

ÉTUDE

*Problématique de l'organisation et du repérage des documents administratifs électroniques dans un contexte de gestion décentralisée des ressources informationnelles**

Sabine Mas
Louise Gagnon-Arguin

Depuis plusieurs décennies, les employés utilisent une interface hiérarchique pour interagir avec les documents administratifs électroniques se trouvant sur leur poste de travail informatique personnel. 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 (Berlin *et al*, 1993; Barreau et Nardi 1995; Berchtold, Biliris et Panagos 1999).

Pour répondre, entre autres, aux besoins d'uniformité dans la gestion des documents, de contextualisation de l'information et de regroupement des documents, les archivistes préconisent l'utilisation d'un schéma de classification institutionnel, déjà conçu pour l'organisation des documents sur support papier, pour nommer et organiser

* Cet article rend compte des résultats d'une thèse réalisée dans le cadre d'un doctorat en sciences de l'information à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal sous la direction de Louise Gagnon-Arguin et la codirection de Michèle Hudon. La thèse complète est disponible en ligne à l'adresse <http://hdl.handle.net/1866/1433> sous le titre *Schémas de classification et repérage des documents administratifs électroniques dans un contexte de gestion décentralisée des ressources informationnelles*.

les répertoires électroniques. Les avantages liés à l'utilisation du schéma de classification institutionnel dans un contexte de gestion décentralisée sont nombreux et ont déjà été signalés. Le schéma de classification institutionnel offre une structure de classification hiérarchique, uniforme et formelle qui permet d'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; Gagnon-Arguin et Bannouri 1998; Héon 1999; Senécal 1999; Dhérent et collaborateurs 2002; Sprehe, McClure et Zellner 2002).

Or, dans la pratique courante, le schéma de classification institutionnel est rarement utilisé par les employés qui mettent eux-mêmes en place leur propre structure classificatoire pour organiser les documents qui se trouvent sur leur poste de travail informatique personnel. C'est l'individualisme qui prévaut le plus souvent dans un tel contexte, encouragé par différents facteurs tels l'absence de leadership, de ressources ou de motivations institutionnelles, la quasi-absence de politiques, normes, méthodes et procédures de gestion des documents électroniques, l'absence de formation et d'encadrement des employés, la croyance que les documents électroniques ne sont pas des documents « officiels », des documents d'archives.

Alors que les créateurs et utilisateurs premiers des documents semblent s'y retrouver (Chapman 1999; Boardman et Sasse 2004), l'utilisation de ces stratégies personnelles d'organisation des documents électroniques nous a amenés à nous poser les questions suivantes : Quelles sont les caractéristiques des structures classificatoires conçues et utilisées par les employés pour organiser leurs documents électroniques? Qu'en est-il de la capacité d'une tierce personne à repérer les documents électroniques organisés selon ces schémas personnels? Existe-t-il des différences sur le plan de la capacité à repérer les documents selon les caractéristiques des schémas de classification utilisés? Si oui, quelles sont les caractéristiques des schémas de classification qui influencent l'efficacité du repérage? Les recherches portant sur les schémas de classification utilisés par des catégories particulières d'employés sont peu nombreuses. Aucune étude empirique n'a été menée à ce jour afin de vérifier dans quelle mesure les schémas de classification personnels permettent ou même facilitent le repérage des documents électroniques par une tierce personne.

Notre recherche avait deux objectifs spécifiques. Le premier objectif consistait à décrire les caractéristiques de différents schémas de classification utilisés par des employés pour organiser et classer les documents administratifs électroniques qui sont sous leur contrôle direct. Le deuxième objectif consistait à vérifier, dans un environnement contrôlé, les différences sur le plan de l'efficacité du repérage de documents électroniques qui sont fonction du schéma de classification utilisé. La réalisation de ces deux objectifs visait à augmenter nos connaissances sur les schémas de classification personnels en contexte de travail et à mesurer empiriquement la capacité de tierces personnes à comprendre et à repérer des documents à l'aide de schémas de classification et de navigation dotés de caractéristiques variées. Ce faisant, nous souhaitions identifier les caractéristiques d'un schéma de classification qui permettent d'augmenter la probabilité de repérer avec succès un document

électronique, tout en minimisant le temps requis pour compléter la tâche ainsi que le risque d'erreur.

Dans cet article, nous résumons d'abord l'état de la littérature relatif à l'organisation personnelle des documents administratifs électroniques et aux principaux facteurs d'influence. Nous exposons ensuite quelques éléments théoriques sur la relation entre le schéma de classification documentaire et le repérage. En terminant, nous présentons les principaux résultats de l'analyse d'un échantillon de schémas de classification personnels et de la simulation du repérage de documents électroniques réalisé à l'aide de cinq schémas de classification présentant des caractéristiques variées.

L'ORGANISATION PERSONNELLE DES DOCUMENTS ADMINISTRATIFS ÉLECTRONIQUES ET FACTEURS D'INFLUENCE

L'organisation personnelle des documents administratifs électroniques est définie comme étant tout processus mis en place par un individu pour faciliter l'organisation et le repérage des documents qui sont sous son contrôle direct en vue de répondre à ses besoins personnels (Lin et Chan 1999; Bergman, Beyth-Marom et Nachmias 2003; Boardman et Sasse 2004). Le concept d'organisation «personnelle» ou «décentralisée» des documents administratifs électroniques est donc distinct de celui d'organisation «institutionnelle» ou «centralisée» des documents administratifs électroniques. Dans un contexte de gestion centralisée, c'est un professionnel en gestion documentaire qui conçoit et contrôle un système d'organisation des documents pour une variété d'individus ayant des besoins différents à l'intérieur d'un même organisme ou d'une même unité (Bergman, Beyth-Marom et Nachmias 2003).

L'organisation personnelle des documents électroniques en milieu de travail est effectuée par des millions d'individus, plusieurs fois par jour (Whittaker, Terveen et Nardi 2000). Elle est fréquemment considérée comme un fardeau (Malone 1983; Lansdale 1988; Whittaker et Sidner 1996) et une source de problèmes et de frustrations (Boardman et Sasse 2004). Si certains individus qui ont eu recours au schéma de classification des documents administratifs pour organiser leurs documents sur support électronique sont satisfaits des résultats, d'autres semblent éprouver des difficultés et préfèrent recourir à des stratégies personnelles d'organisation des documents plus adaptées à leurs besoins (Gagnon-Arguin et Bannouri 1998).

L'outil communément utilisé par les individus pour organiser les documents administratifs électroniques qui sont sous leur contrôle direct possède des appellations variées : «système hiérarchique de fichiers» ou «structure organisationnelle» (Dumais *et al.* 2003), «hiérarchie» ou «hiérarchie sémantique» (Abrams, Baecker et Chignell 1998) «système de classement personnel» (Lansdale 1988), «taxonomie personnelle» (Jones, 2004) ou «schéma de classification personnel» (Jones et Maier 2003; Ferraioli 2005) (Voir Figure 1).

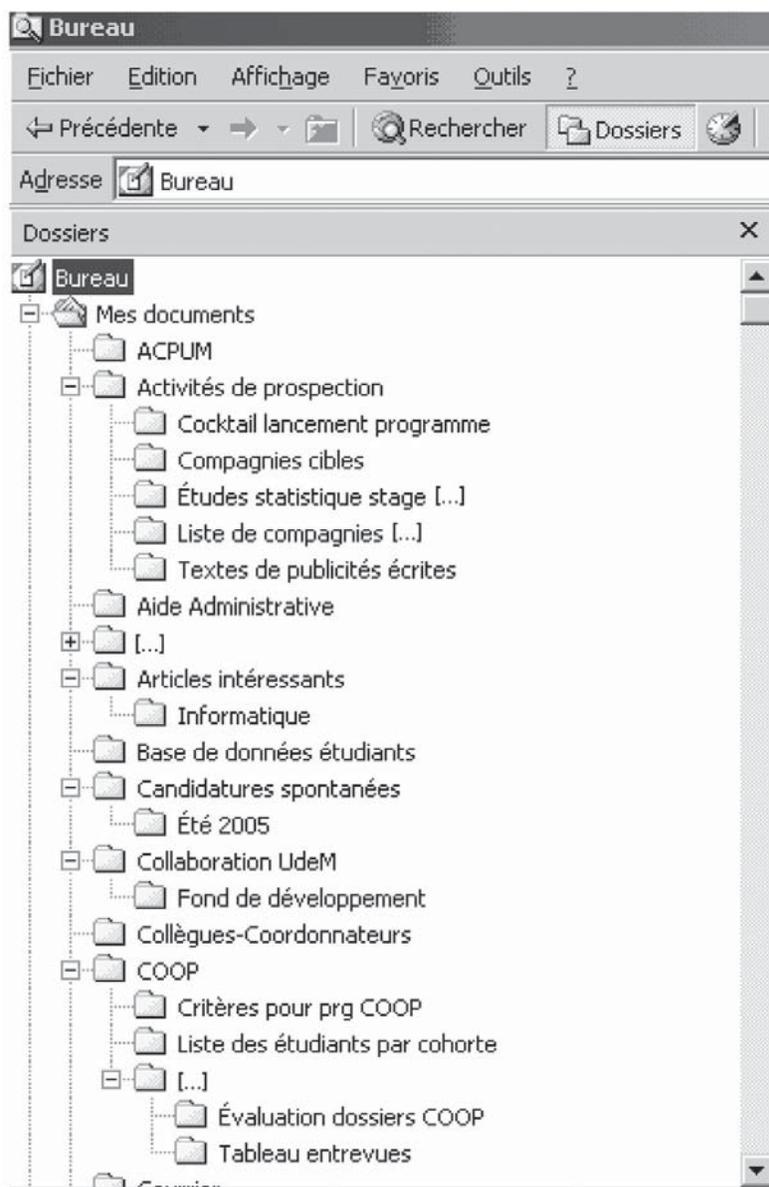


Figure 1. Exemple de schéma de classification personnel (Les caractères [...] ont été ajoutés par les auteures pour masquer des noms de répertoires confidentiels ou nominatifs)

Un schéma de classification personnel des documents administratifs désigne un arrangement hiérarchique de catégories conceptuelles (ou répertoires électroniques) permettant le regroupement de documents selon des règles d'application souvent implicites, relatives par exemple à des projets, à des activités ou à des types de documents (Boardman et Sasse 2004). L'organisation de ces catégories conceptuelles est qualifiée de «personnelle» car elle se base sur la représentation du monde que se fait une personne et vise à répondre à ses propres besoins (Jones et Maier 2003; Jones 2004).

À partir d'une revue de la littérature dans divers domaines reliés à l'organisation des documents électroniques (par exemple, l'archivistique, les sciences de l'information, la psychologie cognitive, l'interaction Homme-Machine et la communication), il est

possible d'identifier trois ensembles de facteurs interreliés entre eux qui influencent le niveau de formalisation d'un schéma de classification personnel des documents. Il s'agit 1) de facteurs humains (les employés), 2) de facteurs documentaires (l'espace documentaire à organiser), et 3) de facteurs institutionnels (Voir Figure 2).

