

Classification à facettes et modèles à base de points de vue : Différences et complémentarité

Sabine Mas

sabine.mas@utt.fr

Aurélien Bénel

aurelien.benel@utt.fr

Jean-Pierre Cahier

jean-pierre.cahier@utt.fr

Manuel Zacklad

manuel.zacklad@utt.fr

Laboratoire ICD/Tech-CICO

Université de Technologie de Troyes, France

Résumé

La modélisation multi-points de vue telle qu'utilisée en ingénierie des connaissances est parfois confondue avec la classification multi-facettes telle que comprise en bibliothéconomie et sciences de l'information. Cet article tente de les distinguer tant au niveau théorique que sur des exemples pratiques. Il décrit ensuite leur utilisation conjointe et leur complémentarité dans la gestion de contenu d'une grande entreprise.

Abstract

Because of their similarities, faceted classification (in Information sciences) and multi-viewpoints modeling (in Knowledge engineering) are often confused. This paper draws their differences both theoretically and on examples. Then it describes their joint use and their complementarity for content management in a large corporation.

1. Facettes et points de vue : Enjeux

Après des décennies axées sur la gestion des données puis celle des documents numérisés, les entreprises doivent aujourd'hui faire face à la production massive de documents nativement numériques. Le *record management* et la gestion des contenus

deviennent ainsi des enjeux cruciaux. Pourtant, malgré les avancées technologiques récentes, plusieurs problèmes majeurs en gestion documentaire subsistent et gravitent principalement autour des moyens utilisés pour décrire et organiser les documents dans le but de mieux les repérer, les partager, les exploiter et ultimement les conserver ou les détruire.

Les moyens utilisés par les entreprises pour gérer leurs documents reposent largement sur la mise à disposition de moteurs de recherche, de langages documentaires et d'instruments de gestion tels que les schémas de classification, les thésaurus et les calendriers de conservation (Sabourin, 2001; Van der Walt, 2004). Malheureusement beaucoup de ces outils, qui ont été conçus pour faciliter des activités de gestion documentaire spécifiques (par exemple, recherche, classification, indexation, gestion du cycle de vie des documents), sont le plus souvent incompatibles, redondants, inefficaces ou insatisfaisants pour l'organisation et la recherche quasi quotidienne de documents dans un environnement électronique. En l'absence d'un système d'organisation systématique et intégré des documents, il devient donc de plus en plus difficile pour les grandes entreprises de maîtriser leur patrimoine documentaire et de se conformer aux lois portant sur l'accès à l'information et sur la transparence du processus décisionnel, d'assurer la défense de leurs droits et la conservation de la trace de leurs activités (Lévesque, 1998-1999; Skupsky, 1999; Guercio, 2002; Association des archivistes français, 2004).

Pour surmonter ces difficultés plusieurs courants de recherche visent à proposer des solutions diverses. Dans le domaine des sciences de l'information on explore, par exemple, la possibilité de remplacer les systèmes de classification traditionnellement hiérarchiques par une infrastructure à facettes permettant d'indexer et de naviguer à travers une collection de ressources documentaires référencées sur le Web (Zins, 2002; Hudon, 2007) ou plus modestement dans l'espace personnel d'un individu (Henderson, 2005) ou d'une organisation (Bibliothèque et Archives Canada, 2007).

Parallèlement, pour faire face aux limitations d'une classification basée sur un hypothétique consensus, de nombreux efforts issus du domaine de l'ingénierie des connaissances ont cherché à rendre compte des "connaissances" d'un domaine selon différents points de vue. En particulier, l'approche du Web socio-sémantique (Cahier, 2005; Zacklad *et al.*, 2007) propose de recourir à des outils d'organisation des documents et des connaissances qui permettent de caractériser les éléments d'une collection à l'aide

de thèmes répartis dans une pluralité de points de vue complémentaires ou concurrents, favorisant ainsi une construction des connaissances, de manière non pas consensuelle mais dialectique (Iacovella, Béné et al., 2006). Autrement dit, alors que les facettes permettent à une même communauté d'organiser et de repérer les documents suivant différentes dimensions préalablement sélectionnées, les points de vue, quant à eux, relèvent d'acteurs qui appartiennent à des communautés ou à des métiers différents (Cahier et al., 2004).

Le présent texte fait état d'une réflexion théorique et méthodologique sur la divergence et la complémentarité possible de ces deux approches pour optimiser l'organisation, le repérage et l'exploitation documentaire. Nous présenterons d'abord les deux approches à partir de projets de recherche réalisés par les auteurs de ce texte dans le domaine des sciences de l'information et de l'ingénierie des connaissances. Nous discuterons ensuite des enjeux méthodologiques et de la possibilité de combiner l'approche par point de vue à une classification à facettes, en vue d'offrir une organisation documentaire autour non seulement du savoir de référence de la communauté mais aussi des expertises des différents acteurs.

2. Illustration des deux types de classification

2.1 Classification à facettes des documents administratifs électroniques

La classification documentaire désigne l'organisation des documents en classes sur la base de caractéristiques ou critères de division communs (Maniez, 1987; Hunter, 2000). Le premier critère de division est un des éléments les plus importants puisque c'est lui qui décrit et organise le fonds documentaire en classes logiques (Kwasnik, 1999). Cependant, même les documents les plus simples ont plusieurs caractéristiques (par exemple, le type de document ou le thème représenté) qui peuvent être utilisées pour faire des regroupements de documents. Ainsi, les possibilités de regroupements peuvent être infinies et les mêmes documents peuvent être classifiés de différentes manières (Langridge, 1992; Mai, 2000).

