utilisation des différents descripteurs géographiques et une autre sur la distinction à établir entre les différents descripteurs relatifs à l'Union européenne.

Un certain nombre d'outils d'aide à l'indexation ont été créés :

- des modèles de fiches (version EN/FR) pour proposer ou modifier des termes dans Eurovoc ;
- une liste interne des programmes communautaires qui ne sont pas inclus dans le thésaurus ;
- la liste des descripteurs créés en tant que descripteurs-candidats (disponibles dans le catalogue) ;
- la liste des descripteurs déjà traités (version EN/FR) qui permet de vérifier rapidement si la proposition a déjà été faite et la solution proposée. Pour un certain nombre de propositions de créations de descripteurs, l'option choisie est de combiner deux ou trois descripteurs plutôt que d'en créer de nouveaux pour traduire le concept. Cette liste comporte actuellement plus de vingt pages. Tous ces documents, guide compris, sont accessibles sur le site intranet de la bibliothèque.

Une fois ce travail établi au niveau du groupe, des réunions au niveau du service sont organisées ponctuellement pour présenter les nouveautés, informer de l'évolution du thésaurus et conseiller si besoin.

Formations

Le groupe "Coordination de l'indexation" est chargé de la partie formation des nouveaux col-

lègues de la bibliothèque. Différents types de formation sont proposés :

- interne, avec une présentation du thésaurus, du guide, de la politique d'indexation du service, de l'indexation dans notre catalogue et illustrée de petits exercices pratiques ;
- interne mais avec un intervenant extérieur pour offrir si nécessaire une formation à l'indexation à un groupe de collègues ;
- externe quand elle est organisée en fonction des besoins exprimés par les indexeurs et des cours disponibles dans les différents pays.

Cette méthode de travail a été expérimentée au cours du changement de système informatique en 1999 et s'est révélée extrêmement positive. Revoir les procédures d'indexation et rappeler régulièrement des règles de base s'avère en effet très utile. Ce type de cours permet à chacun de faire le point sur ses propres méthodes de travail et d'améliorer sa pratique de l'indexation.

Eurovoc dans le cadre des activités du Parlement européen

Eurovoc aux archives du Parlement européen

Eurovoc a été pendant des années l'outil d'indexation de la base de données législative du Parlement européen intitulée EPOQUE. Cette base a été fermée en 1999 et ses données sont intégrées maintenant dans la base d'archives ARCDoc. Tous les documents antérieurs à 1999 sont donc indexés avec Eurovoc. Le service des archives a le projet de reprendre les dossiers législatifs des législatures suivantes et leur attribuer une indexation Eurovoc.

Rôle du Parlement européen au sein des comités Eurovoc

Le Parlement européen joue un rôle important en tant que membre des comités Eurovoc. Au comité de pilotage, c'est lui qui assure la présidence. Il représente non seulement l'institution mais également les parlements nationaux. Il veille ainsi à ce que leurs besoins soient pris en compte et prévoit l'organisation régulière d'une conférence Eurovoc avec ces institutions nationales.

En tant que membre du comité de maintenance, le Parlement européen reçoit les propositions de descripteurs des parlements nationaux. Avant de les transmettre à l'OPOCE, il les étudie et les

filtre selon certains critères : rejet de descripteurs trop nationaux ou trop spécifiques. Les questions linguistiques peuvent parfois poser quelques problèmes de compréhension et de traduction si le concept n'existe pas dans une autre langue. Il est parfois difficile d'évaluer une proposition à partir d'une traduction. Ainsi, les russes envoient leurs propositions en anglais (chacun utilise alors cette langue comme langue pour correspondre).

Ces propositions et celles de la Bibliothèque du Parlement européen sont ensuite transmises au comité de maintenance d'Eurovoc qui les étudie et les intègre ou non dans le thésaurus. En outre, le Parlement européen fait remonter les préoccupations et besoins exprimés par les Parlements nationaux auprès du Comité de maintenance.

Un certain nombre de parlements nationaux se montrent dynamiques dans l'enrichissement d'Eurovoc (Parlements belge, espagnol, portugais, suédois, polonais). Les Parlements russe et croate sont également actifs. Par ailleurs, l'Assemblée Pan-Africaine a établi des contacts avec le Parlement européen et manifeste un grand intérêt pour Eurovoc.

Conférences Eurovoc

Le Parlement européen organise régulièrement une conférence Eurovoc avec les parlements nationaux. Cette conférence, autrefois annuelle, a lieu maintenant tous les deux ans. La prochaine se tiendra à Bruxelles en mars 2006. Cette conférence donne lieu à des échanges entre les utilisateurs et permet de présenter les modifications et les améliorations apportées aux nouvelles versions du thésaurus.

Certains parlements nationaux présentent leur travail avec Eurovoc : les problèmes de traduction rencontrés, les concepts existant dans une langue et pas dans une autre, les niveaux de spécificité qui ne répondent pas toujours aux besoins évoqués, les difficultés liées à la réindexation des bases, etc.

Beaucoup d'utilisateurs - dont la bibliothèque du Parlement européen avec le champ spécifique "descripteur-candidat" - ont développé un système propre à leur organisation. Ceci leur permet d'introduire des concepts propres à leur langue ou spécifiquement nationaux qui ne peuvent être intégrés dans la version officielle du thésaurus.

Divers représentants d'autres thésaurus multilingues et des spécialistes en informatique documentaire sont invités à ces conférences. Leurs interventions se révèlent très enrichissantes pour tous les participants.

Relations avec d'autres organismes concernant Eurovoc

Dans le cadre de projets interinstitutionnels, le Parlement européen, via son représentant au comité de maintenance Eurovoc, a établi des contacts avec d'autres organismes utilisant Eurovoc.

Des chercheurs du Centre commun de recherche de la Commission européenne ont développé un système d'indexation automatique avec Eurovoc. Ce système pourrait se révéler très utile à l'avenir aussi bien pour le Parlement européen que pour d'autres institutions.

Le Congrès espagnol travaille avec ce système et le résultat est tout à fait intéressant. Il permet une indexation rapide des documents. Il nécessite tout de même la supervision d'un indexeur qui élimine les descripteurs trop pointus.

Pour procéder à une indexation automatique, il est préférable d'avoir des textes complets. À la bibliothèque, les résumés ou les sommaires sont incorporés dans les notices lorsque c'est possible mais ce n'est pas suffisant pour indexer automatiquement. Cependant, le Parlement européen a participé à des tests de ce système d'indexation automatique.

Différents organismes travaillant dans le domaine de l'éducation et de la formation professionnelle souhaitent mettre en place un vocabulaire commun au niveau de leurs thésaurus respectifs et utiliseraient Eurovoc comme thésaurus référentiel. Ce projet d'interopérabilité sémantique des thésaurus de l'éducation est en cours d'élaboration.

Conclusion

L'usage d'Eurovoc prend de plus en plus d'ampleur au fil des ans. Le développement d'un vocabulaire commun aux institutions communautaires est un atout extrêmement important. En

harmonisant la description des fonds, il rend les échanges d'informations plus faciles et permet d'effectuer des recherches documentaires plus pertinentes.

L'octroi de licences ne cesse d'augmenter depuis plusieurs années. Cet outil documentaire apparaît de plus en plus utilisé aussi bien par les institutions européennes que par des parlements nationaux et même par des organismes publics ou privés des différents États membres. Sa couverture généraliste et multilingue est largement appréciée. Eurovoc permet également de développer en parallèle d'autres outils adaptés aux besoins de ces organismes.

L'entrée de nouveaux pays membres dans l'Union européenne a permis l'intégration de nouveaux concepts et la création de nouvelles versions linguistiques. Au cours des conférences Eurovoc, les parlements de la Géorgie, de l'Ukraine et de la Turquie ont manifesté un grand intérêt pour le thésaurus. Ainsi, le thésaurus s'ouvre sur le monde même s'il conserve une spécificité européenne propre aux institutions de l'Union. Tout ceci s'avère positif et enrichissant autant pour les gestionnaires du thésaurus que pour ses utilisateurs.

Nouer et développer les relations non seulement avec les parlements nationaux mais aussi avec les autres utilisateurs d'Eurovoc se révèle extrêmement important. C'est grâce à l'apport de chacun que la mise à jour du thésaurus se poursuit, reflétant ainsi l'actualité et les évolutions de la société, afin de répondre au mieux aux besoins de ses usagers.

Isabelle GAUTIER
Parlement Européen
Rue Wiertz, 60
1047 Bruxelles

22 février 2006

Liens utiles

1. Thésaurus Eurovoc : <<http://europa.eu.int/celex/eurovoc/>> (consulté le 22/02/2006).
2. Centre commun de recherche de la Commission européenne (indexation automatique avec Eurovoc) : <<http://www.jrc.it/langtech/Eurovoc.html>> (consulté le 22/02/2006).

CASH FOR KNOWLEDGE ?

ETHICAL IMPLICATIONS OF PATENTING ACADEMIC RESEARCH

Quatrième Forum Éthique de la Fondation Universitaire

Compte-rendu de Simone Jérôme,
Éditrice des *Cahiers de la Documentation*.

▪ Le brevet est-il une information scientifique en soi ? La question n'a aucun sens si l'on sait qu'il s'agit d'un contrat par lequel un (groupe d')état(s) accorde à un inventeur un monopole d'utilisation. Certes, pour exister, le brevet doit informer sur son objet mais rien n'oblige à ce que cette information soit vraie car le critère pris en compte est celui de l'originalité au sens propre c'est-à-dire ce qui n'a encore fait l'objet d'aucune demande de protection. La preuve en est les nombreux brevets pris sur des inventions farfelues défiant les lois fondamentales de la physique entre autres. L'excellente revue anglaise *New Scientist* avait ouvert à cette enseigne une rubrique dont la lecture ne manquait pas de sel. Mais les illuminés restent marginaux et le brevet est maintenant considéré comme une source d'information normale. Son importance est d'autant plus grande que notre civilisation occidentale est marquée par un libéralisme croissant qui se traduit par la marchandisation d'objets, de procédés et parfois même d'idées abstraites, qui jusque là échappaient à sa sphère d'influence. Le brevet est devenu un instrument et à ce titre, les universités sont de plus en plus appelées à en faire usage pour assurer la notoriété autant que la rentabilité de leurs recherches.