L'influence de ces facteurs n'a pas été statistiquement démontrée. Leur identification est cependant utile car elle rend possible la catégorisation des schémas de classification mis en place par les individus, allant de schémas de classification peu structurés (*piles*) à des schémas de classification relativement plus structurés (*files*) (Henderson 2003). Dans le cadre de notre recherche, nous avons tenu compte de ces facteurs de contingence ou d'influence pour l'échantillonnage et l'interprétation des résultats.

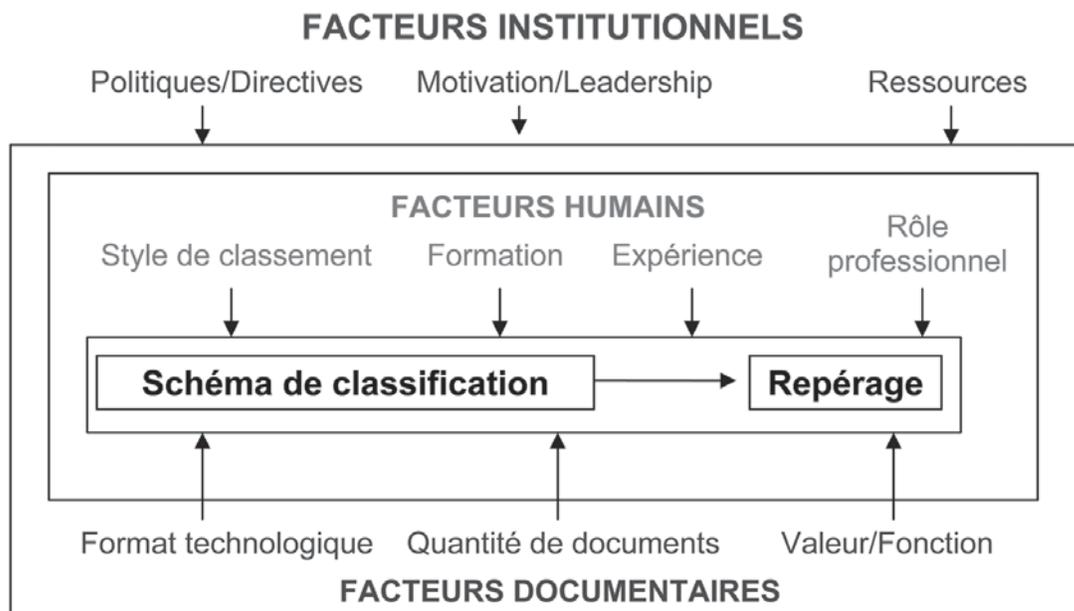


Figure 2. Facteurs d'influence de l'organisation personnelle des documents

Facteurs humains

Plusieurs chercheurs ont proposé une catégorisation des comportements classificatoires des individus dans l'utilisation des différents outils de gestion personnelle de l'information, en mettant l'accent sur l'activité d'organisation (Barreau 1995; Barreau et Nardi 1995; Whittaker et Snider 1996; Bälter 1997; Abrams, Baecker et Chignell 1998; Chapman 1999; Boardman 2001a; Henderson 2003; Ravasio, Schar et Krueger 2004). Ces études ont été guidées par les travaux de Cole (1982) et Malone (1983) qui ont identifié deux stratégies fondamentales d'organisation personnelle des documents : l'« empilement » (*piling*) et le « classement » (*filing*) des documents. Des recherches ultérieures ont montré que ces deux stratégies observées dans l'environnement papier sont aussi valables dans l'environnement électronique (Barreau 1995; Barreau et Nardi 1995; Henderson 2003; Boardman et Sasse 2004).

L'empilement des documents (*piling*) consiste à regrouper, sans ordre particulier et sans titre explicite, au même endroit physique ou spatial, le plus souvent sous forme

de piles, des documents n'ayant le plus souvent aucune relation entre eux (Cole 1982). Ce comportement permet de se rappeler des choses urgentes à faire, d'accéder très rapidement aux documents d'utilisation courante en les ayant «à la portée de la main», ou bien de «déposer» temporairement sur le bureau (virtuel ou non) les documents reçus (Malone 1983; Hertzum 1993; Kidd 1994; Barreau et Nardi 1995; Whittaker et Sidner 1996; Sellen et Harper 2001; Ravasio Schar et Krueger 2004).

Le classement des documents (*filing*) consiste à fournir un effort régulier pour classer les documents à l'aide d'un schéma de classification personnel ou dans des piles de documents relativement bien ordonnées et identifiées (Malone 1983; Whittaker et Snider 1996). Le schéma de classification personnel peut être conçu selon une approche déductive (*top-down*), c'est-à-dire sur la base d'une branche du schéma de classification des documents administratifs institutionnel, ou à partir d'une approche inductive (*bottom-up*), c'est-à-dire de manière incrémentale et informelle, davantage ajustée aux besoins d'un individu (Abrams, Baecker et Chignell 1998). Dans un environnement de travail, cette structure classificatoire personnelle, qui est généralement peu profonde, présente des avantages cognitifs pour les employés. Elle permet de mettre en ordre de grandes quantités de documents qui peuvent être réutilisés ultérieurement, pour des besoins de référence par exemple (Barreau 1995; Whittaker et Sidner 1996; Bergman Beyth-Marom et Nachmias 2003; Henderson 2003). Selon Jones *et al.* (2005), cette pratique permet aux individus de conserver un certain contrôle sur leur espace documentaire personnel (par exemple, s'assurer que tous les documents sont à une seule place) et d'établir des liens entre les documents qui sont sous leur contrôle direct.

Le choix de la stratégie de classification mise en place par les individus peut aussi être influencé, entre autres, par la nature des tâches à réaliser et l'ancienneté dans l'emploi actuel (Kidd 1994; Henderson 2003). La nature des tâches à réaliser ou le rôle professionnel (i.e., exécution ou encadrement) suggère qu'un individu peut avoir besoin d'accéder plus fréquemment à certains types de documents que d'autres, obligeant une organisation plus systématique de ses documents électroniques. L'ancienneté (ou les *années d'expérience*) dans le travail actuel est également considérée comme un facteur pouvant entraîner une accumulation d'une plus grande quantité de documents incitant l'employé à faire l'effort d'implanter un schéma de classification plus élaboré (Lansdale 1988; Barreau 1995; Whittaker et Snider 1996; Jones, Dumais et Bruce 2002; Henderson 2003).

Facteurs documentaires

Metzger et Lallich-Boidin (2004, 12) définissent l'espace documentaire comme un lieu «où s'organisent les collections» et un lieu «de mémoire, de structuration, de mise en relation de documents». Il est établi que le niveau de structuration de l'organisation de l'espace documentaire personnel est influencé par le format technologique et la dimension temporelle de l'information véhiculée par les documents constitutifs de cet espace (Barreau et Nardi 1995; Boardman 2001a; Boardman et Sasse 2004; Ravasio, Schar et Krueger 2004).

Format technologique

Dans le cadre de leur travail, les individus utilisent différents formats technologiques (par exemple, des documents papier, des courriels, des pages Web, des

documents bureautiques) qui peuplent leur espace documentaire personnel (Gagnon-Arguin 1999; Conférence des recteurs et des principaux des universités du Québec 2004; Henderson 2004). Ces formats technologiques sont le plus souvent organisés séparément selon différentes applications ou systèmes d'organisation (par exemple, un système de gestion des documents papier, un système de gestion des répertoires et fichiers, un système de gestion du courriel, un système de gestion des favoris) (Dumais *et al.*, 2003; Henderson 2003; Boardman et Sasse 2004). L'espace documentaire personnel est donc un espace fragmentaire, morcelé parmi les différents systèmes, et obligeant le plus souvent la conception par un même individu de plusieurs schémas de classification possédant des caractéristiques variées pour répondre à différents besoins personnels (Boardman 2001a; Boardman et Sasse 2004; Jones 2007).

Si des recherches récentes ont commencé à s'intéresser aux différents schémas de classification mis en place par un même individu pour organiser ses documents sous les différents systèmes qui lui sont proposés (Alvarado *et al.* 2003; Boardman et Sasse 2004), les recherches précédentes¹ relatives à l'organisation et au repérage des documents se sont le plus souvent limitées à un seul format technologique en raison de la difficulté d'analyse causée par cette diversité d'outils d'organisation et de besoins (Jones et Maier 2003).

Dans le cadre de notre recherche, nous avons restreint l'étude à l'organisation des documents administratifs créés par des applications bureautiques et qui sont localisés dans le système de répertoires et de fichiers du poste informatique personnel d'un individu. Ce choix est lié à des contraintes de temps et au fait que presque tous les documents de bureau sont depuis quelques années créés grâce à l'outil électronique et ne sont plus systématiquement imprimés sur papier ou enregistrés sur un serveur (Skupsky 1999; Dhérent et collaborateurs 2002; Patterson et Sprehe 2002; Conférence des recteurs et des principaux des universités du Québec 2004).

Dimension temporelle

Le niveau de structuration de l'organisation de l'espace documentaire personnel (et par conséquent le repérage à l'intérieur de cet espace) est également lié à la valeur informationnelle (l'utilité personnelle ou institutionnelle) des documents, qui évolue à travers le temps. Les recherches de Cole (1982), de Hertzum (1993) et de Barreau et Nardi (1995) ont permis d'identifier trois types d'information qui s'apparentent aux trois âges du cycle de vie des documents d'archives (Pérotin 1961) : 1) l'information action (*action information*), soit l'information d'intérêt à très court terme, et dont la durée de vie est de quelques heures ou de quelques jours, se manifeste par des documents tenus «à portée de main», le plus souvent empilés sur le bureau (virtuel ou non), 2) l'information de travail (*working information*), soit l'information pertinente pour une tâche, un projet ou une activité et qui est utilisée pendant plusieurs semaines, voire plusieurs mois, se manifeste par des documents vaguement organisés sur des étagères ou dans un système de classeurs ou de répertoires électroniques, et 3) l'information archivée (*archived information*), à savoir l'information de référence qui peut être conservée plusieurs mois ou plusieurs années sur les postes de travail des employés. L'utilité de l'information «archivée» pour l'individu n'est pas clairement établie. Bien que rarement utilisés, les documents qui véhiculent ce type d'information sont le plus souvent organisés et repérés à travers un vaste schéma de classification (Berlin *et*

al,1993; Kidd 1994; Barreau et Nardi 1995; Whittaker et Hirschberg 2001; Henderson 2003; Boardman et Sasse 2004).

Pour Hertzum (1999), l'évolution du niveau de structuration de l'organisation de l'espace documentaire personnel, qui part d'une organisation spatiale vaguement structurée et qui s'appuie sur la capacité de la mémoire pour aller ensuite vers des schémas de classification plus élaborés, suit le cycle de vie des documents qui passent de l'état «actif» (*action information*) à celui d'«inactif» (*archived information*). La caractérisation de cette dimension temporelle est utile dans le cadre de notre recherche, car elle rend également possible la catégorisation des modes de repérage mis en place par les individus allant de modes de repérage spatiaux ou qui s'appuient sur la capacité de la mémoire (*piles*) à des modes de repérage fortement structurés (*files*) (Henderson 2003).