Depuis plusieurs décennies, les employés au sein des entreprises utilisent une interface hiérarchique pour organiser les documents administratifs électroniques se trouvant sur leur poste de travail informatique. Cette interface hiérarchique possède des avantages cognitifs pour l'individu puisqu'elle lui permet de repérer un document en privilégiant la navigation

locale à travers une structure arborescente de répertoires et la localisation spatiale du dossier dans lequel se trouve classé le document (Barreau et Nardi, 1995; Berchtold, Biliris et Panagos, 1999).

Les limites inhérentes à une telle structure hiérarchique, déjà soulignées par plusieurs chercheurs (Barreau, 1995; Dourish et al., 2000), obligent les individus à privilégier comme base de classification et selon des critères très personnels, tantôt le thème du document, tantôt le type de document ou encore l'activité à laquelle il se rattache, alors que le document pourrait de fait être classifié à plusieurs endroits (Henderson, 2005).

Les résultats de l'analyse de 21 schémas de classification hiérarchiques créés et utilisés par des employés de l'Université de Montréal (Mas, 2007) révèlent de nombreuses différences sur le plan logique entre les schémas de classification personnels et les schémas de classification institutionnels. Alors que les schémas de classification personnels permettent d'organiser plus spécifiquement les documents selon leur contenu thématique ou leur forme, le schéma de classification institutionnel vise à organiser les documents selon les fonctions et activités administratives et opérationnelles dans le cadre desquelles ils ont été créés ou reçus. Ce faisant, le schéma de classification institutionnel offre des éléments de contexte qui facilitent le repérage et la compréhension des documents pour l'ensemble du personnel de l'organisme (Conférence des recteurs et des principaux des universités du Québec, 1994; Héon, 1999; Direction des archives de France, 2002; Sprehe, McClure et Zellner, 2002). Le décalage observé entre les exigences institutionnelles et les besoins individuels permet de mieux comprendre la réticence de certains employés à vouloir utiliser la classification institutionnelle basée sur une hiérarchie des fonctions et activités pour l'organisation des documents administratifs électroniques qui sont sous leur contrôle direct (Gagnon-Arguin et Bannouri, 1998).

L'individualisme qui caractérise la classification des documents administratifs électroniques a conduit les archivistes et gestionnaires de documents à revoir leurs méthodes et à proposer des solutions en vue d'uniformiser l'organisation des documents électroniques à caractère institutionnel et prévenir les problèmes ultérieurs de repérage des documents. L'impact positif des approches de classification à facettes pour l'organisation et le repérage de l'information a été confirmé dans un environnement réseauté (Gödert, 1991; Ingwersen et Wormell, 1992; Maisonneuve et Touitou, 2007) et nous est apparu prometteur pour l'organisation des documents adminis-

tratifs électroniques (Hudon et Mas, 2001). Les études et expériences présentement réalisées au sein de plusieurs organisations publiques canadiennes (Bibliothèque et Archives Canada, 2007) semblent indiquer que la souplesse, l'expressivité et la simplicité d'un modèle à facettes (Vickery, 1963), mettant en relation le contenu de l'objet informationnel (plus proche des besoins individuels) et le contexte dans lequel il a été créé (incontournable pour les besoins de gestion documentaire), présenteraient plusieurs avantages : Elles permettraient de surmonter les difficultés auxquelles se heurtent les tentatives de mise en œuvre des schémas de classification institutionnels hiérarchiques et unidimensionnels, d'améliorer la précision de la recherche et de mieux gérer l'information corporative à travers son cycle de vie. Cela favoriserait aussi une plus grande automatisation du processus de classement et d'organisation de l'information dans le but de délester l'utilisateur d'une tâche qu'il accomplit rarement très bien (Gagnon-Arguin et Bannouri, 1998; Mas, 2007).

La notion de « facettes » a souvent été présentée comme l'apport théorique le plus important en sciences de l'information (Maniez, 1999) et a inspiré de nombreuses études dans les dernières décennies. La classification à facettes se présente comme une combinaison de groupes conceptuels complémentaires plutôt que comme une longue liste structurée de sujets. À la différence d'un simple mode hiérarchique et unidimensionnel d'organisation des documents, l'efficacité des classifications à facettes tient à leur capacité à intégrer différentes dimensions d'analyse sur les objets informationnels, à caractériser et à rendre l'accès à l'information plus facile en offrant de multiples voies de navigation vers n'importe quel document.

Selon la théorie des facettes à la base de la *Colon Classification* (Ranganathan, 1965), le contenu de tout document touche à cinq aspects de la réalité : la personnalité (groupe, catégorie de personnes, thème ou objet duquel il est question), la matière (substance, propriété, qualité), l'énergie (opération principale, action que l'on décrit en rapport avec l'objet), l'espace (localisation du sujet dans l'espace) et le temps (place du sujet dans le temps). Ranganathan voit dans l'utilisation de ces perspectives ou facettes une syntaxe idéale pour l'expression et la représentation d'un sujet.

L'ambition d'une formule universelle de description des sujets qui serait applicable à tous les domaines (par exemple, santé, métallurgie) apparaît utopique pour répondre aux exigences d'une classification moderne (Vickery, 1963; Spiteri, 1998). Maniez

(1999) rapporte que pour le Classification Research Group (CRG), qui réunit des théoriciens et des praticiens anglais des classifications, le modèle à facettes est d'autant mieux adapté à une indexation de qualité que le domaine d'application est plus étroit.