Cette transformation des objectifs affecte surtout la recherche appliquée mais il ne faut pas chercher très loin les implications pour la recherche fondamentale elle-même. La bataille autour du brevet du génome humain en est un exemple.

La Fondation Universitaire organisait, le 22 novembre dernier, dans le cadre de son Forum éthique, une après-midi de conférences et de réflexion sur le thème : **Cash for knowledge ? Ethical Implications of Patenting Academic Research**. Ce qui s'est dit lors de cette manifestation ne pouvait qu'intéresser les spécialistes de l'information scientifique. Nous vous en présentons ici un compte-rendu succinct.

Vous verrez que de nombreux problèmes soulevés mériteraient d'être aussi débattus au sein de notre association.

Le Forum Éthique de la Fondation Universitaire

L'Université a connu ces dernières années une mutation sans précédent. Confrontée à une pression grandissante des différentes composantes de la Société, à de nouvelles réalités technologiques, politiques et économiques, elle s'implique activement dans des débats et des activités qui lui étaient la plupart du temps étrangères. Bonne ou mauvaise, cette évolution est un fait et la Fondation Universitaire a eu l'excellente idée de créer en 2002 un "Forum Éthique" qui se veut un lieu où se discutent, librement et toujours de façon rigoureuse, les problèmes soulevés par les changements profonds de l'environnement universitaire dans l'économie de marché. Qu'ils soient chercheurs ou enseignants, mais encore administrateurs, responsables de ressources humaines, auteurs, utilisateurs de services, conseillers politiques, consultants auprès de firmes ou citoyens impliqués dans le débat public, les universitaires y sont appelés à partager leurs convictions et leurs questionnements à partir de problèmes spécifiques choisis dans diverses disciplines pour leurs aspects éthiques et moraux.

Les thèmes ne manquent pas : financement de la recherche, liberté académique, organisation du recrutement, de la publication, de l'utilisation des

langues, évaluation de la recherche...

La Fondation Universitaire a choisi de traiter ces thèmes sous la forme de séminaires d'une demi-journée, largement ouverts à la communauté universitaire où des spécialistes viennent exposer leur vision du problème et répondre aux questions du public.

Quatrième Forum Éthique : "Cash for knowledge ? Ethical Implications of Patenting Academic Research"

Bart Pattyn¹, professeur au Centre européen d'Éthique de la KUL, introduit le sujet du forum.

Sous ce titre provocateur "Cash for knowledge...", la Fondation Universitaire entend ouvrir un débat qui, pour certains, apparaît déjà comme clos. Le credo de la valorisation des résultats de recherche pour faire face à des coûts de fonctionnement de plus en plus élevés, est répété à l'envi dans l'université même et relayé dans le monde politique et économique. La production de savoir apparaît comme le chemin obligé de l'innovation et de la création d'emplois. Aucun doute ne serait possible.

Il n'en fut pas toujours ainsi. Traditionnellement, le savoir ne pouvait pas être un objet de com-

merce, de par sa nature, mais aussi parce que l'usage ne l'affectait pas et que, pour le mettre en lumière, il ne fallait en exclure personne.

Le brevet est-il à même de réconcilier ces deux logiques ? Les défenseurs de l'orientation commerciale disent qu'il est bon de secouer le vieux système par de nouveaux incitants qui augmenteront la pertinence de la recherche. Les opposants estiment que c'est le sens même de la mission académique et sa fonction sociale qui seront perdus.

Il faut admettre que notre société est complexe et que ce qui est vrai ou bon dans un domaine ne l'est pas nécessairement dans l'autre. Les normes sont multiples et il est juste de respecter leur caractère propre et d'éviter la colonisation d'un domaine par un autre. Notre dernier forum avait conclu que le monde académique devait respecter son propre système de valeurs indépendamment des pressions du pouvoir politique. En serait-il autrement des pressions économiques ?

L'université a, selon la tradition, la liberté de choisir ses champs de recherche et les méthodes qui lui permettent d'évaluer ses chercheurs ont rarement à voir avec les retombées économiques de leurs découvertes. Aussi longtemps que cette situation est garantie, il n'est pas à craindre que la communication des résultats de recherche soit biaisée par les intérêts financiers des sponsors et des investisseurs.

Les risques de l'influence commerciale sont peut-être exagérés par des académiques pessimistes mais qui peut le savoir ? Avancer sans règles, n'est-ce pas jouer à l'apprenti sorcier ?

Principes et justifications de la protection par le brevet

Le premier orateur est Alain Strowel², avocat, professeur aux Facultés Saint-Louis. Il dégage les principes de la protection par le brevet.

Longtemps confiné aux bureaux spécialisés, le brevet investit la sphère politique et, via la presse, alimente le débat public.

En se basant sur des cas concrets, l'orateur décrit le système européen de brevet et rappelle ce qui le distingue du droit d'auteur. Ce dernier protège essentiellement des utilisateurs abusifs tandis que le brevet confère un avantage compétitif sur la concurrence.

Il passe ensuite en revue les avantages attribués

au système, à savoir l'incitation à l'innovation, la création de cycles dynamiques favorables à la croissance économique, la promotion du financement de jeunes entreprises, la gestion des risques financiers.

La troisième partie de l'exposé porte sur l'évolution du droit des brevets. De moins en moins de secteurs sont exclus du brevet. On assiste à une dématérialisation du savoir technique, notamment depuis le brevet sur les produits de la biotechnologie et sur les technologies de l'information. Certains domaines restent exclus. C'est le cas du brevet de logiciel, refusé récemment par le Parlement Européen. Une réflexion d'ensemble et un affinement des lois sur le sujet est souhaitable et d'autres voies de divulgation peuvent être choisies, comme la publication.

Les brevets peuvent-ils être justifiés sur le plan de la morale ?

Professeur d'éthique au département de philosophie et de science morale à l'Université de Gand, Sigrid Sterckx se penche sur les trois arguments généralement avancés pour justifier le brevet.

Le premier est celui des droits naturels : l'Homme a un droit de propriété naturel sur ses idées et utiliser celles-ci sans son autorisation est un vol. Cette conception remonte à Locke qui, en 1690, dans son deuxième traité de gouvernement, expose sa théorie de la propriété par le travail. L'Homme a un droit naturel sur les moyens nécessaires à sa survie. Chaque être humain a un droit de propriété sur sa propre personne, par extension sur son travail et sur les choses qui en sont le fruit.

Locke prévoit néanmoins des limites au droit à l'appropriation des ressources : si l'on peut prendre pour répondre au besoin en suffisance, il faut en laisser pour les autres autant et de même qualité. L'on ne peut non plus prendre plus que l'on ne peut utiliser, ce qui serait du gaspillage.

Cette théorie s'applique-t-elle à la propriété intellectuelle ? Des problèmes se posent indéniablement. La valeur d'un objet dû au travail n'est pas la valeur du travail lié à cet objet. Des "objets intellectuels" peuvent naître des idées d'autrui. De plus, l'inventivité peut-elle être gaspillée ? Le brevet laisse-t-il à l'autre "assez et de même qualité" ?

Le deuxième argument est celui de la justice distributive : la justice exige que l'inventeur soit récompensé de son effort et protégé des agissements de ceux qui s'approprieraient les fruits de

son invention.

Mais que faut-il récompenser : la grandeur de l'effort ou la valeur du résultat ? Est-il juste de ne pas récompenser l'effort produit dans des domaines où le brevet ne s'applique pas ? Le brevet n'est-il pas responsable d'une distribution inéquitable des ressources ? Ne limite-t-il pas l'accès à l'information ? Qu'en est-il des autres moyens de récompenser l'effort ?

La recherche est un processus cumulatif qui implique la collaboration. Est-il juste que le "dernier" soit seul récompensé ? L'inventeur indépendant n'est-il pas défavorisé par le système ? L'équité est-elle respectée quand l'intérêt pour la société de l'invention n'est pas pris en compte, quand des brevets manifestement trop larges dans leurs exigences sont accordés ?

L'argument utilitaire est le troisième avancé. Le brevet est une incitation à inventer par la perspective d'un rendement. Sans cela, l'information tombe dans le domaine public et échappe aux lois du marché. Le contrat social qu'implique le brevet est une incitation à divulguer l'information de manière sûre. Il y a un paradoxe dans ce principe : le brevet limite la distribution et l'utilisation d'une invention et annule ainsi l'avantage de l'incitant. L'exigence d'information est rarement prise au sérieux par les offices de brevet. Quels sont les coûts réels du système ? Plutôt que d'être une réponse au sous-investissement, n'est-il pas un encouragement au sur-investissement ? Trop peu de recherches ont été menées sur le sujet, qui repose plus sur des articles de foi que sur le résultat d'expériences.

La preuve par l'économie : l'effet réel des incitants

Le professeur Paul Belleflamme (CORE et Louvain School Management à l'UCL)³ pose la question : le système de la propriété intellectuelle remplit-il sa mission ? Il est censé régir la relation entre la société et un innovateur potentiel dans le but de produire un bien commun : le savoir. L'innovation ne se justifie que si la valeur sociale attendue surpasse le coût social attendu et, partant, si la compensation qu'en attend l'innovateur dépasse la somme qu'il lui en coûte. Le système doit être particulièrement protégé dans le cas où le processus de création est coûteux et dans le cas où le risque, c'est-à-dire la probabilité que la recherche n'aboutisse pas, est grand. D'un point de vue strictement économique, si l'exclusivité n'est pas accordée pour une production donnée, il ne peut y avoir de vente et donc il y a un risque de sous-production. Par contre, l'exclu-

sivité conduit à un monopole et si le client ne veut pas payer le prix imposé, il y a sous-utilisation et donc inefficacité. Les lois de la propriété intellectuelle sont conçues pour instaurer un équilibre entre ces deux situations conflictuelles, c'est pourquoi leur effet est limité dans le temps et leurs exigences arrêtées. Un tel équilibre est forcément imparfait puisque les risques sont contradictoires.