Facteurs institutionnels

Plusieurs facteurs institutionnels influencent et favorisent l'individualisme qui caractérise le plus souvent l'organisation des documents électroniques dans un contexte de gestion décentralisée. La gestion des documents ne constitue généralement pas une priorité pour les organismes en raison d'un manque de motivation institutionnelle ou de ressources (Hedstrom 1996; Sprehe, McClure et Zellner 2002). Cette situation est renforcée par le fait que de nombreuses personnes pensent que les documents électroniques ne sont pas régis par les lois sur les archives ou ont du mal à faire une distinction entre les documents d'archives électroniques et les autres sources d'information électronique telles que les ressources Web, par exemple (Sprehe, McClure et Zellner 2002). Des enquêtes ont permis de constater dans la majorité des organismes la quasi-absence de politiques, normes, méthodes et procédures de gestion complète et intégrée des documents électroniques (Karjalainen *et al.* 2000; Sprehe, McClure et Zellner 2002). Il est établi que l'absence de politique institutionnelle ou de formation en matière de gestion des documents ou d'encadrement des employés (Sprehe, McClure et Zellner 2002), de directives ou de conseils (Conférence des recteurs et des principaux des universités du Québec 1994) sont des facteurs qui conduisent à la manipulation «sans règle ni discipline» des documents électroniques par les employés (Skupsky 1999, 30). En l'absence d'un cadre normalisé, un organisme se retrouve avec autant de façons de faire qu'il y a d'individus (Bergeron 1992; Jones 2007), ce qui entraîne une perte de contrôle sur une partie importante de la mémoire institutionnelle (Karjalainen *et al.* 2000; Sprehe, McClure et Zellner 2002).

Si les individus semblent pouvoir se retrouver plus ou moins efficacement dans leur propre schéma de classification (Chapman 1999; Boardman et Sasse 2004), qu'en est-il de la capacité d'une tierce personne à comprendre et interpréter cette structure classificatoire qualifiée d'«hermétique» (Senécal, 1999)? Autrement dit, quelle est la performance des schémas de classification personnels sur le plan du repérage de documents effectué par autrui? Dans la section suivante, nous expliquons les raisons pour lesquelles nous considérons l'efficacité du repérage comme moyen de comparer et d'évaluer les schémas de classification.

SCHÉMAS DE CLASSIFICATION ET REPÉRAGE

Les schémas de classification constituent une structure généralement hiérarchique de catégories descriptives conçue et utilisée par un individu ou un organisme pour faciliter l'organisation, le repérage et l'utilisation des documents. Nous entendons par repérage d'un document électronique l'accès par la localisation de l'emplacement d'un document (la classe ou le répertoire) à l'intérieur d'une structure arborescente de répertoires électroniques.

Pourquoi avoir choisi, dans le cadre de notre recherche, l'efficacité du repérage afin de mesurer la qualité d'un schéma de classification? Les schémas de classification produits par les employés pour organiser leurs documents électroniques ont pour fonctions principales, rappelons-le, de classer et de repérer ces documents. Le repérage des documents en milieu de travail vise à permettre aux individus d'utiliser et de réutiliser les documents conservés sur leur poste de travail ou sur celui de leurs collègues. Le repérage peut être considéré comme un processus important, car l'impossibilité de repérer un document implique, par exemple, une diminution de la productivité des employés, le ralentissement ou la cessation des activités, voire la perte de la mémoire institutionnelle. L'évaluation fonctionnelle ou opérationnelle de ces schémas de classification passe donc par l'évaluation de l'efficacité du repérage de documents ciblés réalisé par leur moyen.

Lien théorique entre schéma de classification et repérage

Cette section présente les présomptions théoriques qui justifient l'existence d'un lien entre le repérage et le schéma de classification utilisé ainsi que les propriétés ou qualités théoriques des schémas de classification qui visent à faciliter et améliorer ce repérage. La description de ces qualités est pertinente pour notre recherche, car elle permet de mieux comprendre la nature complexe de l'organisation et du repérage des documents. Par la suite, nous justifions le choix de l'efficacité du repérage comme moyen de comparer les schémas de classification des documents administratifs créés par des professionnels en gestion de l'information (schémas institutionnels) et ceux créés par des employés (schémas personnels).

Il existe des principes théoriques consensuels établissant l'existence d'un lien entre le schéma de classification et le repérage. Le postulat selon lequel l'organisation représentée par le schéma de classification facilite et améliore le repérage n'est pas nouveau et a été défendu par Bates (1977), Svenonius (1983), Cochrane (1985), Marcella et Newton (1994) et Lin et Chan (1999), parmi d'autres.

Lin et Chan (1999) expliquent bien le rôle de la classification et du schéma de classification pour l'organisation et le repérage des documents. La structure logique et hiérarchique du schéma de classification montre explicitement les relations entre les différentes classes, ce qu'une liste alphabétique ou aléatoire de termes ne révèle pas. Elle est également un moyen de naviguer systématiquement depuis des sujets généraux jusqu'à des sujets spécifiques, ce qui permet d'augmenter la précision du repérage (Chan 1995; Lin et Chan 1999).

Même s'il fait l'objet de critiques en raison des limites liées à son modèle structurel (hiérarchique), le schéma de classification semble influencer l'efficacité du repérage de

ressources documentaires par un ensemble d'individus. Ce postulat est appuyé par des réalisations concrètes, par exemple l'adoption ou l'adaptation des schémas de classification traditionnels, telle la Classification décimale de Dewey, pour normaliser l'organisation des ressources électroniques sur le Web en vue d'améliorer leur repérage. Dans le même sens, nous avons mentionné que de nombreux organismes utilisent ou recommandent l'utilisation des schémas de classification institutionnels pour construire de façon logique les répertoires et sous-répertoires pour l'organisation des documents électroniques (Dhérent et collaborateurs 2002; Beaupré 2004; Chamard 2004; Conférence des recteurs et des principaux des universités du Québec 2004). Nous savons également que, dans un contexte de gestion décentralisée des ressources informationnelles, les individus préfèrent généralement repérer leurs documents organisés selon un schéma de classification personnel plutôt qu'utiliser les fonctionnalités de recherche proposées par les systèmes d'exploitation en vue d'accéder plus rapidement aux documents recherchés.

La recension des écrits suggère que tous les schémas de classification ne permettent pas un repérage également efficace des documents par un groupe d'individus. Par exemple, les schémas de classification institutionnels, qui utilisent un langage formel, permettraient *a priori* de repérer plus facilement les documents que les schémas de classification personnels qualifiés «d'hermétiques» (Senécal 1999) ou «d'idiosyncrasiques» (Berlin *et al.* 1993; Adar, Karger et Stein 1999; Boardman 2001b; Saffady 2002). La présomption sous-jacente est qu'un schéma de classification qui possède certaines propriétés théoriques permet d'optimiser le repérage et les besoins documentaires de plusieurs individus ou groupes d'individus.

Certaines propriétés théoriques d'un schéma de classification qui ont une influence sur le repérage ont été identifiées par plusieurs chercheurs (Rousseau 1980; Couture et Rousseau 1982; Roberge 1985; Molholt 1995; Héon 1999; Zins 2002; Roberge 2004). Il s'agit principalement des propriétés suivantes :

1. Simplicité
2. Logique
3. Hospitalité
4. Autorité
5. Universalité

Nous décrivons brièvement chacune de ces propriétés en les mettant en relation avec les caractéristiques d'un schéma de classification :

1. *Simplicité* : un schéma de classification doit être simple afin d'être compris rapidement. La simplicité d'un schéma de classification se manifeste par une structure peu profonde (de trois à quatre niveaux maximum) pour faciliter la navigation tout en maintenant un certain niveau de spécificité. Pour être facilement compris, les termes utilisés pour nommer les répertoires (ou classes) doivent être concis et intelligibles à l'intérieur de la hiérarchie² (Hertzum et Frøkjær 1996; Lin et Chan 1999). L'utilisation d'une terminologie contrôlée³ pour la dénomination des catégories (ou répertoires) vise à en diminuer l'ambiguïté⁴ (Hertzum et Frøkjær 1996; Lin et Chan 1999).

2. *Logique* : le schéma de classification doit permettre de couvrir le champ des activités ou sujet de manière exhaustive, exclusive et cohérente (Molholt 1995; Zeller 2004). Cette qualité se manifeste par l'utilisation d'une seule caractéristique de

division logique (ou d'une caractéristique de division dominante) à chaque niveau de la hiérarchie. Un schéma de classification logique facilite théoriquement le repérage en permettant une recherche plus méthodique à l'intérieur d'une structure hiérarchique des répertoires (Maniez 1987; Fallis et Mathiesen 2000; Hunter 2000).

3. *Hospitalité* : le schéma de classification doit être conçu de manière à refléter les développements continus des activités de l'organisme tout en étant stable (Molholt 1995; Hertzum 2004). L'hospitalité d'un schéma de classification permet de s'assurer de l'intégration possible de l'ensemble des documents sans apporter de modifications majeures à la structure logique (au niveau de la macro-structure) (Rousseau 1980). Dans un contexte administratif, cette qualité se manifeste par l'utilisation des fonctions et activités administratives comme premiers critères de division logique (Renaud 1975; Héon 1995; Héon 1999; Sabourin 2001; Roberge 2002; Roberge 2004).

4. *Autorité* : un schéma de classification doit être construit sur la base d'un consensus (Molholt 1995). Cette qualité assure le niveau d'accessibilité du schéma de classification par tous les utilisateurs. Dans un contexte de gestion décentralisée des ressources informationnelles, cette qualité se manifeste notamment par l'adoption par l'individu du schéma de classification institutionnel (ou l'une de ses branches) pour créer, nommer (à partir de l'énoncé des classes) et organiser les répertoires et sous-répertoires électroniques sur son poste de travail informatique (Beaupré 2004; Chamard 2004; Roberge 2004).

5. *Universalité* : un schéma de classification doit pouvoir être utilisé par n'importe quelle application ou environnement (Molholt 1995). Par exemple, dans un contexte de gestion centralisée des ressources informationnelles au sein d'un organisme, cette qualité peut se manifester par l'utilisation d'un schéma de classification unique pour l'organisation de l'ensemble des documents électroniques ou analogiques. Cette qualité favorise l'uniformité dans la classification des documents conservés sur plusieurs applications différentes ou sur différents supports et permet, par conséquent, la constitution de dossiers complets au sein d'un organisme (Roberge 2004).

En résumé, les présomptions théoriques nous permettent de prédire qu'il existe un lien entre l'utilisation d'un schéma de classification institutionnel pour organiser les documents (qui posséderait *a priori* les propriétés théoriques d'un schéma de classification) et l'observation d'un repérage efficace de ces documents comparativement à l'utilisation d'un schéma de classification personnel (qui ne posséderait pas les propriétés théoriques). En effet, un schéma de classification, comme n'importe quel langage, est « porteur » d'information. Il nous communique à travers ses caractéristiques une représentation plus ou moins formelle d'une collection documentaire particulière. Ces différentes caractéristiques interagissent entre elles et jouent un rôle important dans l'interprétation qui est faite du schéma de classification pour des fins de repérage de documents. Notre recherche vise à vérifier non seulement l'influence du schéma de classification sur l'efficacité du repérage, mais surtout l'importance relative de chacune de ses caractéristiques spécifiques au moment du repérage.