Dans le cadre de ce courant de recherche, nous avons réalisé une étude exploratoire pour étudier la possibilité de remplacer la structure classificatoire hiérarchique des documents électroniques d'un coordonnateur de stages de l'Université de Montréal par un schéma de classification à facettes qui devait représenter à la fois le contexte de production et d'utilisation des documents mais aussi le sujet véhiculé par les documents (Mas, Zaher et Zacklad, 2008). Dans le cadre de cette étude, nous nous sommes inspirés du « modèle simplifié » d'analyse de facettes de Spiteri (1998), une combinaison des théories de Ranganathan et du CRG, soit une application d'une approche d'analyse à la fois descendante (*top-down*) et ascendante (*bottom-up*) commune à la conception de tout schéma de classification documentaire.

D'une part, il s'agit de subdiviser le domaine conceptuel à structurer selon les règles de division logique, indépendamment du contenu du fonds documentaire. L'approche déductive (*top-down*) sous-jacente à cette structuration logique du domaine assure la cohérence et les possibilités d'évolution du schéma de classification (Maniez, 1987; Fallis et Mathiesen, 2000).

D'autre part, la garantie d'existence documentaire (*literary warrant*) consiste à tenir compte du nombre de documents offerts sur chaque sujet pour s'assurer d'élaborer un schéma de classification spécifique à l'ensemble documentaire (Langridge, 1992; Cannonne, 1993; Iyer, 1995). Autrement dit, dans un domaine précis, la liste des facettes ne peut être fournie *a priori* dans une formule immuable, elle doit plutôt être recherchée de façon pragmatique par un examen des documents relevant du domaine, l'analyse conceptuelle permettant de ramener le nombre théoriquement illimité des termes d'un domaine à un nombre restreint de catégories fondamentales (Vickery, 1963).

Une analyse des activités formelles et informelles de l'employé participant à l'étude (i.e. un coordonnateur de stages) fut réalisée comme étape préalable à l'identification et à la sélection des facettes pertinentes pour répondre à la fois aux exigences institutionnelles (associer les documents aux activités dans le cadre desquelles ils sont créés ou reçus) et aux besoins particuliers de cet employé (représenter le sujet du document). Par exemple, les coordonnateurs de

stages à l'Université de Montréal peuvent être amenés à participer à différentes activités de nature administrative, professionnelle ou pédagogique reliées à leur fonction. Ils sont non seulement responsables des activités relatives à la coordination générale des stages (par exemple, l'organisation de rencontres avec les étudiants, la planification des stages, l'assignation des étudiants en stage et parfois les visites dans les milieux), ils jouent également le rôle d'intermédiaire entre les milieux de stage (par exemple, établissements de santé), les employeurs (organismes publics ou privés) et les étudiants (inscrits ou non inscrits en stage). Les coordonnateurs de stages peuvent être appelés à participer à la formation des partenaires qui supervisent les étudiants en stage, à la formation des étudiants qui ont de la difficulté en stage, à l'élaboration et à la mise à jour d'outils pédagogiques (par exemple, grilles d'évaluation et guides de l'étudiant). Enfin, les coordonnateurs de stages peuvent être impliqués dans la planification d'événements spéciaux (par exemple, la Journée de l'emploi) ou invités à participer à diverses activités de prospection ou de relations publiques (par exemple, le Salon des études).

La nature des tâches à réaliser suggère qu'un employé peut avoir besoin d'accéder plus fréquemment et plus rapidement à certaines catégories spécifiques de documents dans le cadre de ses activités. Le coordonnateur de stages interrogé a confirmé avoir un rapport quasi quotidien avec certains types de documents. Il est amené à gérer et de manière récurrente (i.e. d'année en année ou par session universitaire) beaucoup de documentation et de correspondance, d'offres de stages, d'offres d'emploi, de banques de données ainsi que de nombreux rapports produits ou reçus. Par ailleurs, il travaille de manière récurrente avec des modèles ou types de documents génériques (par exemple, des formulaires ou des étiquettes), de la documentation de référence sur différents thèmes (par exemple, sur le plagiat) et est appelé à gérer des documents plus spécifiques (par exemple, procès-verbaux, rapports d'étudiants).

L'analyse des concepts représentés dans la structure arborescente des répertoires et de la terminologie utilisée pour désigner explicitement ces concepts (Voir Figure 1) a permis de valider la pertinence et l'exhaustivité des facettes identifiées. L'observation d'un usage répandu de classes qui représentent plusieurs concepts (par exemple, "Dossiers étudiants A-2003") allait dans le sens d'un transfert souhaitable dans ce domaine d'une classification hiérarchique unidimensionnelle vers une classification à facettes.