Des solutions alternatives existent comme le financement par les pouvoirs publics ou le secret. La première introduit des distorsions dans le marché et, si elle n'est pas suffisante, elle perd sa valeur d'incitant. Quant à l'absence de diffusion de l'information, elle a un coût pour la société et elle n'offre pas de protection contre les innovations indépendantes. Depuis les années 80, on assiste à un renforcement des lois sur la propriété intellectuelle, aux USA comme en Europe, à une recherche d'harmonisation au niveau international et à son extension à de nouveaux domaines. Le nombre des brevets accordés a triplé entre 1980 et 2001. L'innovation a crû rapidement. Y a-t-il un lien entre ces tendances ?

En l'absence d'autres méthodes de mesure, le brevet est de plus en plus considéré comme un indicateur de l'innovation. Toutefois, si le rapport avec l'inventivité existe, si l'état de la technologie, en l'absence de données plus explicites, est bien reflété et si des informations importantes peuvent en être tirées, des méthodes standard pour mesurer les effets font défaut et celles qui sont proposées, conduisent à des résultats contradictoires. Beaucoup de brevets n'ont pas d'application industrielle ; certaines inventions échappent au système, soit qu'elles ne soient pas brevetables soit que l'inventeur ait choisi un autre mode de protection ; la perception du brevet et la législation afférente change selon les pays. Le brevet est le meilleur indicateur d'innovation disponible en l'absence d'autres plus fiables. Comment, dès lors, tirer le meilleur parti du brevet en termes d'innovation ? On peut affiner la mesure en pondérant l'importance de chaque brevet par le nombre de références auxquelles il donne lieu dans d'autres brevets, encore que toutes n'aient pas la même valeur, en examinant les efforts de renouvellement et les batailles juridiques auxquelles un brevet peut donner lieu, efforts coûteux qui supposent une valeur intrinsèque, et en s'informant auprès des inventeurs eux-mêmes.

L'activité du brevet, concentrée dans quelques pays, a connu une forte augmentation depuis le début des années 90. Elle connaît aussi une croissance rapide dans les pays émergents.

Pourtant, en dehors des firmes chimiques et

pharmaceutiques, la protection par le brevet est généralement réputée inefficace. Beaucoup de firmes préfèrent le secret ou des stratégies basées sur l'avantage d'être le premier. Les délais de mise en oeuvre, les temps d'apprentissage, les efforts de vente et de service sont autant de techniques que les gestionnaires utilisent pour protéger la propriété industrielle. Pour eux, le brevet ne décourage pas les imitateurs, il est coûteux et l'obligation de dévoiler certains détails peut être préjudiciable.

Pourquoi les firmes prennent-elles la peine de breveter dans ces conditions ? Le brevet sert à mesurer la productivité d'une division Recherche et Développement. Il répond à une demande des investisseurs pour bloquer la concurrence et disposer d'actifs à vendre en cas de faillite. Les portefeuilles de brevets sont ainsi utilisés comme outils de commerce. Les grandes firmes tirent de plus en plus de revenus de leurs licences d'exploitation.

Et l'effet sur l'innovation ? L'accroissement du nombre des brevets correspond à une augmentation des budgets de R&D mais la prolifération gêne-t-elle ou accélère-t-elle l'innovation ? Cela dépend du point de vue. Elle l'encourage par la création d'un "marché" de la technologie mais elle la stérilise en créant des barrières, de coût notamment, et en limitant l'accès public. La question reste donc ouverte.

Les universités et les brevets : de l'"Open Science" à l'"Open Innovation"

Gilles Cappart, président de ProTon Europe⁴, révèle que 70% de la Recherche/Développement "pré-compétitive" est effectuée en Europe par des organismes de recherche publics. Alors que 70% des brevets US sont basés sur des recherches émanant de telles institutions, les entreprises européennes interrogées leur attribuent seulement 5% de leurs innovations. Le transfert de savoir technologique vers l'économie ne serait donc pas efficace en Europe.

Dans le modèle scientifique traditionnel, l'Université ne touche aucun droit de propriété intellectuelle si ce n'est celui, symbolique, de la citation et des droits d'auteur sur certaines de ses productions. Il n'y a aucun incitant pour investir dans les applications de ses découvertes puisque aucune protection n'existe. La contribution à l'innovation n'est qu'indirecte et repose presque entièrement sur l'industrie.

Les universités doivent-elles jouer un rôle actif dans l'innovation ? Les états exercent une pres-

sion grandissante dans ce sens et ont adopté des législations ad hoc. Beaucoup de programmes de financement de la recherche, européens, nationaux et régionaux, sont orientés vers de possibles applications commerciales. En Angleterre, l'évaluation de la recherche inclut déjà des indicateurs de transfert technologique. Cette tendance est faite pour durer, il en va de la reconnaissance et de la survie de la recherche.

Les découvertes et les idées ne peuvent être brevetées, seules les applications techniques le sont. Le brevet protège les investissements pour développer des applications ; il décrit les inventions pour permettre leur amélioration et la recherche subséquente. L'Université ne peut se protéger par le secret de fabrication ; elle doit publier sans délai pour participer au réseau mondial de la recherche scientifique. Elle ne peut exploiter directement ses inventions ; elle doit vendre des licences. Ses inventions en sont au premier stade de développement et doivent subir des améliorations avant de devenir économiquement attrayantes. Le brevet réconcilie la publication et l'innovation.

Comment le brevet peut-il devenir un instrument du transfert de savoir ? Il augmente la probabilité qu'une découverte conduise à des applications. Il favorise les conditions de licences compatibles avec le bien public. Il motive à la fois les inventeurs et les chercheurs. Il oblige à une professionnalisation de la gestion de la propriété intellectuelle. L'auteur compare les systèmes américain et européen et en conclut que les universités feraient mieux de breveter aux Etats-Unis uniquement ! Plusieurs modèles sont possibles. Le premier est la prise de licence, aussi plus efficace aux Etats-Unis depuis la loi Bayh-Dole. L'auteur regrette, entre autres, le manque d'intérêt de l'industrie européenne pour la recherche publique. Le modèle d'interaction est plus dans la continuité des relations existantes, il concerne des développements à court terme. La négociation de la propriété intellectuelle est un levier pour augmenter la capacité de recherche des institutions. Enfin le modèle "Spin-out" permet d'utiliser la technologie pour introduire de nouveaux concepts dans l'économie. Il vise à la rajeunir à condition de lui en laisser le temps. Le processus est lent mais il marche en Europe.

Réconcilier les politiques de brevet avec la mission de l'Université

Pour Geertrui Van Overwalle, professeur de droit à la KUL, l'information et l'acquisition de données sont fondamentales pour le progrès de la science et l'innovation. Les universités sont des institu-

tions-clés de "l'économie du savoir" et sont soumises à des pressions pour investir dans la production et le progrès scientifique. Traditionnellement, ce dernier concept remontant aux XVI^{ème}-XVII^{ème} siècles est lié à un idéal de dissémination libre et ouverte de l'information. Bien qu'aucune règle n'ait jamais été édictée à ce sujet, les valeurs propres à cet idéal font partie de la culture académique et sont assumées par la communauté scientifique. Maintenant, les universités sont fermement invitées à monnayer le potentiel commercial de leur recherche et à participer activement à l'économie de marché. L'une des voies proposées est la protection du savoir académique par le brevet. La restriction d'usage remplace la liberté d'usage créant ainsi un grave dilemme pour le chercheur. Les deux positions peuvent-elles être conciliées et comment ?

L'auteur ne prend pas parti sur les aspects moraux de la question, d'ailleurs dépassés par les politiques actuelles qui encouragent la prise de brevet dans le monde académique, une tendance probablement irréversible. Elle propose dès lors deux séries de mesures indispensables.

Il faut protéger la recherche individuelle et collective des torts que le dépôt de brevet peut causer dans la pratique, que ce soit le retard à la publication des résultats ou leur manipulation. La nouvelle culture académique doit se doter de règles claires, éventuellement d'un code de conduite pour prévenir abus et conflits d'intérêt. Des mécanismes d'intervention et d'arbitrage doivent être mis en place pour régulariser les relations entre les centres académiques et leurs partenaires industriels.

Un deuxième train de mesures dicté par le réalisme visera à réconcilier les propositions antagonistes de privatisation et de partage du savoir en créant de nouvelles catégories de brevets. Le brevet public garantit l'usage libre et sans réserve du savoir divulgué ; les licences sont accordées sans acquiescer de droit. Le brevet "libre accès" assure un usage libre à condition que les améliorations du brevet soient elles aussi partagées ; les licences ne peuvent en aucun cas porter sur les fondements de la technologie et de ses améliorations. Ce système ne génère pas de revenus directs mais des avantages indirects. Ces deux types de brevets ne générant pas de rentrées financières, il est peu probable qu'ils s'imposent en dehors de cas où le financement est assuré par service public. Dans le système du brevet à deux niveaux par contre, le niveau des droits à acquiescer dépend du statut des utilisateurs ou du but qu'ils poursuivent. Ainsi les droits sont nuls pour des projets humanitaires ou de développement tandis qu'ils sont calculés de manière habituelle pour toute compagnie les uti-

lisant dans un projet commercial. Pour que ce dernier système fonctionne, il faut absolument que le principe en soit clairement énoncé dès le départ des négociations et mieux encore qu'il soit incorporé aux codes de conduite nationaux applicables aux universités.

Cet exposé est en outre émaillé d'exemples et se complète d'une bibliographie fournie.

La société mondiale de l'information : l'Université et le brevet

Christopher May, professeur d'économie politique à l'Université de Lancaster ⁵ part du même constat sur la société mondiale de l'information pour parler des interactions des universités avec le système des droits de propriété intellectuelle, encore que, pour lui, l'information dans laquelle notre société est plongée jusqu'à saturation a changé les relations sociales plus que la substance de la société elle-même et que ses effets sont loin d'être universels. Jusqu'il y a peu, les universités et les entreprises s'ignoraient dans la plupart des cas. La recherche de valeur ajoutée pour les entreprises et la pression sur les budgets de recherche pour les universités ont provoqué un nouvel intérêt pour la propriété intellectuelle et son exploitation marchande. Aux États-Unis, la loi Bayh-Dole en 1980 a donné aux universités le droit de commercialiser des licences pour des inventions ayant bénéficié d'un financement public. Une tendance similaire, encore que moins marquée, prévaut aussi en Europe.