Facteurs d'influence de la qualité du repérage

Dans notre recherche, le repérage des documents électroniques dans un système de classification hiérarchique est considéré comme le produit de l'interrelation de deux

principaux éléments puisque nous demandons à 1) un individu (ou agent humain) d'utiliser 2) un mécanisme de repérage donné (une structure arborescente de répertoires considérée comme un schéma de classification) pour repérer des documents. La section suivante décrit ces deux éléments.

Agent humain

Les recherches sur les pratiques de repérage observées dans le domaine de la gestion personnelle de l'information consignées sur support électronique ont permis de distinguer deux situations distinctes : 1) le repérage est effectué par le créateur et organisateur même de l'espace documentaire, et 2) le repérage est effectué par une tierce personne étrangère au contenu et à l'organisation de cet espace documentaire.

Dans la situation où le repérage est réalisé par le créateur et organisateur de l'espace documentaire, Lansdale (1988) et Berlin *et al.* (1993) rapportent que le repérage consiste à localiser un document déjà manipulé et qui a été personnellement nommé et classé à l'intérieur de la structure arborescente des répertoires. Dans ce cas, le repérage est basé sur des processus cognitifs tels que la *mémorisation* d'une ou de plusieurs caractéristiques du document recherché (Lansdale et Edmonds 1992; Berlin *et al.* 1993; Barreau et Nardi 1995; Adar, Karger et Stein 1999; Farrell *et al.* 2002; Alvarado *et al.* 2003; Ravasio, Schar et Krueger 2004). Lansdale et Edmonds (1992) indiquent qu'il peut s'agir de l'apparence du document, du contexte temporel (« *We were handling that just before Christmas*») ou de sa relation avec l'information contextuelle (par exemple, le contexte de production ou de réception passé ou présent). D'autres individus, au contraire, privilégient l'effort cognitif lié à la *classification* pour réduire l'effort cognitif lié au repérage basé sur la capacité de la mémoire (Lansdale 1988). Quelle que soit l'approche (mémorisation ou classification), Chapman (1999) et Boardman et Sasse (2004) observent que la situation dans laquelle un individu est incapable de repérer ses propres documents électroniques reste occasionnelle tout en étant perçue comme extrêmement frustrante. L'échec au niveau du repérage d'un document serait lié à la destruction, à l'archivage des documents, à une organisation documentaire désordonnée ou à une erreur de classement (Boardman et Sasse 2004).

Le repérage effectué par une tierce personne implique de localiser un document qui n'a probablement jamais été entrevu ni manipulé, mais qui se trouve classé là, «quelque part», dans le schéma de classification créé par un collègue ou un ancien employé, depuis plusieurs mois, ou plusieurs années, dans le cadre d'un contexte de production particulier (Hertzum 1999). Le repérage d'un document s'appuie ici sur la capacité de cette personne à comprendre et à interpréter un schéma de classification conçu par autrui (Hertzum 1999). Dans ce cas, le processus de repérage peut être qualifié de subjectif puisqu'il existe plusieurs interprétations possibles d'un même schéma en raison du «caractère pluriel de toute lecture» (Senécal 1997-1998, 53). Les recherches sur la notion de genre permettent cependant d'anticiper que la qualité de l'interprétation d'un schéma de classification personnel dépend de l'appartenance de l'interprète à la même communauté de discours⁵ que le créateur et utilisateur principal de ce schéma (Orlikowski et Yates 1994). Berlin *et al.* (1993) donnent l'exemple de documents relatifs au congrès INTERCHI («*International Conference on Human Factors in Computing Systems*») qui seraient en fait classés sous un répertoire intitulé CHI («*Conference on*

Computer-Human Interaction»). Une telle simplification ne serait significative que pour les individus (par exemple, les membres d'une même équipe ou communauté scientifique) qui connaîtraient déjà la relation entre INTERCHI et CHI.

Alors que des recherches empiriques ont évalué le repérage de documents électroniques par le producteur même de l'espace documentaire (Barreau et Nardi 1995; Nardi et Barreau 1997; Chapman 1999; Boardman et Sasse 2004; Ravasio, Schar et Krueger 2004), aucune ne s'est intéressée au repérage réalisé par une tierce personne.

Mécanismes de repérage des documents

La façon dont les documents peuvent être repérés est étroitement liée à la capacité des mécanismes de repérage proposés aux individus (Chu, 2003). Gagnon-Arguin et Bannouri (1998) notent que, dans le cas des documents électroniques, le repérage n'est pas aussi lié à la classification que pour les documents sur support papier, les systèmes informatiques proposant d'autres moyens pour effectuer cette tâche. Dans le même sens, Barreau et Nardi (1995) et Ravasio, Schar et Krueger (2004) identifient deux mécanismes fondamentalement différents conçus et proposés aux individus pour effectuer le repérage des documents électroniques entreposés sur le poste informatique d'un individu : 1) l'utilisation d'expressions logiques, c'est-à-dire l'utilisation d'un outil de recherche proposé par le système d'exploitation, ou 2) l'accès aux documents à partir de la structure arborescente de répertoires.

Dans le cas d'un repérage «logique» (i.e. qui utilise des expressions logiques), qui est aussi appelé «recherche» (Boardman et Sasse 2004) ou «téléportage» (Alvarado *et al.* 2003), un individu se «téléporte» directement au document recherché à partir de la saisie de mots-clés ou du nom du fichier dans un outil de recherche.

Le repérage «manuel» (ou sensoriel⁶) des documents est aussi désigné par les termes «navigation» (Barreau et Nardi 1995; Boardman et Sasse 2004) ou «orientation» (Alvarado *et al.* 2003). Ce mécanisme de repérage implique l'utilisation de l'information contextuelle pour se rapprocher du document cible recherché à travers l'arborescence de répertoires, souvent par une série d'étapes (O'Day et Jeffries 1993; Alvarado *et al.* 2003; Ravasio, Schar et Krueger 2004). L'individu doit d'abord identifier le répertoire où le document est classé et doit ensuite afficher la liste des documents contenus dans ce répertoire. L'identification du document recherché s'effectue à partir de l'interprétation de la signification des noms de répertoires et de fichiers ou d'autres attributs donnés par le système (par exemple, la date de modification ou la taille du fichier) (Jones, Bruce et Dumais 2001; Boardman et Sasse 2004).

Barreau et Nardi (1995), Alvarado *et al.* (2003) et Ravasio, Schar et Krueger (2004) rapportent que les utilisateurs sont généralement très réticents à faire le repérage au moyen des outils de recherche en raison de l'effort cognitif requis pour spécifier complètement le besoin documentaire et toutes ses métadonnées (par exemple, la fonctionnalité de recherche de Windows oblige le plus souvent à mémoriser les noms exacts des fichiers correspondants aux documents recherchés). Il apparaît dans les études empiriques que seuls les individus expérimentés dans l'utilisation d'un ordinateur personnel utilisent occasionnellement ces outils, et cela en dernier recours, c'est-à-dire lorsqu'ils n'ont aucune idée de l'endroit où le document peut être entreposé, lorsqu'ils

ont besoin d'accéder à une information très précise ou lorsqu'ils ont besoin de repérer un document dans un espace de stockage partagé dont le contenu documentaire est mal connu (Nardi et Barreau 1997; Alvarado *et al.* 2003; Boardman et Sasse 2004; Ravasio, Schar et Krueger 2004).

Dans notre recherche, nous évaluons prioritairement l'efficacité du repérage manuel d'un ensemble donné de documents à partir de la navigation dans la structure arborescente de répertoires et sous-répertoires, que nous considérons comme un outil d'organisation intellectuelle et physique (i.e. comme un schéma de classification) des espaces documentaires électroniques personnels.

Mesures de l'efficacité du repérage

L'utilisation du repérage de documents comme moyen d'évaluer ou de mesurer la qualité (ou qualimétrie) des schémas de classification (ou tout autre outil de repérage) est connue en informatique et en sciences de l'information (Zins 2002). Plusieurs études ont utilisé le repérage pour évaluer la performance opérationnelle des systèmes de recherche automatique de l'information (Cleverdon 1967; Lancaster 1968; Salton 1989; Borlund et Ingwersen 1998), des catalogues et autres instruments de repérage en ligne (Krikelas 1969; Lytle 1980; Hertzum et Frøkjær 1996; Ribeiro 1996; Hildreth 2000), ou des systèmes personnels de classification (Malone 1983; Barreau et Nardi 1995; Chapman 1999; Alvarado *et al.* 2003; Dumais *et al.* 2003; Boardman et Sasse 2004; Ravasio, Schar et Krueger 2004). La précision (i.e. la proportion de documents pertinents retrouvés parmi tous les documents retrouvés par le système de recherche d'information) et le rappel (i.e. la proportion de documents pertinents retrouvés par un système de recherche d'information parmi tous les documents pertinents existant dans une base de documents) sont les deux critères utilisées dans le domaine de recherche d'information (RI)⁷ pour mesurer l'efficacité des systèmes de repérage automatisé de l'information.

Dans notre recherche, qui vise à mesurer et comparer l'efficacité du repérage d'un même ensemble de documents préalablement ciblés à l'aide de différents schémas de classification, et non pas l'efficacité d'un système de repérage en lui-même indépendamment des autres, les notions classiques de précision et de rappel ne s'appliquent pas (en raison de l'absence, dans notre recherche, d'une notion qui est centrale en IR, soit la notion de pertinence à une requête). Les mesures utilisées dans les travaux de Toms et Campbell (1999) (i.e. le *nombre de documents correctement identifiés* et le *temps requis pour l'identification*), Ribeiro (1996) et Chapman (1999) (i.e. le *nombre d'essais*) sont plus appropriées pour évaluer les résultats du repérage de documents déjà ciblés. Alors que le nombre de documents correctement identifiés s'apparente au calcul de la précision du repérage, le temps requis pour le repérage et le nombre de tentatives pour repérer un document sont des indicateurs de mesure de l'efficacité d'un système (analyse des coûts du repérage en termes de temps et d'effort) (Ribeiro 1996).

Par conséquent, nous avons retenu les trois variables suivantes pour analyser et mesurer l'efficacité du repérage de documents dans un environnement contrôlé : 1) la proportion de documents cibles repérés, 2) le temps moyen requis (mesuré en secondes) pour repérer un document et 3) la proportion de documents repérés dès le premier essai. Ces trois variables qui mesurent le succès, la rapidité et la facilité du

repérage de documents constituent des critères pertinents d'évaluation de l'efficacité du repérage réalisé à l'aide de différents schémas de classification.

L'ORGANISATION PERSONNELLE DES DOCUMENTS ADMINISTRATIFS ÉLECTRONIQUES ET REPÉRAGE : RÉSULTATS D'UNE ÉTUDE EMPIRIQUE

Les assises théoriques qui précèdent ont permis de mener une étude empirique à partir de situations réelles d'organisation et de repérage de documents administratifs électroniques. Nous en présentons la méthodologie et les résultats.

Méthodologie

Un protocole de recherche en trois étapes fut réalisé pour atteindre les deux objectifs spécifiques que nous nous étions fixés, soit de 1) décrire les caractéristiques de différents schémas de classification utilisés par des employés pour organiser et classer les documents administratifs électroniques qui sont sous leur contrôle immédiat et de 2) vérifier, dans un environnement contrôlé, les différences sur le plan de l'efficacité du repérage de documents électroniques en fonction du schéma de classification employé.