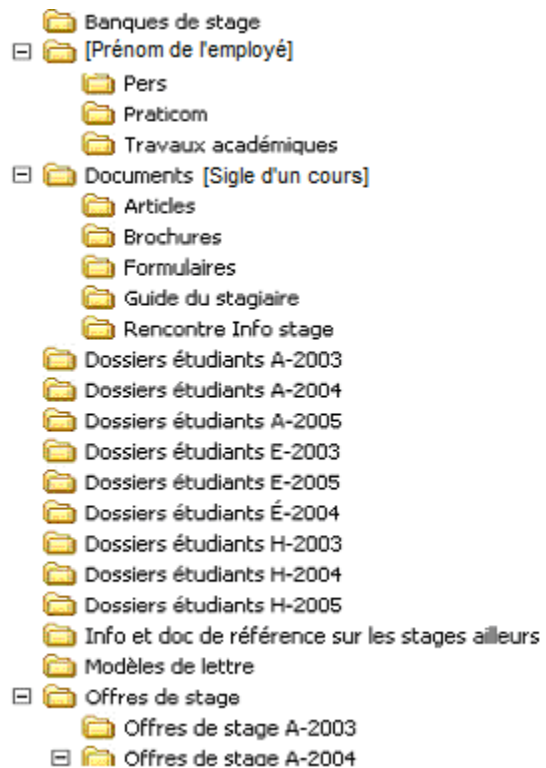


Figure 1. Extrait de la structure classificatoire arborescente du coordonnateur de stage participant à l'étude.

Suite à l'analyse des activités documentarisées et des concepts utilisés par l'employé pour classifier ses documents, un modèle classificatoire a été développé en vue d'offrir à l'employé la possibilité de classifier et de repérer ses documents selon six facettes :

1. Activité
2. Type de document
3. Session universitaire
4. Provenance du document
5. Destinataire du document
6. Statut de l'étudiant

Une première utilisation et évaluation de ce prototype par l'employé lui-même a donné des résultats forts encourageants qui ont été rapportés dans Mas, Zaher et Zacklad (2008). Une évaluation de la maquette est prévue auprès d'un plus grand nombre de coordonnateurs de stages de l'Université de Montréal.

En résumé, la conception d'une classification à facettes appliquée à l'organisation décentralisée des

documents administratifs électroniques met en œuvre un raisonnement relevant de la formalisation référentielle (dans un domaine donné) (Bachimont, 2004). Cette conception est basée sur une analyse qui inclut tour à tour des méthodes descendantes et ascendantes, effectuée par un expert en gestion de l'information et des documents, en vue d'identifier et de sélectionner les facettes parmi un ensemble de dimensions conventionnelles ou mesurables : auteur du document, destinataire du document, type de document, références administratives des dates, des lieux, des dénominations de personnes, etc. Le but d'une telle démarche est de proposer un langage classificatoire post-coordonné logique, référentiel et stable à travers le temps, comportant un nombre limité de facettes prédéfinies, et permettant une classification à la fois multidimensionnelle mais aussi consensuelle et standardisée de l'ensemble des documents au sein d'une même entreprise par des acteurs appartenant à un même métier.

On peut toutefois s'interroger sur la reconnaissance de cette manière unilatérale (conçue par un seul expert) et formelle de décrire et d'organiser les documents complexes, et destinée à des catégories d'employés poursuivant des objectifs similaires mais ayant une sémantique singulière (l'univers du discours de l'archiviste-classificationniste ne coïncide pas toujours avec les auteurs du document ni avec les utilisateurs), des besoins informationnels différents, des connaissances et des références pouvant être très spécifiques et évolutifs. Parmi les multiples informations véhiculées de façon explicite ou implicite, structurée ou non structurée, dans les documents, on ne peut complètement prédire lesquelles vont constituer des critères de référencement ou de recherche déterminants pour les différents employés (par exemple, durée du stage mentionné dans le rapport, coût du stage pour l'étudiant, sujet du stage, nombre et nature des intervenants impliqués, etc.). Les concepteurs de schémas de classification à facettes pourraient tirer profit à accorder, dans un cadre plus décentralisé, une place croissante au point de vue des employés. La participation des utilisateurs à la définition des catégories de classification et des thèmes utilisés pour classer n'est certes pas nouvelle en soi. L'aspect novateur de l'approche multi-points de vue que nous nous proposons de présenter dans la section suivante réside dans les modalités collaboratives et participatives par lesquelles les individus peuvent jouer un rôle actif dans le référencement des documents qu'ils créent, reçoivent et utilisent dans le cadre de leurs activités.

2.2 Classification multi-points de vue visant à partager un corpus de ressources pédagogiques

Dans le cadre de l'exploration d'un système de classification documentaire qui se doit d'être à la fois assez général et cohérent pour répondre aux besoins de l'entreprise et assez précis et dynamique pour répondre aux besoins des employés, les auteurs de ce texte travaillent activement sur les apports potentiels du Web « socio-sémantique » (Cahier *et al.*, 2004), très proche de la notion de « Pragmatic Web » (Schoop, De Moore et Dietz, 2006), un sous-domaine potentiellement prometteur et fécond du Web sémantique.

Les applications du Web socio-sémantique permettent aux communautés de co-produire et d'utiliser des artefacts organisationnels symboliques, de type « cartes », « nuages de mots » ou « index partagés » pour rendre les activités et connaissances collectives à la fois plus visibles et interactives.

L'hypothèse du Web socio-sémantique est de considérer que la sémantique des notions échangées sur un réseau (intranet et internet), celle qui sera la plus utile et la plus efficace, sera celle qui sera co-construite et définie à l'intersection de plusieurs points de vue portés par une chaîne d'acteurs, de rôles et de métiers différents dans des situations pragmatiques où tous ces acteurs sont porteurs de différentes interprétations (Cahier, 2005). L'idée est d'encourager le débat autour de documents décrits par des « points de vue » construits par des personnes ou des équipes autonomes (Iacovella, Bénéel *et al.*, 2006). Dans ce cadre conceptuel le degré de formalisation sémantique des concepts dépend directement du degré d'accord entre les participants à la communauté. Le succès récent des outils inscrits sous la bannière marketing du Web 2.0, qui relèvent largement des principes du Web socio-sémantique, semble en confirmer la justesse.