Certains aspects de la science pure débouchent sur des applications et se voient de fait impliqués dans le commerce. La recherche de propriété et la production de biens de consommation se font pressantes. Certain contenu intellectuel, qui de tout temps faisait partie du domaine public (savoir dont on ne peut s'attribuer la propriété), peut être re-qualifié comme propriété "relative au commerce" donc privatisée et traitée comme bien de consommation. Cet état de fait est souvent banalisé mais la frontière entre le savoir commercialisable et celui qui ne l'est pas, est flou et dépend des secteurs et des technologies qui l'utilisent.

La loi organise les relations sociales entre celui qui possède et celui qui ne possède pas. En ce sens, pour ce qui est de la propriété intellectuelle, elle crée la rareté là où elle n'existait pas et fait monter les prix. L'exemple de la transformation de la recherche bio-médicale en industrie de la biotechnologie illustre bien ce passage de l'effort scientifique gratuit à l'industrialisation

d'un secteur. Ainsi le rôle du brevet comme support de l'innovation cède la place au brevet, agent de commercialisation. Les universités sont coincées entre deux demandes sociales contradictoires : produire du savoir et de l'information comme service public et supporter le développement économique générateur de bien-être selon la logique capitaliste.

Le savoir n'est plus un tout complexe et multifacettes mais se découpe en modules à vendre et à acheter, ce qui le prive de son caractère social. Les conséquences peuvent s'avérer dramatiques comme dans le cas des médicaments contre le SIDA refusés aux pays pauvres. Le savoir est un élément du pouvoir économique dans la société de l'information.

Si l'université utilise le brevet pour protéger sa recherche, l'industrie pourrait logiquement revendiquer une limite de la brevetabilité. L'histoire nous montre que c'est au contraire l'industrie qui pousse à toujours plus de brevets. Pour les industriels, le droit à l'exclusivité et l'avantage qu'il confère sur le marché sont plus importants que le prix des licences sans compter des apports externes : savoirs traditionnels, bio-piraterie...

Par ailleurs, les universités peuvent voir d'un autre oeil leur rapport avec la société de l'information en se plaçant comme pourvoyeurs d'innovation et de savoir en dehors de tout but commercial. Les entrepreneurs peuvent alors utiliser ce savoir dans leurs activités sans se l'approprier (ex: MIT OpenCourseWare pilot project). Cette ouverture s'inspire de l'exemple du logiciel libre. Elle concerne l'aspect social de la dissémination du savoir et n'affecte pas la recherche du profit sauf par la voie d'un monopole.

La difficulté de ce modèle est que les autorités de financement ne peuvent plus exiger de rendus par ce biais. En fait, le modèle d'une institution impliquée socialement, soutenant le domaine public du savoir pour le bien général, est celui sur lequel l'institution a toujours fonctionné mais sans faire d'effort pour communiquer largement et librement le savoir produit.

Une université humaniste a-t-elle un futur ? Si jusqu'à présent un juste équilibre entre les droits individuels et les bénéfices sociétaux de l'accès au public a été maintenu, il est devenu résiduel par rapport à une marchandisation de plus en plus mise en avant. La défense de l'intérêt public devrait dès lors impliquer une participation dans des stratégies de diffusion plus actives. Par bien des aspects, la libéralisation s'attaque au bien public que les lois sur la propriété intellectuelle sont censées protéger. L'université doit être à l'avant-garde de sa défense. Pour l'auteur, la

question de la propriété intellectuelle n'est pas liée aux intérêts de l'université mais à ceux d'acteurs industriels privés. Ce sont les intérêts économiques qui demandent d'étendre le brevet dans les domaines de la science fondamentale. Le rôle de l'OMPI vise à harmoniser les règles du brevet à l'échelon international. Le traité en préparation entend imposer une limite aux restrictions que les états pourraient imposer compte tenu de leur situation propre. La difficulté d'en finaliser le texte devrait inciter les universités à s'impliquer dans ce débat.

Les forces politiques opposées au domaine public sont puissantes et la logique du commerce est largement acceptée comme une stratégie propre à garantir une utilisation efficace des ressources (ici le savoir et l'information). La nécessité d'un domaine public de valorisation scientifique n'est pas un argument de défense suffisant. Les universités doivent se positionner et mettre en pratique leur rôle de créateurs d'un savoir "ouvert" comme il a déjà été démontré dans le domaine du développement de logiciel qui ne remet pas en cause la propriété intellectuelle mais en limite l'application à des domaines d'intérêt général. Elles seraient bien avisées de s'inspirer de cette expérience. Elles doivent entretenir deux logiques contrastantes : celle de la marchandisation pour protéger certaines de leurs productions et celle de l'ouverture de certains de leurs travaux au libre accès. Même si leur situation financière ne s'améliore pas dans un proche futur, ce n'est pas une raison pour perdre de vue leur rôle social et ne pas défendre l'accès à des innovations importantes sur ce plan. Brader la valeur sociale de l'université conduirait à perdre tout espoir de la rétablir dans son rôle non-commercial.

Conclusions

Des questions qui suivirent la présentation des orateurs, il apparaît que le sujet interpelle le monde universitaire même si sa perception varie d'un domaine à un autre. Les problèmes éthiques soulevés restent nombreux et plusieurs intervenants ont posé la question d'un code de conduite universel. La proposition est d'autant plus urgente que les accords ADPIC de l'OMC consacrent définitivement les aspects commerciaux de la propriété intellectuelle à l'échelle mondiale. Le débat reste ouvert car tout le monde n'a pas encore pris conscience des enjeux. Certaines universités s'y impliquent plus que d'autres. L'initiative de la Fondation Universitaire vient ainsi à son heure.

La plupart des documents originaux sont accessibles sur le site de la Fondation Universitaire⁶.
Les textes des contributions seront publiés dans la revue *Ethical Perspectives*⁷.
Le prochain Forum Éthique aura lieu le jeudi 23 novembre 2006. Le thème en sera : "The end of free entry?
Can university admission tests and numerus clausus rules make our higher education system more cost-efficient and more socially responsible? "

Notes

¹ <<http://www.kuleuven.ac.be/oce/page.php?LAN=E&ID=2&FILE=persfiche&>> (consulté le 5 mars 2006)

² <http://www3.fusl.ac.be/4D ACTION/rechw_detail_chercheur/92527/F> (consulté le 5 mars 2006)

³ <<http://www.core.ucl.ac.be:16080/~pbel/>> (consulté le 5 mars 2006)

⁴ <http://www.cordis.lu/finance/src/g2g_proton.htm> (consulté le 5 mars 2006)

⁵ <<http://www.lancs.ac.uk/fss/politics/people/may/may.htm>> (consulté le 5 mars 2006)

⁶ <<http://www.fondationuniversitaire.be/fr/forum4.php>> (consulté le 5 mars 2006)

⁷ <<http://poj.peeters-leuven.be/content.php?url=journal.php&code=EP>> (consulté le 5 mars 2006)



MANUEL DE BIBLIOGRAPHIE GÉNÉRALE

Marie-Hélène PRÉVOTEAU et Jean-Claude
UTARD – Novembre 2005 – Format : 17 x
24 cm – 524 p. – 44 € - ISBN 2-7654-0912-9 –
ISSN 0184-0886.

La recherche bibliographique et documentaire est une activité essentielle et en constant renouvellement dans nos sociétés de l'information et de la communication. C'est également une discipline où s'affirment, depuis longtemps, le savoir et la compétence des bibliothécaires et des documentalistes.

La nouvelle édition de ce manuel prend en compte l'ensemble des modifications que le développement d'Internet et des moteurs de recherche, la mise en ligne sur le Web des bases de données et des catalogues de bibliothèques, mais aussi l'évolution des logiciels dans les outils documentaires, les transformations éditoriales des produits bibliographiques et les mutations des demandes des publics apportent à la bibliographie et au travail des professionnels de la documentation.

La méthodologie de la recherche documentaire, les modes d'organisation, de classement et d'accès ou d'interrogation des bibliographies imprimées ou informatisées sont étudiées en premier et introduisent le chapitre consacré aux outils de recherche sur Internet.

Les chapitres suivants mettent l'accent sur la recherche d'une information synthétique qui répond à la fonction de renseignement et d'accueil des publics diversifiés qui fréquentent bibliothèques et centres de documentation. Les

encyclopédies, les dictionnaires, les annuaires imprimés, sur cédéroms ou en ligne rendent disponible cette information.

Les bibliographies et catalogues permettant d'acquérir, d'identifier et de localiser livres, revues, documents audiovisuels, articles de périodiques et cédéroms sont ensuite décrits et leur utilisation est détaillée. Les outils de repérages des thèses et des documents numérisés, en particulier ceux qui s'inscrivent dans le mouvement des archives ouvertes, sont également présentés.

Les instruments cités, références de base dans leur spécialité, sont accompagnés de descriptions et commentaires méthodologiques, enrichis de fac-similés et de nombreuses copies d'écran présentant certaines de leurs caractéristiques essentielles. L'ouverture sur l'international, rendue possible par la mise en ligne sur Internet de nombre de ces outils, est une des caractéristiques de cette nouvelle édition.

Très pratique, ce manuel s'adresse à tous les étudiants bibliothécaires ou documentalistes, mais aussi à tous les professionnels du livre et de la documentation déjà en poste et confrontés à l'accroissement des instruments et des tâches bibliographiques.

LA LECTURE NUMÉRIQUE : RÉALITÉS, ENJEUX ET PERSPECTIVES

Divers collaborateurs - Coordonné par Claire
BELISLE – Avril 2004 – Format : 16 x 24 cm – 293 p.
– 40 € - ISBN 2-910227-51-0 – ISSN 1621-3084.

En se déplaçant progressivement du papier vers l'écran, les pratiques de lecture s'adaptent et se transforment. Dans quelle mesure la place prise par l'information dans la société et la prépondérance des technologies et des supports numériques, en bouleversant les habitudes liées aux textes, mettent-elles en péril certaines de ces pratiques ? Comment lirons-nous demain ?