La première étape du projet a consisté dans l'identification et la description des caractéristiques de 21 schémas de classification personnels utilisés par des employés pour organiser les documents administratifs électroniques qui sont sous leur contrôle. Cette caractérisation de schémas de classification personnels, effectuée à l'aide d'un modèle d'analyse, nous a permis d'approfondir notre connaissance de ce type de schémas. Dans la deuxième étape de notre projet, nous avons procédé à plusieurs manipulations des caractéristiques d'un schéma de classification personnel. L'objectif de cette opération était de constituer un corpus de cinq schémas de classification présentant des caractéristiques variées au plan structurel, logique et sémantique en vue de permettre des comparaisons. Dans la troisième étape, une simulation a été réalisée pour collecter des données sur l'efficacité du repérage d'un ensemble de documents électroniques réalisé à l'aide de schémas de classification ayant des caractéristiques variées.

Pour chacune de ces étapes, nous précisons quelles ont été les méthodes de collecte et d'analyse des données appliquées, ainsi que les procédures d'échantillonnage et la taille de l'échantillon utilisées.

Étape 1 : Caractérisation des schémas de classification

Le premier objectif de notre recherche était de décrire les caractéristiques de schémas de classification personnels des employés. En l'absence d'un instrument d'analyse spécifique des schémas de classification personnels, nous avons conçu un modèle ou grille d'analyse qui tient compte des caractéristiques multidimensionnelles de ces langages classificatoires particuliers. La littérature scientifique a servi de cadre de référence pour la conception de cette grille d'analyse des schémas de classification personnels. Plus précisément, nous nous sommes penchées sur la théorie de l'organisation des connaissances (Maniez 1987 ; Iyer 1995), les méthodes de travail et les résultats de recherches dans les domaines de la gestion personnelle de l'information

(Jones, Dumais et Bruc, 2002; Gonçalves et Jorge 2003; Boardman et Sasse 2004; Henderson 2003; Khoo *et al.* 2007) et des sciences de l'information (Van der Walt 1998; Hudon 2001; Zins 2002). Cette grille d'analyse nous a permis d'examiner et de décrire de manière uniforme et systématique les caractéristiques structurelles (par exemple, le nombre de classes principales), logiques (par exemple, le critère de division privilégié) et sémantiques (par exemple, les stratégies d'abréviation observées dans la dénomination des classes) des schémas de classification personnels.

Dans le domaine de la gestion personnelle des ressources informationnelles («*Personal information management*»), les schémas de classification personnels ont été caractérisés à partir d'indicateurs de mesure de leur *structure hiérarchique ou arborescente* (Gonçalves et Jorge 2003; Boardman et Sasse 2004; Henderson 2004; Khoo *et al.* 2007). Les principaux indicateurs de mesure utilisés sont relatifs à l'étendue (i.e. au *nombre de classes principales* et au *nombre moyen de classes à chaque niveau*) et à la profondeur (par exemple, le *nombre maximum* et le *nombre moyen de niveaux hiérarchiques*). D'autres indicateurs sont relatifs à la complexité (i.e. au *nombre total de classes*) et à l'équilibre de la structure arborescente (i.e. à l'*écart-type de la moyenne du nombre de classes à chaque niveau*) (Gonçalves et Jorge 2003; Henderson 2004).

La théorie de la classification (Maniez 1987; Richmond 1990; Langridge 1992; Iyer 1995; Bowker et Star 2000) confirme la dimension *structurelle* des schémas de classification puisqu'il y est fait mention de profondeur (par exemple, le *nombre de niveaux*) et d'étendue (par exemple, le *nombre de classes principales*) de la structure. En sciences de l'information, les caractéristiques structurelles des schémas de classification conçus pour organiser les ressources Web ont ainsi été étudiées en regard des règles théoriques de classification (Van der Walt 1998; Hudon 2000; Zins 2002). Les principaux indicateurs utilisés par ces chercheurs sont essentiellement le *nombre de classes* et le *nombre de niveaux hiérarchiques*.

Est-ce que ces indicateurs sont suffisants pour caractériser un schéma de classification? On ne peut limiter l'analyse des schémas de classification uniquement à leur dimension structurelle. L'étude de la théorie de la classification nous apprend que la classification doit aussi être logique et compréhensible.

Un schéma de classification doit théoriquement respecter le principe de division logique qui consiste à utiliser *une seule caractéristique* à chaque étape ou niveau de division, ce qui permet d'assurer l'exclusivité mutuelle des classes (Maniez 1987; Van Der Walt 1998; Bowker et Star 2000). Selon ce principe, les différentes notions sont réparties dans un ordre logique à l'intérieur de classes qui s'emboîtent du général au plus précis, ce qui permet de répondre généralement aux attentes des utilisateurs, de fournir un élément de cohérence important et de permettre une exploration méthodique (Maniez 1987; Maniez 1991; Langridge 1992; Canonne 1993; Molholt 1995; Jacob 2004). Notre modèle prévoit une analyse des concepts représentés par les énoncés des classes utilisées pour classer les documents. Cette analyse fait la distinction entre les concepts représentés par les classes du premier et deuxième niveau hiérarchique en vue d'identifier les critères de division logique qui sont privilégiés par les employés et de vérifier les résultats de Boardman et Sasse (2004), Henderson (2005) et Khoo *et al.* (2007).

Le modèle d'analyse de la dimension logique des schémas de classification examine l'ordre de succession des classes. Cet ordre de succession des classes peut être

sans importance puisque cela n'affecte en rien la logique du schéma de classification. Toutefois, la similitude de l'ordre de succession des classes principales d'un schéma de classification à un autre permet une adaptation et donc une utilisation plus rapide d'un schéma de classification dont la structure est familière. Nous avons examiné dans notre recherche l'ordre de succession des classes privilégié dans les schémas de classification personnels observés.

En résumé, nous avons retenu les indicateurs de mesure suivants pour décrire la dimension logique des schémas de classification utilisés par les employés pour organiser et classer leurs documents électroniques : 1) critère de division logique aux premier et deuxième niveaux hiérarchiques et 2) ordre de succession des classes.

Un schéma de classification est aussi un langage ou un instrument de communication qui comprend des concepts et des classes porteurs d'un contenu *sémantique*. Les termes (simples ou composés) utilisés pour représenter les concepts doivent permettre la compréhension du contenu des classes et des documents qui y sont rattachés (Dumais et Landauer 1983). La conception et l'application de règles théoriques ou de conventions relatives à la dénomination des répertoires visent à améliorer la capacité à comprendre la signification de la structure hiérarchique et logique et à repérer des documents.

Alors que plusieurs chercheurs se sont intéressés aux différentes approches utilisées par les individus pour nommer leurs documents électroniques (Carroll 1982; Chapman 1999), très peu de recherches ont porté sur la dénomination même des répertoires électroniques (Gonçalves et Jorge 2003). Nous avons examiné les variables utilisées par les chercheurs qui se sont intéressés aux différentes approches employées par les individus pour nommer leurs documents électroniques (Carroll 1982; Chapman 1999). L'examen des résultats de leurs travaux a permis d'identifier trois types de données que nous avons jugées pertinentes pour caractériser les schémas de classification sur le plan sémantique : 1) la longueur moyenne (en nombre de caractères) des noms de classes, 2) les stratégies d'abréviations utilisées et 3) le degré de redondance structurelle et conceptuelle.

Avec le modèle d'analyse que nous venons de décrire, nous avons étudié les caractéristiques de 21 schémas de classification utilisés par des employés de l'Université de Montréal, soit des techniciennes à la coordination de travail de bureau (onze schémas de classification personnels dont cinq partagés avec des collègues de travail) et des coordonnateurs de stage (dix schémas de classification personnels dont deux schémas partagés). Trois raisons sous-tendent le choix de ces deux catégories particulières d'employés :

1. le rapport quasi-quotidien que ces deux catégories d'employés entretiennent avec des documents électroniques dans le cadre de leurs activités respectives, ce qui permet de supposer :
2. l'application régulière d'une classification pour organiser un corpus volumineux de documents électroniques, et d'envisager :
3. l'existence de différences et de ressemblances dans les caractéristiques des schémas de classification utilisés dues à la variété et à l'homogénéité des activités auxquelles participent chaque technicienne et coordonnateur de stage.

Nous avons rencontré individuellement chaque employé sur son lieu de travail, devant son poste informatique. L'objectif de ces rencontres était de collecter, en présence de l'employé, des données de nature quantitative (par exemple, le nombre total de classes) et qualitative (i.e. la description du contenu des répertoires) sur le schéma de classification personnel utilisé pour l'organisation des documents électroniques. Nous avons réalisé des saisies d'écran de l'arborescence des classes situées dans le répertoire racine « Mes Documents » (ou son équivalent). L'objectif de cette opération était d'obtenir une « photo » de la structure arborescente des classes et sous-classes créées par l'employé. Préalable à cette opération, l'employé avait été invité à indiquer les classes ou sous-classes (généralement une ou deux classes par schéma) dont il ne désirait pas faire la présentation ou la description pour des raisons de confidentialité ou de respect de la vie privée. Ces classes n'ont pas été incluses dans l'analyse des schémas de classification personnels.

Pour accélérer la collecte des données quantitatives relatives aux caractéristiques structurelles des schémas de classification, nous avons utilisé, avec l'assentiment de ses concepteurs, le programme informatique d'analyse de l'espace documentaire personnel (PDS) (Gonçalves et Jorge 2003).

Ce logiciel a permis une collecte et une analyse statistique automatisées d'une partie des données sur les caractéristiques structurelles des schémas de classification personnels (i.e. le nombre total de classes et de sous-classes, le nombre total de fichiers, le nombre moyen de classes à chaque niveau hiérarchique et l'écart-type de la moyenne du nombre de classes à chaque niveau hiérarchique).

Étape 2 : Manipulation des schémas de classification

Le deuxième objectif de notre recherche consistait à vérifier les différences entre différents schémas de classification quant à leur efficacité pour le repérage de documents électroniques. La réalisation de cet objectif supposait la possibilité de disposer d'un échantillon de schémas de classification présentant des variations suffisantes aux plans structurel, logique et sémantique pour fins de comparaison. Or, les caractéristiques structurelles, logiques et sémantiques relativement semblables des schémas de classification personnels observés ne permettaient pas de neutraliser l'effet des autres caractéristiques lorsqu'on vérifie l'influence d'une caractéristique donnée. Afin de pouvoir réaliser notre deuxième objectif de recherche, nous avons envisagé de compléter l'analyse de schémas existants par la création de schémas de classification artificiels. Le but de cette manipulation était d'assurer la représentativité de différents types de schémas de classification et de tester, dans une deuxième phase de collecte de données, leur influence sur l'efficacité du repérage de documents électroniques. Pour ce faire, nous avons sélectionné le schéma de classification personnel (Schéma A) d'un coordonnateur de stage, que nous avons jugé représentatif des différentes caractéristiques observées dans les schémas de classification personnels (Voir Figure 3).