Pour opérationnaliser le concept de Web socio-sémantique, le laboratoire Tech-CICO de l'Université de technologie de Troyes en France (Zacklad *et al.*, 2007) a spécifié le modèle et le protocole « Hypertopic » permettant de matérialiser la diversité des points de vue en présence et l'évolution rapide de la sédimentation sémantique des notions échangées. Pour ce faire, ils ont été inspirés par le standard ISO de représentation des connaissances « Topic Map ». Les spécifications d'Hypertopic ont été opérationnalisées à l'aide de XML et ont servi de base pour le développement de plusieurs architectures Open Source.

Hypertopic est directement conçu pour faciliter des activités d'enquête (Dewey, 1967) réalisées par les membres d'une communauté ou d'une organisation impliqués dans des tâches récurrentes et pour partie commune, médiatisées par le Web socio-sémantique. Dès que ces tâches ont un certain niveau de complexité, de récurrence et de durabilité, il est important de mettre en œuvre des stratégies de gestion des connaissances (Dieng-Kuntz *et al.*, 2001; Zacklad et Grundstein, 2001; Ermine, 2003), facilitant la transférabilité du connaissant (*knowing*) et du connu (*known*) vers d'autres acteurs.

Hypertopic est conçu comme un hypertexte autour des *items*. Cet item est la charnière d'une triple représentation avec pour conséquence un usage croisé plus efficace des ressources documentaires, des données informatisées et du Web. En effet, dans l'approche Hypertopic, la caractérisation documentaire par des ressources, la spécification référentielle par des attributs standards (par exemple, nom de l'auteur ou date de la dernière modification si l'item est un document) et la thématisation heuristique (*thèmes* indexant l'item) se conjuguent au service de formes hybrides de recherche ouverte d'informations centrées sur cet item.

Dans le cas du projet Cogdoc, site expérimental où un réseau d'enseignants distants mutualisent des supports de cours sur le thème du document au sens large destinés aux étudiants, le modèle Hypertopic est utilisé pour des caractérisations ou thématisations concurrentes de l'item « séance de cours ». Ces « points de vue » sont proposés par les enseignants en sciences de l'information et en ingénierie des connaissances, mais aussi reprennent les thématisations des catégories officielles des catalogues d'université (les intitulés du cours et du programme d'enseignement) et celles qui émanent des élèves cherchant des modules de formation.

La figure 2 montre l'application CogDoc réalisée selon les principes du Web socio-sémantique avec le client Agoræ et le serveur Argos de l'infrastructure Hypertopic. Cette application permet de naviguer dans les ressources documentaires selon un hypertexte structuré hiérarchisé multi-points de vue qui facilite la recherche ouverte d'information au sein des corpus. En assistant la navigation, le protocole Hypertopic permet de réaliser des associations entre thèmes, ressources documentaires et attributs standardisés, dont on peut considérer qu'elles relèvent d'inférences pragmatiques et non plus formelles. Les points de vue sont notamment ceux des cursus offerts par les différents établissements d'enseignement, ici celui de l'Université de Techno-

logie de Troyes. Comme on le voit dans l'interface ci-dessous, l'objectif est de corrélérer une description hiérarchique des items selon différents points de vue (i.e., point de vue enseignant, étudiant, université) avec une description folksonomique. Les « Tags » permettent aux utilisateurs de proposer d'autres sous-thèmes selon une méthode « ascendante ». Ceux-ci peuvent ensuite être intégrés aux ontologies sémiotiques (Zacklad *et al.*, 2007) associées aux points de vue.

The figure consists of two screenshots of the Hypertopic Agoræ client interface. Both screenshots show a navigation bar at the top with 'Espace personnel', 'Personnel', 'Enseignants', and 'Universités'. The top screenshot displays a page for 'Cycle d'ingénieur : Systèmes d'information et télécommunications' with a 'Nuage de mots' (word cloud) on the right containing terms like 'numérique', 'document', 'électronique', 'Dewey', 'Document', 'Ergonomie et interfaces homme-machine', 'Groupware et workflow', 'HTTP', 'Ingénierie des connaissances', 'Modélisation des connaissances', and 'connaissances'. The bottom screenshot displays a page for 'Histoire d'une bibliothèque numérique : l'ABU' with a 'Fiche descriptive' section showing 'auteur : Pierre Cubaud' and 'dernière modification : 2004', and a 'Ressource(s) documentaire(s)' section showing a document thumbnail titled 'A.B.U.' with a description: 'Un projet collaboratif - copies numériques - copies liées de modifications'. A 'Nuage de mots' on the right of the bottom screenshot contains 'Web sémantique, document numérique et interfaces avancées'.

Figure 2. Le client Hypertopic Agoræ : exploration selon le point de vue des cursus de l'Université de Technologie de Troyes (a) *point de vue* (un établissement) (b) *thème* (un cursus) (b') *sous-thèmes* (intitulés des cours), (c) *item* (un cours), (d) *attribut standard/valeur* (auteur, dernière modification...) (e) *ressource documentaire* (ici des transparents). Il est à remarquer que cet écran propose aussi un nuage folksonomique de « tags » et propose à l'utilisateur d'annoter l'item courant par un nouveau tag.