En analysant les caractéristiques des nouveaux médias et en rendant compte des expériences de lecture d'oeuvres numériques, cet ouvrage fait le point sur la place actuelle et les perspectives de la lecture sur les supports numériques, et particulièrement sur les livres électroniques.

Les divers chapitres donnent des informations factuelles et des points de repère sur la lecture

sur écran, sur son avenir et sur les conditions de son utilisation notamment par les bibliothèques.

Les apports des différents spécialistes qui ont contribué à cet ensemble constituent un précieux outil pour appréhender l'impact culturel de la lecture à l'écran, la transformation des tâches et des rôles des auteurs, éditeurs, libraires et bibliothécaires.

BIBLIOTHÈQUES ET DOCUMENTS NUMÉRIQUES. CONCEPTS, COMPOSANTES, TECHNIQUES ET ENJEUX

Alain JACQUESSON et Alexis RIVIER –
Décembre 2005 – Format : 17 x 24 cm –
573 p. – 42 € – ISBN 10 2-7654-0915-3 – ISBN 13
978-2-7654-0915-1 – ISSN 0184-0886.

Les bibliothèques numériques mettent en cause un des acquis les plus fondateurs de notre civilisation : l'imprimé. Production et diffusion des textes sont en pleine mutation ; auteurs et lecteurs ne sauraient travailler sans le recours à l'informatique. Les collections imprimées conservées depuis des siècles sont de plus en plus rendues accessibles par la numérisation.

Depuis la première édition de ce livre, le sujet a provoqué maints débats parmi les professionnels. Une société informatique en vue a même candidement annoncé sa volonté de diffuser ainsi le patrimoine de l'humanité. La presse s'est emparée du concept de bibliothèque numérique et l'a répandu dans l'opinion publique.

Une synthèse en français paraît d'autant plus nécessaire. Clarifier les enjeux et expliciter les techniques permet de mieux comprendre l'usage et la nécessité des documents électroniques, matière première des bibliothèques numériques. Les sciences exactes et biomédicales ne conçoivent tout simplement plus leurs activités sans les périodiques électroniques. Les sciences humaines les rejoignent sur ce terrain et ont constitué d'importants corpus de textes classiques aussi bien que d'œuvres peu connues.

Comment les bibliothèques réagissent-elles ? Elles doivent s'adapter. L'intégration des documents numériques dans le traitement bibliographique est devenue nécessaire, avec toutes les questions que cela pose selon qu'ils sont loués, achetés, numérisés par la bibliothèque elle-même ou encore dispersés sur Internet. La préservation à long terme des supports numériques est également un souci central.

Pour les lecteurs, les avantages sont certains. Outre un accès plus aisé aux documents, même fragiles et rares, se développe une manière de travailler plus souple : étudiants et chercheurs constituent des corpus personnels, annotent électroniquement les textes, tissent leurs propres réseaux de liens hypertextuels.

Dépassées par les bibliothèques numériques les bibliothèques traditionnelles seraient-elles vouées à disparaître, absorbées par Internet ? Les auteurs de ce livre démontrent qu'il n'en est rien. Mais un nouvel espace documentaire, intégrant à la fois imprimé et électronique a vu le jour.



RECHERCHE D'INFORMATION ET TRAITEMENT DE LA LANGUE : FONDEMENTS LINGUISTIQUES ET APPLICATIONS

Geneviève LALLICH-BOIDIN et Dominique MARET – Préface de Serge CHAMBAUD –
Novembre 2005 – Format : 16 x 24 cm – 288 p. –
ISBN 2-910227-60-X – ISSN 01834-9962.

Aujourd'hui les recherches documentaires dans des collections numérisées sont devenues incontournables et quotidiennes pour qui doit s'informer, étudier ou se former. Si l'utilisation des outils et des moteurs est familière, les mécanismes qu'ils emploient ne sont eux pas forcément connus.

Cet ouvrage est consacré aux connaissances fondamentales, à savoir, les bases de la linguistique et de la logique informatique pour comprendre les traitements auxquels les requêtes et les textes sont soumis. Il réserve une part importante aux techniques existantes comme la recherche sur la proximité phonétique, syntaxique ou sémantique, étendue au multilinguisme. Des notions apparemment banales comme mot vide, terme, synonyme, lemmatisation sont expliquées dans ce contexte.

Des exercices corrigés et des exemples issus d'applications diverses comme les pages jaunes de l'annuaire, les bases de brevets, et même la Toile appuient les apports théoriques.

Table des matières

Sommaire

Préface de Serge Chambaud

Introduction

1. La segmentation d'un texte
2. Lemmatisation : de la surface à la syntaxe
3. Syntaxe
4. Au-delà de la syntaxe
5. L'affixation, créatrice de termes
6. Terminologie et lexicalisation
7. Les grammaires de réécriture

8. Recherche d'information par des noms propres et des dénominations
9. Recherche d'information en langue naturelle
10. Ressources linguistiques multilingues
11. Indexation de documents et de rubriques
12. Extraction de données à partir de textes

Bibliographie

Conclusion

Index général

Table des matières

BIBLIOTHEEK- & ARCHIEFGIDS
Vol. 81, nr 6 (december 2005)

- *Bewaarbibliotheken in Vlaanderen: ideeën voor een beleid* – Ludo SIMONS – (p. 5-13) – (1+14 ref.).

Het is dringend noodzakelijk dat de Vlaamse overheid, in overleg en samenwerking met de provinciale en lokale besturen, werk maakt van een beleid voor de Vlaamse bewaarbibliotheken. Het gaat hier om een zeer disparate groep van grote en kleine instellingen, die alle samen een belangrijk documentair patrimonium bewaren, maar waarbinnen vaak expertise ontbreekt op het stuk van preservering, conservering, restauratie en bibliografische ontsluiting, om nog niet te spreken van financieel en technologisch veel ingrijpender processen zoals digitalisering. Dit netwerk van bibliotheken zou ten eerste gebaat zijn met de oprichting van een coördinatiecentrum, dat de samenwerking tussen deze instellingen bevordert en de mogelijkheden schept voor de professionalisering ervan via structurele en projectsubsidies. De Stadsbibliotheek Antwerpen, die zich kandidaat heeft gesteld om de zetel te worden van dit coördinatiecentrum, verdient op grond van haar grote traditie, haar geheel eigen profiel als "erfgoedbibliotheek en haar bereidheid tot dienstbare samenwerking binnen het netwerk, tot deze rol geroepen te worden.

(J.H.)

- *Bewaarbibliotheken, e-erfgoed en digitale cultuur* – Jeroen WALTERUS – (p.19-25) – (54 ref.).

De omgang met cultureel erfgoed staat onder druk van technologische innovatie, die de motor is van de maatschappelijke omslag naar een "digitale of technologische cultuur". Erfgoedinstellingen zoals bewaarbibliotheken, erkennen deze evolutie en wensen een inhaalbeweging voor de aanmaak, de ontsluiting en de representatie van "erfgoed" of digitaal erfgoed. De strategische opties die hierbij genomen worden, kunnen het best ingepast worden in een integrale beleidaanpak van e-cultuur, met het oog op een open, digitale cultuurpro-

ductie en -verspreiding in Vlaanderen. Een essentiële voorwaarde is dat bewaarbibliotheken kunnen participeren aan expertiseplatformen voor digitaal erfgoed en dat de overheid hiertoe de nodige ondersteunende en coördinerende maatregelen treft. In dit artikel wordt breed geteenderd op een aantal ontwikkelingen en evoluties rond digitaal erfgoed en e-cultuur die voor bewaarbibliotheken van belang zijn.

(J.H.)

- *Bespiegelingen over het wiede- en afvoerbeleid in openbare bibliotheken* – Jean-Paul DENAYER – (p. 32-34).

De levensloop van een collectie item in de lokale bibliotheek eindigt met de afvoer. Het wieden en afvoeren van de collectie is om meerdere redenen een essentiële taak voor het bibliotheekbeleid. In het bibliotheekdecreet wordt dan ook vooropgesteld dat de collectie "actueel, pluriform en representatief" moet zijn. De brochure opgesteld door het Streekgericht Bibliotheekbeleid van provincies en VGC en door het VCOB wil hier een aantal aanbevelingen meegeven. Wieden en afvoeren zijn delen van een drieluik. Een goed wiede- en afvoerbeleid wordt ondersteund door een performant bewaarbeleid

(J.H.)

- *Universitaire bewaarbibliotheken onder druk* – Dirk VAN ELDERE en Martine DE REU – (p. 39-43) – (6 réf.).

Deze bijdrage gaat in op de plaats van de bewaarbibliotheek in het universitaire bibliotheeklandschap van Leuven en Gent. Ruimte- en personeelsgebrek dwingen de conservatoren om gerichter te reflecteren over collectievorming. Deze houding is al langer ingeburgerd in bijvoorbeeld de archiefwereld. De Vlaamse conservatoren beseffen dat geen enkele bewaarbibliotheek in Vlaanderen geheel alleen alle opdrachten kan vervullen. Samenwerking en concrete afspraken tot taakverdeling zijn bijgevolg absoluut vereist.

(J.H.)

BIBLIOTHEEK- & ARCHIEFGIDS
Vol. 82, nr 1 (februari 2005)

- *Na het jaar 2000 en de euro nu ISBN 13* – Rosemie CALLEWAERT – (p. 7-14).

In 2005 is de publicaties gestart van de nieuwe ISBN standaard die in 2007 geïmplementeerd zal worden. Een overgangperiode is voorzien opdat iedereen zich zou kunnen voorbereiden op het gebruik

van boekidentificatienummers met 13 in plaats van 10 karakters. Zijn bibliotheken klaar voor de komst van de nieuwe ISBN-standaard, en hoe kunnen zij zich voorbereiden op de omschakeling?

Een standaard omschakelingsprocedure voorschrijven is niet mogelijk omdat de manier waarop het ISBN gebruikt wordt per organisatie kan verschillen. Het is wel mogelijk enkele aandachtspunten en aanbevelingen op te sommen zodat onaangename verrassingen in 2007 vermeden worden.

(J.H.)

- *De auteurswet en de bibliotheek* – Julien VAN BORM – (p.15-24) – (17 ref.).