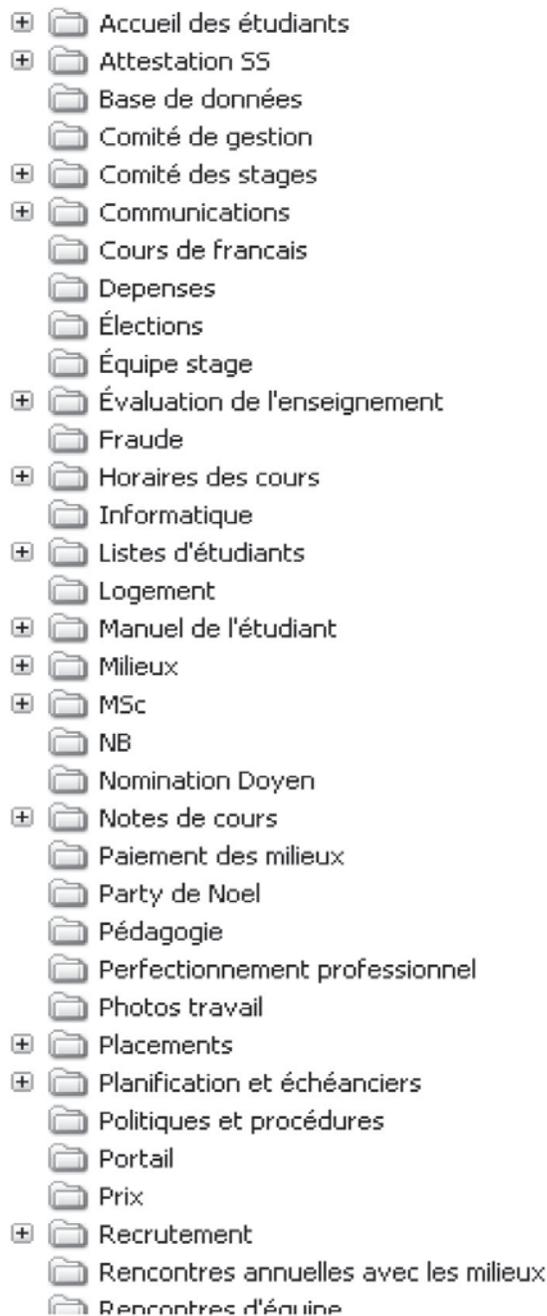


Figure 3. Schéma de classification personnel d'un coordonnateur de stage (Schéma A)

Nous avons manipulé ce schéma personnel pour créer artificiellement quatre schémas de classification présentant des variations significatives sur le plan structurel, logique et sémantique (Voir Annexe 1). Les manipulations effectuées sont les suivantes :

1) Manipulation de la sémantique (schéma B) : L'objectif de cette manipulation était de modifier la dimension sémantique tout en maintenant la structure classificatoire et la logique du schéma de classification de l'employé (schéma A). L'approche retenue a été de rajouter la description ou la définition de chacune des classes et sous-classes

telles que communiquées par l'employé en entrevue (Voir Figure 4). Cette approche est comparable à l'utilisation de notes d'application dans les schémas de classification institutionnels. La création d'un tel schéma de classification visait à permettre d'évaluer l'influence de la dimension sémantique sur le repérage des documents.

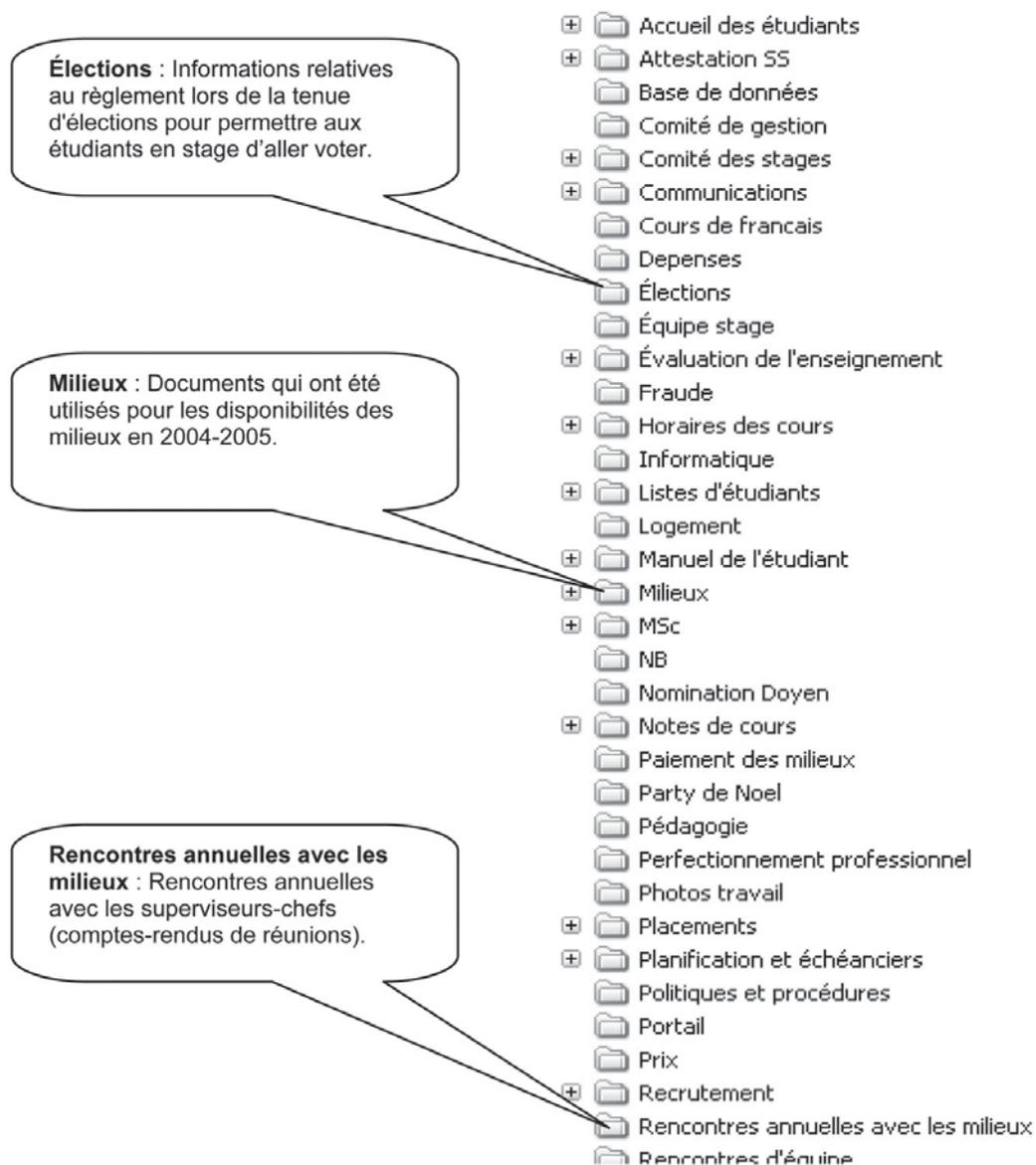


Figure 4. Schéma de classification manipulé au plan sémantique (Schéma B)

2) Manipulation de la structure et de la sémantique (schéma C) : L'objectif de cette manipulation était de supprimer toute structure hiérarchique ou arborescente. Cette approche modifie la dimension sémantique en augmentant : 1) la longueur moyenne (en nombre de caractères) des intitulés des classes (i.e., une longueur moyenne de 37.9 caractères), 2) le niveau de redondance structurelle et conceptuelle (90 % des classes sont redondantes) et 3) la proportion de classes abrégées (48 %) (Voir Figure 5). La création d'un tel schéma de classification visait à permettre d'évaluer l'influence

de l'absence de structure arborescente ainsi que l'influence de la propriété sémantique des noms des classes électroniques sur le repérage des documents.



Figure 5. Schéma de classification manipulé aux plans sémantique et structurel (Schéma C)

3) Manipulation de la structure et de la logique (schéma D) : L'objectif de la manipulation des dimensions logique et structurelle était de modifier la structure et la logique tout en maintenant la dimension sémantique du schéma A. Les caractéristiques du schéma de classification produit étaient conformes aux règles conventionnelles de

conception d'un schéma de classification des documents administratifs électroniques (par exemple, un nombre peu élevé de classes principales, une division logique basée sur les activités de l'employé) tout en conservant la sémantique du schéma original de l'employé (Voir Figure 6). Cette manipulation consistait à utiliser le schéma de classification préalablement conçu et conforme aux exigences archivistiques (Voir Figure 7) tout en maintenant la sémantique de l'employé. La création d'un tel schéma de classification visait à permettre d'évaluer l'influence de la propriété sémantique des noms des classes sur le repérage des documents.

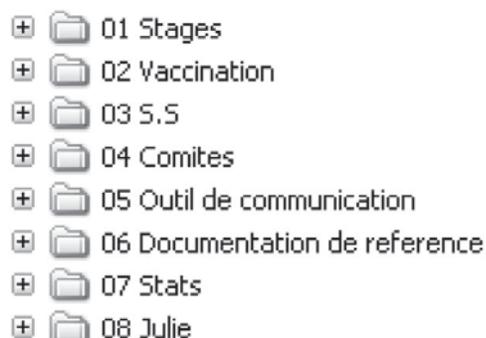


Figure 6. Schéma de classification manipulé aux plans structurel et logique (Schéma D)

4) Manipulation de la structure, de la logique et de la sémantique (schéma E) : L'objectif de cette manipulation était de modifier simultanément l'ensemble des caractéristiques logique, structurelle et sémantique du schéma de classification produit et utilisé par l'employé (schéma A) pour produire un schéma de classification davantage conforme aux règles théoriques de conception d'un schéma de classification des documents administratifs électroniques (par exemple, un nombre de classes principales peu élevé, le critère de division basé sur les activités de l'employé, peu d'abréviations) (Voir Figure 7). Ce schéma de classification a été validé auprès d'un archiviste de la Division des archives de l'Université de Montréal.

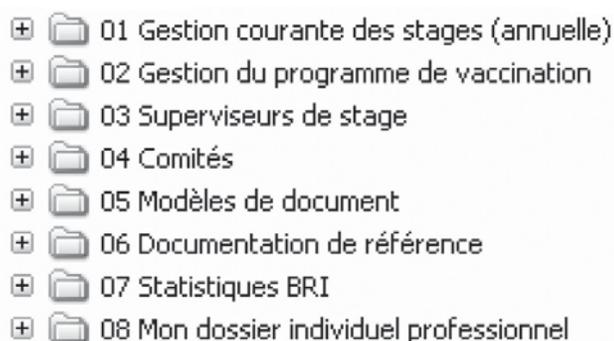


Figure 7. Schéma de classification manipulé aux plans structurel, logique et sémantique (Schéma E)

L'efficacité du repérage d'un même ensemble de documents électroniques réalisé à l'aide de chacun de ces cinq schémas de classification a été mesurée dans le cadre d'une simulation. La section suivante présente cette dernière étape de notre protocole de recherche.