L'approche constructiviste et collaborative proposée comporte le risque d'être confrontée à une classification incohérente et non consensuelle. L'usage des points de vue devient, selon l'approche du Web socio-sémantique, un moyen important pour contenir le désordre et le relativisme que certains redoutent comme conséquence de l'ouverture de la formalité sémiotique (Zacklad *et al.*, 2007). Assez fréquemment, dans une situation donnée, sans pour autant

être en accord sur tout, les acteurs finissent par s'accorder sur une sorte de schéma de classification général perfectible impliquant un nombre raisonnable de points de vue. Sans nier les éventuelles contradictions, ce « schéma pluriel » permettrait de rendre plus claire et plus efficace pour certains usages, la structure sémantique nécessaire à l'activité dans cette situation (Cahier, 2005, p.6).

3 Vers une approche hybride

3.1 Illustration : organisation de documents techniques au sein d'une grande entreprise

À EDF (Électricité de France), nombre des activités effectuées sur un site de production nucléaire (diagnostic, maintenance, suivi) demandent de disposer sur le terrain des informations pertinentes. Ces informations sont accessibles sous forme de données et de rapports produits depuis plus de cinquante ans dans un contexte hautement réglementé.

Par exemple, une recherche pourtant limitée à la base de documents de Recherche & Développement, renvoie déjà une liste de plusieurs milliers de documents sur un sujet comme le diagnostic des transformateurs de puissance. Anciens rapports d'activité, rapports techniques récents et documents de conception sont renvoyés indifféremment à l'employé qui pourtant, en fonction de ses besoins, leur accorde une valeur bien différente.

L'objectif de notre projet avec EDF est de déployer une infrastructure hybride bénéficiant de la complémentarité entre multi-facettes et multi-points de vue (Voir Tableau 1).

Pour la classification selon les exigences managériales, le projet s'appuiera sur une approche multi-facettes des systèmes d'organisation des connaissances combinant une description cohérente, conventionnelle et standardisée du contenu des documents (par exemple, Lieu, Date, Entités) et du contexte de leur création (par exemple, Fonction, Activité, Métier, Rôle).

Ces facettes institutionnelles seront complétées par une méthode d'organisation ascendante inscrite dans l'approche du Web socio-sémantique. Celle-ci permettra d'offrir d'une part, une vision plus subjective et orientée métier des documents, et d'autre part, les moyens de les organiser de manière itérative par les différentes communautés qui les lisent, les analysent et les réutilisent. Cette deuxième approche sera préférée lorsque la complexité des contenus documentaires ne permet pas d'utiliser des bibliothèques de taxinomies prédéfinies, lorsque

celles-ci restent à concevoir ou encore quand que le métier de l'organisation étudiée est trop spécifique.

	Classification multi-facettes	Classification multi-points de vue
Analyse	Descendante et ascendante	Ascendante
Modalité de conception	Externe (mono-expert)	Participative (folksonomique) et collaborative (multi-experts)
Validation	Objective	Inter-subjective
Longévité	Pérenne	Évolutive
Perspective	Managériale	Métier

Tableau 1. Principales caractéristiques méthodologiques et propriétés des deux approches

Le questionnement que soulève ce projet est actuellement le suivant : les deux approches se renforcent-elles ou se concurrencent-elles ? Comment les usagers interpréteront-ils les différentes facettes et points de vue en présence ?

3.2 Enjeux méthodologiques et perspectives

Nous avons mentionné que la classification à facettes qui s'appuie sur une sémantique qui se veut consensuelle et stable ne nous semble pouvoir exprimer qu'une partie des contenus sémiotiques véhiculés par les documents. La représentation de certaines dimensions documentaires conventionnelles ou observables est pertinente mais loin d'être suffisante dès lors que les transactions et utilisations ont une certaine complexité; complexité qui correspond à des situations cognitivement ou socialement conflictuelles dans lesquelles la terminologie est en constante évolution. Dans ces situations, il nous paraît nécessaire de compléter l'objectivité et la pérennité d'une classification à facettes par une approche plus subjective et évolutive par points de vue.

Cette complémentarité théorique reste toutefois à être mise en œuvre. Outre la difficile intégration de ces deux dimensions dans une même interface documentaire, nous devons également régler la question de l'articulation de deux processus de conception extrêmement différents. En effet, les classifications à facettes sont conçues lors de la phase initiale d'un projet par des professionnels en sciences de l'information s'appuyant à la fois sur des bonnes pratiques reconnues (facettes dites de contenu) et sur des analyses, représentations et décompositions des processus métiers et administratifs de l'organisation (facettes dites de contexte). Les points de vue, au contraire sont co-construits par les employés eux-mêmes, c'est-à-dire sans passer par les profession-

nels de l'information qui jouent alors plutôt un rôle d'animation, de consolidation et de gouvernance.

4. Conclusion

La classification à facettes, applicable dans certains domaines de connaissances, restreints, stables et aisément formalisables, ne nous semble pas suffisante pour décrire et organiser l'ensemble des ressources documentaires liées plus spécifiquement aux processus métiers d'une entreprise, à la fois très évolutives et porteuses de contenus aussi riches que diversifiés. Nous postulons qu'une classification multi-points de vue et les méthodes préconisées dans le cadre du Web socio-sémantique, basées sur des approches de type Web 2.0, permettraient d'améliorer progressivement la description formelle des ressources documentaires à partir des annotations déposées par les utilisateurs des documents. Elles contribueraient également aux problématiques d'agrégation de contenus et de connaissances. Ce postulat reste à étayer par des expérimentations dans de prochains travaux.