De Belgische auteurswet van 30 juni 1994 is op 22 mei 2005 ingrijpend gewijzigd om ze in overeenstemming te brengen met de Europese richtlijn 2001/29/EC van 22 mei 2002 over de harmonisering van bepaalde aspecten van het auteursrecht in de informatie maatschappij. Dit artikel geeft een overzicht van de Belgische auteurswet voor privégebruik en voor gebruik door instellingen van onderwijs en onderzoek en de sociaal-culturele sector. De klemtoon ligt bij de bibliotheken: reproductie op papier, reproductie op gelijk welke drager, mededeling aan het publiek, leen...

Het artikel eindigt met een pleidooi en een vraag.

Een pleidooi voor meer transparantie in de wet zelf en bij de beheersvennootschappen die de rechten innen en verdelen.

Een vraag naar meer aandacht en bijstand van de overheid in deze materie voor onderwijs en onderzoek, de sociaal-culturele sector en de bibliotheken.

Auteursrecht en de afwikkeling daarvan is immers de enige business van de beheersvennootschappen, het is een core business voor de verenigingen van auteurs en uitgevers, het kan niet meer zijn dan een side business voor al diegenen die geacht worden de wet correct toe te passen en de vereiste vergoedingen te betalen. Deze geledingen in de maatschappij rekenen erop dat de overheid hun belangen behartigt.

(J.H.)

- *Package deals uitgekapt... of de bibliothecaris verpakt? "Koppelverkoop" in de elektronische bibliotheek* – Patricia HELLRIEGEL en Kaat VAN WONTERGHEM – (p.25-31) – (25 ref.).

Elektronische tijdschriftenovereenkomsten zoals ze door bibliotheekconsortia wereldwijd worden gesloten, bieden toegang tot grote groepen tijdschriften van één bepaalde uitgever. De aantrekkingskracht die uitgaat van

deze pakketten met grote aantallen kwaliteits-tijdschriften, is onweerstaanbaar gebleken voor wetenschappelijke bibliotheken en de onderzoeksgemeenschap die zij vertegenwoordigen, zo ook in Vlaanderen. In een bibliotheekcontext waar toegang stilaan de overhand krijgt op bezit, is de laatste jaren steeds duidelijker geworden dat deze wijze van "collectievorming" enkele negatieve effecten heeft. Bibliotheken voelen steeds meer hun grip op budget en inhoud van hun collectie verminderen. Waar is de grens? De bibliothecaris mag zich niet langer laat inpakken. Maar kan hij "neen" leren zeggen tegen de bundeling van tijdschriftencollecties en weigeren akkoord te gaan met "gedwongen acquisitie". Het moment is aangebroken om gemeenschappelijk werk te maken van de alternatieven die al geruime tijd voorhanden zijn.

(J.H.)

DOCUMENTALISTE – SCIENCES DE L'INFORMATION Vol. 42, n°4-5 (octobre 2005)

- *Gérer et diffuser les thèses électroniques : un choix politique pour un enjeu scientifique* – Diane LE HÉNAFF-STIETEL et Catherine THIOLON ; avec une contribution de Magalie PRUDON – (p.272-282) – (9 réf.).

À l'heure où, au sein de la communauté scientifique, s'affirme le mouvement pour les archives ouvertes, le libre accès aux thèses est devenu une préoccupation majeure des universités, écoles d'ingénieurs et autres établissements d'enseignement supérieur et de recherche. La diffusion des thèses électroniques s'inscrit dans ce souci de rendre visibles et accessibles des documents scientifiques validés mais non publiés. Après avoir rappelé les modalités techniques de dépôt, de traitement et de diffusion des thèses, cette étude évoque le projet français STAR de dépôt, signalement et archivage de ce type de document, et insiste sur les enjeux scientifiques d'une politique nationale de diffusion électronique des thèses.

(J-M. R.)

- *Certification et labellisation : un problème de confiance. Bref panorama de la situation actuelle* – Éric SUTTER – (p. 284-290) – (6 réf.).

Au carrefour de la démarche marketing et de la démarche qualité, la certification a pour but de mettre en confiance la clientèle ou les usagers. Mais derrière ce terme se cachent plusieurs systèmes ou dispositifs qu'il est nécessaire de connaître et d'analyser. Du point de vue du

producteur de services d'information, il convient de savoir quel type de certificat obtenir; du point de vue de l'acheteur ou de l'utilisateur de services d'information, il faut aussi savoir ce que le certificat recouvre réellement. Cette étude présente succinctement les principaux systèmes de certification qui existent dans le champ de l'information professionnelle, en les situant les uns par rapport aux autres.

(J-M. R.)

- *Droit des bibliothèques – Droit des usagers* (Congrès annuel de l'ABF, Grenoble, 17-21 juin 2005) – Michèle BATTISTI – (p. 292-296) – (7 réf.).

En une période où le projet de loi sur le droit d'auteur est sur le point d'être examiné par les parlementaires français, le cinquante et unième congrès de l'Association des bibliothécaires français (ABF) entendait cette année s'interroger sur les conséquences pour les bibliothèques publiques et leurs usagers de l'"invasion du droit". À côté de nombreux ateliers consacrés à divers autres aspects de la profession, une série de tables rondes a abordé cette question complexe par différentes approches : droit de prêt, droit d'auteur, libertés et interdits, situation dans les archives et les musées, regards sur le Danemark et la République tchèque, droits et expression des usagers.

(J-M. R.)

- *Libraries : a voyage of discovery; linking the future to the past* (IFLA 2005, congrès mondial des bibliothèques et de l'information, Oslo, 14-18 août 2005) – Mireille LAMOUREUX et Michèle BATTISTI – (p. 297-309) – (21 réf.).

Le 71^e congrès de la Fédération internationale des associations de bibliothécaires et des bibliothèques (IFLA) s'est tenu cette année à Oslo. Les questions abordées au cours des séances plénières présentées ici ont permis d'approfondir la réflexion des participants sur les liens entre développement, pluralisme culturel, solidarité et démocratie, savoir et pouvoir. La dimension internationale a notamment été évoquée : mondialisation, coopération Nord-Sud. Parmi les nombreuses sessions spécialisées organisées simultanément, ce compte rendu se penche sur celles qui étaient consacrées aux bibliothèques et à la liberté d'expression, aux problèmes juridiques de la profession, à la gestion des bibliothèques, aux services offerts aux usagers, à la formation des professionnels et du public, et à de nouveaux outils pour les associations du secteur de l'information et des bibliothèques.

(J-M. R.)

- *Politiques documentaires et gestion des coûts* (Journée d'étude Poldoc, Paris, 20 septembre 2005) – Michèle BATTISTI – (p. 310-313) – (3 réf.).

La quatrième journée d'étude organisée à Paris par Poldoc, Groupe de recherche bibliéconomique appliquée aux outils de politique documentaire, a abordé la dimension financière de la stratégie des bibliothèques. Des approches transversales ont mis en relief l'impact des coûts sur la stratégie documentaire et celui des budgets d'acquisition sur le budget d'un établissement. Une étude sur la position des acquéreurs face à la baisse de leur pouvoir d'achat a été présentée, ainsi que des outils : les instruments de mutualisation que sont les consortiums et des outils de gestion développés par des bibliothèques.

(J-M. R.)

EDUCACIÓN Y BIBLIOTHECA

Vol. 17, n°150 (noviembre-diciembre 2005)

Ce 150^e numéro passe en revue, avec un commentaire approprié, tous ceux qui ont été édités depuis le lancement de la revue en 1989.

(J.H.)

EDUCACIÓN Y BIBLIOTHECA

Vol. 17, n°151 (enero-febrero 2006)

À signaler pour son originalité l'article - signé par Francisco SOLANO, (p. 39-46) - intitulé *Memorias de editores - Los hijos de Gutenberg con los mercaderes del templo*. (Mémoires d'éditeurs - les fils de Gutenberg avec les marchands du temple). Cela pourrait se traduire par "du livre élément de culture, au livre simple produit commercial".

(J.H.)

LECTURES

Vol. 25, n°143 (novembre-décembre 2005)

- *L'affaire Google Print* – Philippe ALLARD – (p. 15-16).

En quelques lignes, l'auteur fait le point sur l'arrivée de Google dans le domaine de la numérisation d'ouvrages. Il fait état des réactions suscitées dans le monde par ce projet.

(J.H.)

- *Evaluer une bibliothèque* – Divers auteurs - (p. 17-36).

Dans ce dossier centré sur la refonte du *Rapport d'activités annuel des bibliothèques* proposée par la Communauté française, les différents auteurs - tous actifs dans le domaine de la bibliothèque - nous donnent leur point de vue sur cette question particulièrement actuelle.

(J.H.)

LECTURES

Vol. 25, n°144 (janvier-février 2006)

- *Le format PDF* – Philippe ALLARD – (p. 17-18).

Le format PDF est connu de tous et partant fort utilisé depuis l'apparition de sa première version en 1992. En quoi il consiste, ce qu'il permet de faire, comment le créer, le convertir, l'utiliser pour l'archivage, autant de points succinctement examinés par l'auteur.

(J.H.)

INFORMATION –

WISSENSCHAFT & PRAXIS

Vol. 56, nr 5-6 (August-September 2005)

- *Webkataloge: eine alte, neue Alternative für Wirtschaft und Wissenschaft zur Informationsbeschaffung im WWW* – Jürgen SCHAMBERGER – (p. 295-302) – (26 ref.).

Le Web, et Internet de manière plus générale, sont devenus les médias les plus couramment utilisés par les entreprises et leurs collaborateurs pour accomplir leurs tâches quotidiennes.

Plusieurs études démontrent cependant les nombreux échecs dans l'obtention d'informations pertinentes sur Internet. Avec à l'origine un usage prédominant de moteurs de recherche à orientation commerciale, l'inéadéquation des résultats de recherche conduit à des coûts élevés en termes de temps de production et de compétitivité pour les entreprises.

Une solution possible se dégage par l'utilisation d'annuaires Web spécialisés. En plus de ces outils de recherche "redécouverts", il existe au sein des entreprises un besoin réel de nouvelles manières de qualifier et d'éduquer leur personnel en vue d'une utilisation plus efficace de la toile mondiale.

(C.B.)