Étape 3 – Simulation du repérage

L'efficacité du repérage de documents a été mesurée à l'aide de trois variables : le taux de succès du repérage, le temps moyen requis (en secondes) pour repérer les documents et la proportion de documents repérés dès le premier essai. Nous voulions savoir s'il était possible de repérer un document avec la même efficacité, quel que soit le schéma de classification utilisé pour ce faire, ou s'il existait des différences statistiquement significatives quant à l'efficacité au repérage des cinq schémas testés.

Les cinq schémas de classification ont été utilisés par 70 répondants (20 hommes et 50 femmes) pour repérer 20 documents. Les répondants étaient principalement des étudiants ou des diplômés de l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal. Le choix de cette institution comme base de recrutement a été guidé par la nécessité de contrôler les variables étrangères (par exemple, l'habileté à naviguer dans une structure arborescente de répertoires) en visant l'homogénéité du groupe de répondants. Chaque individu qui répondait aux critères d'inclusion a été ensuite assigné aléatoirement⁸ à l'un des 14 groupes afin de diminuer l'influence que des variables étrangères à la recherche pourraient avoir sur les variables étudiées.

Nous avons reproduit sur un ordinateur portable HP doté du système d'exploitation Windows XP professionnel les cinq structures arborescentes des répertoires correspondant aux cinq schémas de classification testés. En vue de minimiser les distractions, toutes les séances de simulation ont eu lieu dans le même environnement soit un local isolé de l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal, dans des conditions ambiantes identiques.

Tous les répondants ont repéré les 20 documents en alternant l'utilisation des différents types de schémas de classification. Pour neutraliser les effets d'apprentissage et de fatigue, l'ordre des documents à repérer ne variait pas, seule l'utilisation des schémas de classification variait systématiquement selon le groupe d'appartenance du répondant. La durée moyenne d'une séance de repérage (comprenant l'explication de la procédure et le repérage des documents proprement dit) était d'une heure et vingt minutes. Les réponses fournies par chaque répondant ont été compilées sur une grille d'observation (ou feuille de réponse) qui mentionne pour chacun des documents à repérer, le schéma de classification testé, le nombre d'essais et le temps requis (mesuré en secondes) par le répondant pour repérer le document. Les commentaires des répondants relatifs à leur satisfaction dans l'utilisation de chacun des schémas de classification ont été aussi colligés sur la grille d'observation.

Les données inscrites sur les grilles d'observation ont été retranscrites dans un fichier Excel pour des fins d'analyse. L'analyse de variance (ANOVA)⁹ à mesures répétées a été réalisée par l'équipe du Service de consultation en méthodes quantitatives de l'Université de Montréal en utilisant le logiciel d'analyse statistique en sciences sociales SPSS («*Statistical package for social sciences*»). Le test de Bonferroni¹⁰ pour les

comparaisons sur l'ensemble des schémas de classification a permis de garantir le niveau de confiance global des résultats à un niveau significatif de 5% de marge d'erreur (i.e. on accepte de se tromper 1 fois sur 20) ($p < 0.05$).

Résultats – Phase 1 : Caractérisation des schémas

La structure, la logique et la sémantique des 21 schémas de classification personnels furent examinées. Nous présentons ici un résumé des caractéristiques observées pour chacune des dimensions analysées. On trouvera une description plus détaillée de ces résultats et de leur interprétation dans Mas (2007).

Structure

Nous rappelons que les caractéristiques structurelles des schémas de classification personnels ont été décrites à partir de l'analyse de l'étendue (i.e. du nombre de classes principales et du nombre moyen de classes à chaque niveau hiérarchique), de la profondeur (i.e. du nombre maximum de niveaux hiérarchiques et du nombre moyen de niveaux hiérarchiques), de l'équilibre (i.e. de l'écart-type de la moyenne du nombre de classes à chaque niveau) et de la complexité des structures (i.e. du nombre total de classes et de sous-classes).

Les schémas de classification observés présentent des structures peu profondes avec un nombre moyen de niveaux de 1.67 (écart-type = 0.46). Le nombre total de classes et de sous-classes varie considérablement, allant d'un minimum de 13 à un maximum de 3541 classes avec une moyenne de 355.05 classes. L'écart-type de 742.25 confirme cette dispersion sur le plan de la complexité des structures. Le nombre de classes principales varie d'un minimum de 6 à 105 avec une moyenne de 28.14 classes principales (écart-type = 22.93) par schéma de classification. Le nombre moyen de classes à chaque niveau est généralement constant et varie entre 2 et 3 (écart-type = 0.89). Ces résultats correspondent à une macrostructure généralement étendue (nombreuses classes principales) et à une microstructure peu étendue (petit nombre de classes à chaque niveau inférieur) (Khoo *et al.* 2007). L'écart-type de la moyenne du nombre de classes à chaque niveau montre que les schémas de classification des techniciennes et les schémas de classification non partagés des coordonnateurs de stages présentent une structure peu équilibrée, avec un écart-type qui varie entre 6 et 7. Autrement dit, l'étendue de la structure n'est pas uniforme, certaines classes se subdivisent en une seule sous-classe (fréquent) alors que d'autres classes peuvent contenir plus d'une vingtaine de sous-classes (plus rare). Seuls les schémas de classification partagés de deux coordonnateurs de stage présentent un très faible écart-type de 2.57 correspondant à une structure plus équilibrée.

Les caractéristiques structurelles des schémas de classification personnels observés ont été comprises et interprétées dans Mas (2007) en tenant compte de plusieurs contextes. Il s'agit du contexte documentaire (par exemple, la quantité de documents à organiser), du contexte professionnel (par exemple, la nature du poste occupé), du contexte cognitif (par exemple, la localisation spatiale des classes) et du contexte pragmatique (par exemple, le besoin d'accéder rapidement à l'information). Les résultats confirment la nécessité de concevoir des schémas de classification très peu profonds dans un contexte de gestion décentralisée des ressources informationnelles

pour répondre à la fois à des besoins d'économie cognitive (i.e. faciliter la navigation) et pragmatiques (i.e. accéder rapidement à l'information). De façon surprenante, les résultats nous apprennent que les employés s'accommodent d'une structure déséquilibrée, d'un nombre élevé de classes principales et de l'utilisation d'un schéma de classification identique pour l'organisation de l'ensemble des documents quel que soit leur support (analogiques et électroniques). De plus, les résultats indiquent que les employés ayant suivi une formation en gestion des documents sont plus enclins à concevoir et à utiliser des schémas de classification très complexes pour organiser les documents administratifs électroniques qui sont sous leur contrôle direct. Enfin, les résultats révèlent qu'un schéma de classification personnel est bien plus qu'un outil d'organisation documentaire pour les employés, puisqu'il peut être utilisé par ces derniers comme mémoire de travail (par exemple, comme rappel de l'existence d'une documentation sur support papier sur un même sujet) ou comme outil de planification de tâches (par exemple, pour l'évaluation d'un document) ou d'activités à venir (par exemple, pour l'évaluation des étudiants).

Logique

Les données qualitatives relatives à la dimension logique de chaque schéma de classification personnel furent établies par examen et interprétation manuels. Deux sous-ensembles de données ont été collectés et analysés.

Le premier sous-ensemble de données décrit les critères de division utilisés aux niveaux supérieurs de la hiérarchie. L'analyse des concepts représentés dans la dénomination des classes des 21 schémas de classification a révélé sept critères de division utilisés par les employés pour organiser les documents électroniques : *Activité* («Communications»), *Thème* («Plagiat»), *Type de contenu* («Rapport»), *Projet* («Révision Programme»), *Temps* («Aut2003»), *Organisme décisionnel ou consultatif* («Comité de gestion») et *Mélange* (désigne une division qui ne permet pas d'identifier clairement le critère de division logique prépondérant) («PV comité MScPhD»). L'étude des caractéristiques logiques des schémas de classification observés montre qu'aucun critère de division logique ne prédomine de manière significative aux premier et deuxième niveaux hiérarchiques. Cela dit, l'analyse conceptuelle a mis en évidence une classification qui privilégie aux niveaux supérieurs soit un regroupement par *thème* (par exemple, «Bourses», «Postes ouverts», «Programmes d'études») (47.10% des schémas de classification, n=10), soit un regroupement qui *mélange* ou combine plusieurs concepts (par exemple, «Listes et Notes 2003», «Internat HSC H06», «Lettres étudiants grève») (28.57%, n=6).

Un deuxième sous-ensemble de données détermine l'ordre de succession des classes principales. L'ordre privilégié par l'ensemble des employés interrogés est l'ordre alphabétique (80.95% des schémas de classification, n=17). L'ordre numérique a été observé dans seulement quatre schémas de classification de documents électroniques reprenant la structure du schéma de classification utilisé pour organiser les documents papier.

À la lumière de ces résultats, le non-respect du principe de division logique semble être la règle dans un contexte d'organisation personnelle des ressources informationnelles. La variété des critères de division logique utilisés par les employés

pour organiser les documents administratifs électroniques qui sont sous leur contrôle direct peut être expliquée par différents facteurs professionnels, cognitifs et pragmatiques tels que la nature des tâches à réaliser (et les besoins documentaires qui y sont rattachés), le manque de temps, la difficulté cognitive à catégoriser et le caractère partagé ou non d'un schéma de classification personnel. L'ordre de succession alphabétique des classes proposé par défaut dans l'environnement Windows semble satisfaire les employés. Cependant, cet ordonnancement alphabétique ne permet pas de contrôler l'ordre de présentation de classes souvent nommées différemment d'un individu à l'autre. Or, selon plusieurs employés ayant été amenés à effectuer de fréquents remplacements, une présentation uniforme et familière des classes principales des schémas de classification personnels en faciliterait la compréhension et donc l'utilisation.

Sémantique

L'analyse sémantique nous a permis d'examiner trois types de données : la longueur moyenne (en nombre de caractères) des noms de classes, les stratégies d'abréviation utilisées par les employés et le degré de redondance structurelle (répétition d'une sous-chaîne de caractères) et conceptuelle (répétition d'un même concept).

Les schémas de classification personnels observés se caractérisent sur le plan sémantique par des classes ayant des intitulés composés d'un nombre moyen de caractères relativement homogène d'un schéma à un autre (i.e., 16 caractères, écart-type=1.85). Les résultats montrent que la longueur moyenne (en nombre de caractères) des noms de classes des schémas de classification observés pourrait varier selon le caractère partagé ou non d'un schéma de classification. Les schémas de classification non partagés observés ont généralement des noms plus courts (moyenne de 15.13 caractères) que les schémas de classification partagés (moyenne de 18.16 caractères). On note l'usage privilégié des acronymes dans les stratégies d'abréviation des techniciennes. En ce qui concerne les coordonnateurs de stage, les stratégies d'abréviation privilégiées sont différentes selon que le schéma de classification est partagé ou non. Ainsi, l'usage des acronymes est également répandu pour les schémas de classification non partagés des coordonnateurs de stage alors que l'usage de codes (par exemple, le sigle d'un cours) est plus fréquent dans leurs schémas de classification partagés. De plus, on observe une proportion moyenne de classes non abrégées faible pour les schémas de classification partagés des techniciennes (48%, n=5) alors qu'inversement, le partage d'un schéma de classification d'un coordonnateur de stage ne semble pas diminuer cette proportion (68.5%, n=2). Les résultats montrent que la proportion de classes redondantes structurellement et conceptuellement varie selon le caractère partagé ou non d'un schéma de classification. Les schémas de classification non partagés observés ont généralement un degré de redondance structurelle et conceptuelle moins élevé (14.21% de l'ensemble des classes sont redondantes en moyenne) que les schémas de classification partagés (31.84% de l'ensemble des classes sont redondantes en moyenne).