L'aspect le plus novateur d'une classification multi-facettes associée à une classification multi-points de vue pour l'organisation et l'exploitation du patrimoine documentaire d'une entreprise est à situer au plan méthodologique. Il s'agirait de combiner : (1) des méthodes inspirés de la théorie des facettes, de l'archivistique et du *record management* permettant de remplir à la fois des objectifs de gestion à long terme de la documentation sous un angle institutionnel, avec une analyse fine et précise des contenus documentaires dans une perspective administrative et opérationnelle, et (2) des méthodes de conception et de gestion des évolutions de certaines facettes de contenu, basées sur une approche d'ingénierie socio-cognitive ascendante, permettant de mettre à contribution les producteurs et les utilisateurs de la base dans la définition et la maintenance des systèmes d'organisation des connaissances. Une méthodologie d'instanciation de ce modèle hybride dans un contexte d'entreprise est en cours de développement.

Références et bibliographie

Association des archivistes français. 2004. *Abrégé d'archivistique*. Paris : Association des archivistes français.

Bachimont, B. 2004. *Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Mémoire d'Habilitation à

Diriger des Recherches. Université de Technologie de Compiègne.

- Barreau, D. K. 1995. Context as a factor in personal information management systems. *Journal of the American Society for Information Science* 46, no. 5: 327-39.
- Barreau, D. K., et B. A. Nardi. 1995. Finding and reminding : File organization from the desktop. *SIGCHI Bulletin* 27, no. 3: 39-43.
- Berchtold, S., Alexandros B., et E. Panagos. 1999. SaveMe : A system for archiving electronic documents using messaging groupware. *Proceedings of the international joint conference on Work activities coordination and collaboration*, San Francisco, California, United States, 167-176.
- Bibliothèque et Archives Canada. 2007. *Examen sommaire SCAOS des petits organismes*. <http://www.gestioninformation.gc.ca/docs/bascs-scaos/bascs-scaostb_f.asp>.
- Cahier, J.P. 2005. Ontologies sémiotiques pour le Web socio-sémantique : étude de la gestion coopérative des connaissances avec des cartes hypertextuelles. Thèse de doctorat. Troyes : Université de Technologie de Troyes. <<http://cahier.tech-cico.fr/publi/cahier-phd.pdf>>
- Cahier, J-P., Zacklad, M. et Monceaux, A. 2004. Une application du Web socio-sémantique à la définition d'un annuaire métier en ingénierie. In *Actes du colloque Ingénierie des Connaissances, Lyon, France, mai 2004*. <<http://cat.inist.fr/?aModele=afficheN&cpsidt=16602714>>
- Canonne, A. 1993. *Vocabulaire élémentaire des classifications*. Liège : Éditions du CÉFAL.
- Conférence des recteurs et des principaux des universités du Québec (CRÉPUQ). 1994. *La gestion des archives informatiques*. Sainte-Foy, Québec : Presses de l'Université du Québec.
- Dewey, J. 1967. *Logique, la théorie de l'enquête*. Presses Universitaires de France.
- Dieng-Kuntz, R., O. Corby, F. Gandon, A. Giboin, J. Golebiowska, et N. Matta. 2001. *Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire du Knowledge Management*. Dunod.
- Direction des archives de France. 2002. *Les archives électroniques : manuel pratique*. Dhérent, C. et collaborateurs. Direction des archives de France.
- Dourish, P., Edwards, W.K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., Terry, D.B. et J. Thornton. 2000. Extending document management systems with user-specific active properties.