- *Informationspathologien im Kontext informationsintensiver Dienstleistungen* – Stefan ZILLICH – (p. 309-312) – (4 ref.).

Pour les entreprises dont la principale activité est le traitement et la fourniture d'information, l'information "matière première" constitue tant un service qu'un contenu. Dans cette industrie liée à l'information, des problèmes spécifiques peuvent apparaître qui empêchent la manipulation efficace de l'information, et qui peuvent conduire à des "pathologies informationnelles". L'article aborde les volets organisationnel et psychologique des processus des flux d'information, développant un modèle adapté aux fournisseurs de services d'information, p. ex. les organismes de consultance ou de services financiers.

(C.B.)

INFORMATION –

WISSENSCHAFT & PRAXIS

Vol. 56, nr 7 (Oktober 2005)

- *Patentinformation in einem erweiterten Europa - Konzentriere Dich auf Deinen Kunden!* – Dieter GEIB – (p. 365-370).

L'article dresse un compte-rendu du congrès PATLIB organisé par l'Office des Brevets européen et qui s'est déroulé du 17 au 19 mai 2005 à Sinaia (Roumanie), avec comme mot d'ordre "Focus on Your Customer".

(C.B.)

- *Management von E-Journalen und Onlinedatenbanken an der Universität Graz - Anforderungen und Lösungen* – Helmut HARTMANN, Christian GÜNTHER – (p. 373-375).

Dans le cadre d'une bibliothèque universitaire, les auteurs décrivent, après identification des besoins, la mise en place d'une solution destinée à la gestion des ressources électroniques (périodiques électroniques et bases de données). Ils dressent le bilan de trois ans d'utilisation du produit "HAN" (Hidden Automatic Navigator).

(C.B.)

- *Persönliche Literaturverwaltung im Umbruch - Vom Bibliographie-Management zum Social Bookmarking* – Hans-Christoph HOBHOM – (p. 385-388) – (5 ref.; 7 logiciels).

L'auteur établit un tour d'horizon des différents logiciels disponibles pour la gestion de bibliographies, les "Bibliographic formatting softwares". Sept logiciels sont présentés :

Reference Manager, EndNote, Procite, RefViz,
WriteNote, RefWorks et Connotea.

(C.B.)

**INFORMATION –
WISSENSCHAFT & PRAXIS**
Vol. 56, nr 8 (November-Dezember 2005)

- *Zielgruppen für barrierefreies Internet* -
Beate SCHULTE – (p. 405-412) – (24 ref.).

L'entièreté de ce numéro est consacrée à la question de l'accessibilité de l'Internet aux personnes handicapées. Les auteurs des divers articles appréhendent ce sujet particulièrement important sous plusieurs aspects : social, légal, celui du discours.

Ils analysent aussi les protocoles utilisés, la qualité des sites proposés, le recours aux documents en PDF, une technique de tests simples d'accessibilité, la conception de sites Web adaptés, la réalisation d'une offre d'accès libre et les exigences de ladite accessibilité

(J.H.)

**INFORMATION –
WISSENSCHAFT & PRAXIS**
Vol. 57 nr 1 (Januar-Februar 2006)

- *Einsatz und Erfolg von Competitive Intelligence (CI) im deutschen Mittelstand* -
Ragna SEIDLER-DE ALWIS – (p. 7-14) – (41 + 40 ref.).

Étude relative à l'application, l'utilisation et la réalisation de l'intelligence économique (IE) dans les moyennes entreprises allemandes. L'article présente une évaluation de l'application de l'IE et son développement au sein d'entreprises et donne une analyse de l'implémentation et de la mise en oeuvre de l'IE par le staff de la société par rapport au recours à des fournisseurs extérieurs. Les résultats sont basés sur des recherches dans la littérature, dans les journaux, dans les bases de données en ligne, dans des sources trouvées sur Internet tout comme à la suite d'interviews téléphoniques. On met en avant les avantages et les désavantages de l'utilisation de l'IE, de même qu'on traite de son externalisation. La réalisation pratique dans les moyennes entreprises allemandes est analysée avec prise en considération des voies et moyens de la fourniture d'information et des responsabilités dans les entreprises considérées.

(J.H.)

- *Fachangestellte/r für Medien- und Informationsdienste. Ein Anspruchsvoller Serviceberuf* – Barbara MÜLLER-HEIDEN – (p. 19-22) – (10 ref.).

Analyse des répercussions (favorables) causées par le décret de 1998 sur la formation professionnelle des employés de services liés à l'information. Les exigences de cette formation et les développements qui en découlent sont détaillés ainsi que les avantages qu'elle présente.

(J.H.)

- *Stand und Perspektiven von Informationswissenschaft und -praxis* – Stefanie HÖHFELD und Sonja WEBER – (p. 27-29).

Une communication au sujet de "l'état et des perspectives de la Science de l'information" faite à la Heinrich-Heine-University Düsseldorf présente les méthodes et approches scientifiques et pratiques de cette science.

Des enseignants ainsi que des experts de l'industrie de l'information en Allemagne font un tour d'horizon sur la recherche actuelle et sur les applications réalisées dans ce domaine.

(J.H.)

- *Dimensionen der Zeitschriftenzientometrie am Beispiel von "Buch und Bibliothek"* – Kerstin JUCHEM, Christian SCHLÖGL und Wolfgang G. STOCK – (p. 31-37) – (12 ref.).

Les dimensions de la scientométrie des revues sont :

1. la production de la revue (articles, auteurs),
2. le contenu (thèmes),
3. la réception (lecteurs),
4. la communication formelle (références, citations),
5. la rédaction (politique éditoriale, publication).

La revue "BuB" (Forum für Bibliothek und Information) – qui a la plus grande circulation en Allemagne est analysée scientométriquement. La période considérée s'étend de 1990 à 2003. Contrairement à ce qui était attendu au regard des lois de l'informétrie "BuB" offre de remarquables caractéristiques spécifiques. Il n'y a pas de Top-auteurs mais un groupe homogène de ceux-ci domine. Les références citent en grande majorité la revue elle-même. En fait cette dernière est comme une île où peu d'informations étrangères viennent s'échouer. Si la production d'articles se partage exactement entre les auteurs féminins et masculins, il y a des différences spécifiques de genre. Les premiers écrivent des articles courts et des rapports de conférence, les autres par contre

produisent des articles longs et des analyses de la littérature.

(J.H.)

- *Digitales Archiv der TU Berlin : Übergang auf OPUS und URN* – Rüdiger SCHNEEMAM – (p. 41-43).

Les publications digitales ne contrôlent pas encore l'exploitation scientifique, mais elles en sont un important composant. Il est de la responsabilité des bibliothèques universitaires de les archiver de manière conviviale et facile à retrouver et aussi de les offrir à la communauté scientifique via des moteurs de recherche adaptés. On décrit les plus importantes normes qui doivent être respectées pour ce type d'archives. On explique aussi pourquoi le traitement des repositories est fait par les bibliothèques. Enfin dans un domaine plus limité, on motive le choix du système OPUS par la TU Berlin avec ses perspectives.

(J.H.)

- *Informationsfachleute sollen Berater werden* – Vera MÜNCH – (p. 45-50).

Dans son compte-rendu du congrès *Online Information 2005* (Londres, 29 novembre au 1 décembre) l'auteur explique pourquoi les personnes impliquées dans le domaine de l'information - principalement celle accessible par l'internet - se transforment en "Conseillers en information". Les possibilités d'un tel accès ne cessent de croître par la mise à disposition de tous d'endroits de connexion à Internet. La multiplicité des sources, la variété des moteurs de recherche et surtout l'évolution constante des techniques relatives à l'information telles :

les Wikis, le Podcasting, les Blogs, et tant d'autres risquent de faire perdre pied à l'utilisateur, sans compter encore sur la mise au point de nouveaux appareils de plus en plus performants. Un des rôles majeurs des spécialistes du domaine est donc d'encadrer les dits usagers afin de leur permettre d'obtenir des réponses pertinentes aux recherches qu'ils exécutent.

(J.H.)

- *Patentrecht, Patentrecherche und Patentanalyse : Mittel auf dem Wege zur Patentaufwertung* – Dieter GEIß – (p. 51-57).

L'auteur, spécialiste allemand dans le domaine de la documentation relative aux brevets, nous donne ici un compte-rendu fouillé de ce qui s'est dit au cours du 27^e Colloque sur la question tenu à l'Université d'Illmenau les 2 et 3 juin 2005. Il fait état des principales communications proposées sur le sujet et des diverses questions que celui-ci soulève.

(J.H.)

- *Mit Usability-tests den Geschäftserfolg von Web-Portalen verbessern* – Anke HOFFMAN - (p. 59-60).

Habituellement les portails du Web ne prennent pas suffisamment en considération les besoins réels des utilisateurs. C'est dans ce sens que des tests d'usabilité peuvent venir à leur secours. Ces différents tests sont passés en revue et on montre sur quoi ils peuvent déboucher.

(J.H.)

ARTICLES - ARTIKELS

2005/1

- Ce n'est pas qu'un texte d'humeur (1^e partie) 3-5
Pierre LASZLO
- De bedreiging van ons papieren erfgoed: massaontzuring als remedie 6-18
Laurent MEESE
- Les archives électroniques : un défi pour l'avenir ? 19-21
Véronique FILLIEUX, Cathy SCHOUKENS et Évelyne VANDEVOORDE
- Apprendre à utiliser et à intégrer l'Internet dans les pratiques d'information et de communication des médecins : l'exemple du Cameroun 22-31
Samuel TIETSE

DOSSIER

L'intelligence économique - Het economische intelligentie

- De la veille à l'intelligence économique 32-34
Marc BORRY
- L'intelligence économique en pratique : de l'analyse stratégique à la veille concurrentielle 35-37
Julie ABRAHAM
- L'attracteur informationnel 38-46
Alain TIHON
- De l'heuristique à l'intelligence économique en passant par la bibliométrie 47-50
Michel DORBAN

2005/2

- Ce n'est pas qu'un texte d'humeur (2^e partie) 5-8
Pierre LASZLO
- Bouwen aan de bibliotheek van Babel 9-18
Dirk VAN EYLEN
- Construire la bibliothèque de Babel 19-28
Dirk VAN EYLEN

- L'accès des sites Internet aux personnes déficientes visuelles 29-34
Tamara BEECKMANS
- Open Linking et OpenURL 35-47
Philippe MOTTET

2005/3

NUMÉRO SPÉCIAL
SPECIAAL NUMMER

Stratégie de préservation et de conservation pour les archives et les collections pour les
bibliothèques, analogiques et numériques
Een preserveerings- en conserveringsstrategie voor analoge en digitale bibliotheekcollecties
en archiefbestanden

- Het verval van papieren collecties 7-35
Guy DE WITTE
- Archiveren van websites: een kwestie van waardering en "capture" 36-53
Filip BOUDREZ
- A digital preservation handbook: best practice and its dissemination 54-59
Neil BEAGRIE
- Een conserveringsbeleid voor geluidsarchief met toepassingen voor het
stadsarchief Antwerpen 60-67
Liesbeth BAATEN

2005/4

- "Lier ou ne pas lier" : pour un usage responsable de l'hyperlien 4-12
Michèle BATTISTI
- Wikipédia : une source d'information extraordinaire ou une pseudo-
encyclopédie peu fiable 13-22
Guy DELSAUT
- Le coût de l'accès aux connaissances scientifiques : la solution du libre
accès 23-28
Annick CASTIAUX
- L'Union Européenne : équation à 25 inconnues 29-36
Maxime WOTQUENNE

AUTEURS

ABRAHAM, Julie	1/35	BEECKMANS, Tamara	2/29
BAATEN, Liesbeth	3/60	BORRY, Marc	1/32
BATTISTI, Michèle	4/4	BOUDREZ, Filip	3/36
BEAGRIE, Neil	3/84	CASTIAUX, Annick	4/23

DE WITTE, Guy	3/7	SCHOUKENS, Cathy	1/19
DELSAUT, Guy	4/13	TIETSE, Samuel	1/22
DORBAN, Michel	1/47	TIHON, Alain	1/38
FILLIEUX, Véronique	1/19	VANDEVOORDE, Évelyne	1/19
LASZLO, Pierre	1/3 & 2/5	VAN EYLEN, Dirk	2/9 & 2/19
MEESE, Laurent	1/6	WOTQUENNE, Maxime	4/29
MOTTET, Philippe	2/35		

Instructions aux auteurs

Le comité de rédaction des *Cahiers de la Documentation* accepte, en vue de les publier, des **articles originaux**, traitant de tout sujet appartenant au domaine des sciences de l'information et de la documentation. Les articles seront écrits dans l'une des trois langues nationales ou éventuellement en anglais. Le comité ouvrira également les pages de la revue aux textes des exposés présentés dans le cadre des manifestations de l'Association Belge de Documentation.

Le comité peut toutefois décider d'accueillir dans ses colonnes des textes déjà publiés ailleurs, soit de sa propre initiative, soit à la demande de l'auteur. Il s'assurera, dans ce cas, que les autorisations de reproduire ont été obtenues des auteurs et des ayant droits.

Le contenu de l'article doit être rédigé avec un outil de traitement de texte **compatible avec le standard Microsoft Word** et envoyé de préférence sous forme d'un fichier attaché "fichier.rtf". Si un auteur se trouve dans l'impossibilité de répondre à ce critère, il devra d'abord en avvertir le comité de rédaction et choisir avec lui le mode de transmission.

Afin de respecter une unité de présentation, les auteurs voudront bien se conformer aux critères de présentation repris dans le document publié sur notre site à l'adresse : <<http://www.abd-bvd.be/cah/instructions.pdf>>.

Les auteurs devront faire parvenir les versions électroniques finales de leur article à l'adresse électronique suivante : <cahiers-bladen@abd-bvd.net>.

Sauf exception à discuter avec le comité de rédaction, la revue demande de lui réserver l'exclusivité de l'article : un même article ne peut être proposé simultanément à différentes revues.

Les auteurs respectent les délais auxquels ils se sont engagés vis-à-vis de la rédaction.

Les auteurs signent un document qui règle la situation des droits d'auteur pour une utilisation future des articles, y compris la mise à disposition du public sur le site de l'ABD. Un formulaire standard leur est soumis. Toute dérogation doit être clairement indiquée dans le document signé.

Les références bibliographiques seront présentées selon la méthode des citations numériques, c'est-à-dire numérotées en chiffres arabes entre parenthèses dans le texte et reprises, classées par ordre croissant des numéros d'appel dans la liste bibliographique publiée à la fin de l'article.

Les références respecteront pour chaque type de document les règles minimales reprises dans la norme ISO 690-1987 <<http://www.collectionscanada.ca/iso/tc46sc9/standard/690-1f.htm#9.2>>.

- **Périodiques** : Liste des auteurs, incluant au moins les initiales de leurs prénoms, suivie du titre de l'article, du nom du périodique (sans abréviation) en italique, de l'année de publication, du numéro du volume, du numéro dans le volume et du numéro des pages.

Exemple : Albrechtsen, Hanne ; Jacob, Elin K. The dynamics of classification systems as boundary objects for cooperation in the electronic library. *Library Trends*, 1998, Vol. 47, n° 2, pp. 293-312.

- **Ouvrages** : Liste des auteurs, incluant au moins les initiales de leurs prénoms, suivie du titre de l'ouvrage en italique, du lieu de la publication, du nom de la société éditrice et de l'année de publication.

Exemple : Juanals, Brigitte. *La culture de l'information, du livre au numérique*. Paris : Hermès, 2003, 230 p.

- **Liens (URL)** : Les références à des sites Web doivent reprendre leur adresse URL entre les signes <>. Elles seront suivies de la mention : consulté le « date », entre parenthèses, la date étant celle de la dernière consultation du site et éventuellement par la mention « consultable uniquement sur abonnement » si cela s'avère être le cas.

Exemple : <<http://www.ala.org/ala/acrl/acrlstandards/standards.pdf>> (consulté le 11 juin 2005)

Pour les documents plus spécifiques, à savoir les extraits d'ouvrages, les documents de littérature grise, les actes de colloques..., les auteurs sont priés de consulter le document complet publié sur le site de l'ABD.

Auteursaanbevelingen

Het publicatiecomité van de *Bladen voor Documentatie* aanvaardt **originele artikels** die in brede zin informatie of documentatie behandelen, om te publiceren in het tijdschrift. Alleen artikels, geschreven in een van de officiële landstalen of in het Engels, komen in aanmerking voor publicatie. Het comité zal eveneens uiteenzettingen van presentaties, georganiseerd in het kader van evenementen van de Belgische Vereniging voor Documentatie, opnemen in zijn tijdschrift.

Het comité kan altijd beslissen om reeds elders gepubliceerde teksten in dezelfde of in gewijzigde vorm op te nemen. In dit geval wordt voor de nodige autorisaties gezorgd.

Het ingediende artikel dient **Microsoft Word compatibel** te zijn en bij voorkeur als "rtf-bijlage" overgemaakt. Indien dit onmogelijk blijkt, zal de auteur hierover vooraf overleggen met het publicatiecomité.

Om uniformiteit in de ingediende teksten te bekomen, dienen de auteurs zich te houden aan de criteria zoals deze zijn opgenomen in het document dat zich bevindt op onze website op het volgende adres:
<<http://www.abd-bvd.be/cah/auteursaanbevelingen.pdf>>.

De auteurs verzenden de finale versie van hun artikel elektronisch naar: <cahiers-bladen@abd-bvd.net>.

Behalve indien anders afgesproken met het publicatiecomité, vraagt het tijdschrift de exclusiviteit van het artikel te waarborgen: eenzelfde artikel kan niet gelijktijdig in verschillende tijdschriften voorgesteld worden.

De auteurs respecteren de tijdstippen waartoe zij zich verbonden hebben ten aanzien van het publicatiecomité.

De auteurs tekenen een document waarbij het auteursrecht voor een verder gebruik van het artikel geregeld wordt en een akkoord voor een vermelding op de BVD website. Een standaard formulier wordt hen hiervoor aangeboden. Iedere afwijking dient duidelijk in, het ondertekend document opgenomen te zijn.

De bibliografische referenties worden aangeduid volgens de methode van digitale referenties, het is te zeggen in Arabische cijfers beginnend met het cijfer één tussen haakjes in de tekst en olopend hernomen en uitgelegd aan het einde van het artikel.

De referenties voor elk type van document houden rekening met de minimale omschrijving uit de ISO-norm 690-1987: <<http://www.collectionscanada.ca/iso/tc46sc9/standard/690-1f.htm>>.

- **Artikels:** Lijst van auteurs, met inbegrip minstens van de initialen van hun voornamen, gevolgd door de titel van het artikel, de titel van het tijdschrift (zonder afkorting) in cursief, het jaar van publicatie, het volume, het tijdschriftnummer en de paginavermelding

Voorbeeld: Meese, Laurent. De bedreiging van ons papieren erfgoed : massaontzuring als remedie? *Cahiers de la documentation = Bladen voor documentatie*, 2005, nr. 1, pp. 6-18.

- **Monografieën:** Lijst van auteurs, met inbegrip minstens van de initialen van hun voornamen, gevolgd door de titel van het boek in cursief, de plaats van uitgifte, de uitgever(s), het jaar van uitgifte en het aantal pagina's

Voorbeeld: Lienardy, Anne ; Van Damme, Philippe. *Inter folia : Handboek voor de conservatie en de restauratie van papier*. Brussel : Koninklijk Instituut voor het Kunstpatrimonium (KIK-IRPA), 1989, 248 p.

- **Linken (URL):** De referenties van de websites dienen het volledige adres tussen scherpe haakjes te vermelden, gevolgd door de vermelding van datum van raadpleging tussen haakjes, de datum is deze van de meest recente raadpleging, eventueel met vermelding dat de site alleen toegankelijk is voor abonnees, indien dit het geval zou zijn.

Voorbeeld: <<http://www.ala.org/ala/acrl/acrlstandards/standards.pdf>> (bezoekt op 25 januari 2005)

Voor meer specifieke documenten zoals uittreksels, grijze literatuur, verslagen en colloquia... worden de auteurs gevraagd het volledige document, gepubliceerd op de BVD website, te raadplegen.