- ACM Transaction on Information Systems 18, no. 2: 140-170.
- Ermine, J.-L. 2003. *La gestion des connaissances*. Paris : Hermes science.
- Fallis, D. et K. Mathiesen. 2000. Consistency rules for classification schemes (or how to organize your beanie babies). In *International ISKO Conference : Dynamism and stability in knowledge organization proceedings of the sixth International ISKO Conference, 10-13 July 2000*, Toronto, Canada, 339-344. Wurzberg : Ergon Verlag.
- Gagnon-Arguin, L., et R. Bannouri. 1998. La classification des documents institutionnels dans les ministères du gouvernement du Québec et le défi informatique : rapport de recherche. Montréal : École de bibliothéconomie et des sciences de l'information. <<http://www.tresor.gouv.qc.ca/doc/acrobat/ingenierie9.pdf>>
- Gödert, W. 1991. Facet classification in online retrieval. *International Classification*, no. 2: 98-109.
- Guercio, M. 2002. Records classification and content management : Old functions and new requirements in the legislations and standards for electronic record-keeping systems. In *Proceedings of the DLM Forum 2002. Access and preservation of electronic information : Best practices and solutions, May 6-8, 2002, Barcelona, Spain*. Luxembourg : Office for Official Publications of the European Communities, 432-442.
- Henderson, S. 2005. Genre, task, topic and time : Facets of personal digital document management. In *Proceedings of the 6th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction, Auckland, New Zealand*, 75-82. <<http://doi.acm.org/10.1145/1073943.1073957>>
- Héon, G. 1999. La classification. In Couture, C. et collaborateurs. *Les fonctions de l'archivistique contemporaine*. Sainte-Foy, Québec : Presses de l'Université du Québec, 219-254.
- Hudon, M. 2007. Proposal for an alternate structure to organize and access resources in a virtual library in education. In *Actes du 35ème congrès de l'Association canadienne des sciences de l'information, Université McGill, Montréal, Canada, 10-12 mai 2007*. <http://www.cais-acs.ca/proceedings/2007/hudon_2007.pdf>
- Hudon, M., et S. Mas. 2001. *Analyse des facettes pour la classification des documents institutionnels au gouvernement du Québec*. Rapport présenté pour le Groupe de travail en classification et indexation. Montréal : École de bibliothéconomie et des sciences de l'information, Université de Montréal. (Collection en ingénierie documentaire; 13). <http://www.msg.gouv.qc.ca/fr/publications/enligne/administration/ingenierie/classification_analyse.pdf>
- Hunter, E. J. 2000. Do we still need classification? In Marcella, R., et A. Maltby (eds.). *The future of classification*. London : Gower, 1-17.
- Iacovella A., Bénel A., Pétard X., et B. Helly. 2006. Corpus scientifiques numérisés : Savoirs de référence et points de vue des experts. In R. T. Pédaudque (Ed.). *La redocumentarisation du monde*. Toulouse : Cepaduès. <http://benel.tech-cico.fr/publi/iacovella_PEDAUQUE_06.pdf>
- Ingwersen, P. et I. Wormell. 1992. Ranganathan in the perspective of advanced information retrieval. *Libri*, no. 42: 184-201.
- Iyer, H. 1995. *Classificatory structures concepts, relations and representation*. Frankfurt/Main : Indeks Verlag.
- Kwasnik, B. H. 1999. The role of classification in knowledge representation and discovery. *Library Trends* 48, no. 1, 22-47.
- Langridge, D. W. 1992. *Classification : Its kinds, elements, systems, and applications*. London : Bowker-Saur.
- Lévesque, M. 1998-1999. Les documents électroniques et le calendrier de conservation, ou les considérations d'un pauvre archiviste qui essaie de suivre l'évolution technologique. *Archives (Revue de l'Association des archivistes du Québec)* 30, no. 3-4: 39-54.
- Mai, J.-E. 2000. Likeness : A pragmatic approach. In *Proceedings of the 6th International ISKO Conference, July 10-13, 2000, Toronto, Canada*. Würzburg, Germany : Ergon, 23-27.
- Maisonneuve, M., et C. Touitou. 2007. Une nouvelle famille d'OPAC : Navigation à facettes et nuages de mots. *Bulletin des bibliothèques de France* 52, no. 6: 12-19.
- Maniez, J. 1987. *Les langages documentaires et classificatoires : conception, construction et utilisation dans les systèmes documentaires*. Paris : Éditions d'Organisation.
- Maniez, J. 1999. Des classifications aux thésaurus : du bon usage des facettes. *Documentaliste - Sciences de l'information* 36, no. 4-5: 249-62.
- Mas, S. 2007. *Schémas de classification et repérage des documents administratifs électroniques dans un contexte de gestion décentralisée des ressources informationnelles*. Thèse de doctorat.

- Montréal : Université de Montréal.
<<http://hdl.handle.net/1866/1433>>
- Mas, S., Zaher, H. et M. Zacklad. 2008. Design and evaluation of multi-viewed knowledge system for administrative electronic document organization. In *10th biennial International Conference of the International Society for Knowledge Organization (ISKO) : Culture and Identity in Knowledge Organization*, 138-144.
- Ranganathan, S. R. 1965. *The Colon classification*. New Brunswick, N.J : Rutgers University Press.
- Sabourin, P. 2001. Constructing a function-based records classification system : Business activity structure classification system. *Archivaria* 51: 137-154.
- Schoop, M., De Moore, A. et J. L. G. Dietz. 2006. The pragmatic web : A manifesto. *Communication of the ACM* 49, no. 5: 75-76.
- Skupsky, D. S. 1999. Applying records retention to electronic records. *The Information Management Journal* 33, no. 3: 28-35.
- Spiteri, L. 1998. A simplified model for facet analysis. *Canadian Journal of Information and Library Science* 23: 1-30.
- Sprehe, T. J., McClure, C.R. et P. Zellner. 2002. The role of situational factors in managing U.S. federal recordkeeping. *Government Information Quarterly* 19: 289-305.
- Van der Walt, M. S. 2004. A classification scheme for the organization of electronic documents in small, medium and micro enterprises (SMMEs). *Knowledge Organization* 31, no. 1: 26-38.
- Vickery, B. C. 1963. *La classification à facettes : guide pour la construction et l'utilisation de schémas spéciaux rédigé*. Documentation et Information. Paris : Gauthier-Villars.
- Zacklad, M. et M. Grundstein. 2001. *Management des connaissances*. Stanmore : Hermes Science Europe LTD.
- Zacklad, M., Bénel, A., Cahier, J-P., Zaher, L'H., Lejeune, et C., Zhou. 2007. Hypertopic : une méta-sémiotique et un protocole pour le Web socio-sémantique. In *Actes des 18eme journées francophones d'ingénierie des connaissances (IC2007)*. Francky Trichet (Eds.). Cépaduès, 217-228. <http://publi.techcico.fr/files/zacklad_IC_07.pdf >
- Zins, C. 2002. Models for classifying Internet resources. *Knowledge Organisation* 29, no. 1: 20-28.