Les employés qui utilisent des schémas de classification personnels partagés avec des collègues de travail ou qui ont été amenés à effectuer fréquemment des remplacements ont dit être soucieux et attentifs à créer des classes dont l'intitulé soit compréhensible par autrui. Les classes « significatives » se caractérisent par des intitulés généralement plus longs et spécifiques (i.e. d'un degré de redondance élevé) que

ceux utilisés pour désigner les classes qui ne sont pas destinées à être communiquées ou partagées. Par ailleurs, dans un contexte de travail exécuté rapidement et souvent interrompu, les employés peuvent utiliser un nombre important d'abréviations pour nommer plus rapidement leurs classes. La proportion de classes abrégées peut augmenter au fil du temps, au point de rendre ces classes peu explicites pour un individu non familier avec les abréviations utilisées. Il serait intéressant d'effectuer une analyse lexicale et d'identifier systématiquement les abréviations les plus fréquemment utilisées par une même catégorie d'employés pour nommer leur domaine d'activités. Existe-t-il une nomenclature personnelle qui repose sur l'emploi d'un certain nombre de désignations officielles, à l'exclusion d'autres? Suite à une telle analyse, il serait possible d'envisager la conception et l'emploi d'un référentiel commun mis en place par la direction (par exemple, une liste d'abréviations ou une liste de sigles) qui permettrait d'officialiser et de définir les acronymes, troncatures et autres expressions abrégées qui ont un sens au niveau local ou au niveau institutionnel.

Résultats – Phase 2 : Simulation du repérage

Nous rappelons que l'efficacité du repérage de documents a été mesurée à l'aide de trois variables : la proportion de documents repérés, le temps moyen requis (en secondes) pour repérer les documents et la proportion de documents repérés dès le premier essai. Nous voulions savoir s'il était possible de repérer avec la même efficacité un document organisé selon cinq schémas de classification distincts aux plans structurel, logique et sémantique ou s'il existait des différences statistiquement significatives entre les cinq schémas testés.

Schémas de classification et proportion de documents repérés (Succès)

Les résultats révèlent que les répertoires électroniques organisés à l'aide d'un schéma de classification dont la macrostructure est peu étendue, dont la logique est partiellement basée sur des classes d'activité et dont la sémantique utilise peu d'abréviations (schéma E), permettent de repérer significativement plus de documents, à la différence d'un schéma de classification personnel (schéma A) ($p = 0.032$), d'un schéma de classification à un seul niveau hiérarchique (schéma C) ($p < 0.001$) et d'un schéma dont la structure est peu étendue et dont la logique est partiellement basée sur des classes d'activité (schéma D) ($p = 0.009$). De manière surprenante, les différences observées dans la proportion moyenne de documents repérés ne sont pas significatives entre le schéma E et un schéma de classification personnel dont les classes sont définies (schéma B) ($p = 0.379$) (Voir Figure 8).

Ces résultats suggèrent que l'utilisation d'un schéma de classification personnel qui comprendrait une définition des classes et sous-classes ou qu'une « visite guidée du schéma » avant le départ d'un employé pourrait augmenter sensiblement le succès du repérage de documents classés par quelqu'un d'autre. En l'absence de définition des classes (schéma A), le même schéma ne permet qu'à 48.6% des répondants de retrouver tous les documents. Quand le même schéma de classification personnel est « expliqué », on peut s'attendre à ce qu'un individu autre que le propriétaire du schéma soit capable de repérer un document dans 52.9% des cas, même si la logique et la structure demeurent très personnelles.

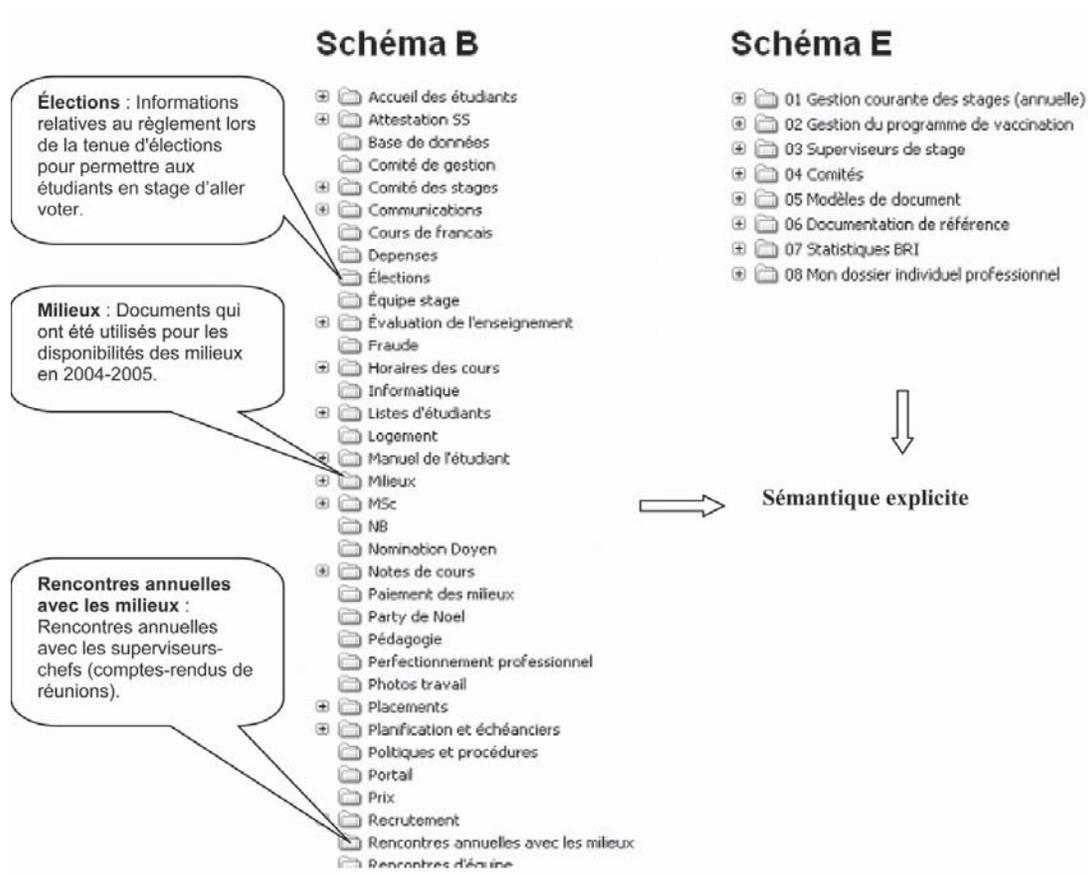


Figure 8. Schémas de classification qui obtiennent une proportion significativement plus élevée de documents repérés (Influence de la sémantique sur le succès du repérage)

Schémas de classification et temps moyen (en secondes) requis pour le repérage (Temps)

Les résultats indiquent qu'il est plus rapide de retrouver les documents en utilisant le schéma de classification dont la macrostructure est peu étendue, la logique est partiellement basée sur des classes d'activité et la sémantique utilise peu d'abréviations (i.e., le schéma E). Les résultats révèlent que la différence entre les temps moyens de réussite est significative entre le schéma E et les schémas A ($p = 0.002$), B ($p < 0.001$) et C ($p = 0.037$). Cette différence n'est toutefois pas significative entre le schéma E et le schéma D ($p = 1$) (schéma de classification personnel dont la structure est peu étendue et la logique est partiellement basée sur des classes d'activité) (Voir Figure 9).

Ces résultats suggèrent qu'un schéma de classification dont la macrostructure est peu étendue, dont la logique est partiellement basée sur des classes d'activité et dont la sémantique utilise peu d'abréviations, permet de repérer les documents aussi rapidement qu'un schéma de classification présentant les mêmes caractéristiques structurelles et logiques tout en ayant des caractéristiques sémantiques différentes (par exemple, une proportion plus élevée de classes abrégées).

Ces résultats nous montrent l'importance d'une macrostructure peu étendue (i.e. peu développée au niveau horizontal) basée, par exemple, sur un regroupement

Schéma D

- 01 Stages
- 02 Vaccination
- 03 S.S
- 04 Comités
- 05 Outil de communication
- 06 Documentation de référence
- 07 Stats
- 08 Julie

Schéma E

- 01 Gestion courante des stages (annuelle)
- 02 Gestion du programme de vaccination
- 03 Superviseurs de stage
- 04 Comités
- 05 Modèles de document
- 06 Documentation de référence
- 07 Statistiques BRI
- 08 Mon dossier individuel professionnel

Figure 9. Schémas de classification qui permettent de manière significative de repérer plus rapidement les documents (Influence de la structure et de la logique sur la rapidité du repérage).

des documents par type de contenu ou par activité pour accélérer le repérage des documents (schéma E et schéma D). L'importance de reconnaître la logique des langages classificatoires qui permet une exploration méthodique et un repérage plus rapide a déjà été soulignée dans les écrits (Hunter 2000). Ce qui est plus surprenant est que la présence de ces deux caractéristiques structurelles et logiques particulières pourrait accélérer non seulement le processus de lecture et de discrimination des classes, mais offrirait également l'avantage de permettre aux répondants de procéder par élimination. Les hésitations entre le choix des classes et sous-classes pourraient être plus rares, même si la signification de certaines classes peut demeurer ambiguë. Ces assertions pourraient être vérifiées dans une recherche ultérieure au moyen d'un protocole d'analyse verbale.

Schémas de classification et proportion de documents repérés dès le premier essai (Essai)

Les résultats montrent que l'utilisation d'un schéma de classification dont la macrostructure est peu étendue, dont la logique est partiellement basée sur des classes d'activité et dont la sémantique utilise peu d'abréviations (i.e. le schéma E) permet de localiser dès le premier essai une proportion significativement plus élevée (79%) de documents (Voir Figure 10). Les résultats révèlent que la différence entre les temps moyens des réussites est significative entre le schéma E et tous les autres schémas testés (schémas A, B, C et D) ($p < 0.001$).

Certaines améliorations des caractéristiques structurelles, logiques et sémantiques peuvent cependant être apportées au schéma E en regard des commentaires des répondants. Par exemple, plusieurs répondants ont qualifié l'ordre numérique de succession des classes du schéma E de «mélangeant». D'autres répondants considèrent que le nombre de classes de premier niveau est insuffisant (i.e. macrostructure trop «étroite») et oblige à «descendre dans la hiérarchie» (i.e. structure trop profonde). Enfin, la dénomination de certaines classes demeure ambiguë (par exemple, «Gestion courante des stages (annuelles)» qui pourrait laisser croire qu'il s'agit de documents relatifs uniquement à l'année en cours.

Schéma E

- ⊕ 01 Gestion courante des stages (annuelle)
- ⊕ 02 Gestion du programme de vaccination
- ⊕ 03 Superviseurs de stage
- ⊕ 04 Comités
- ⊕ 05 Modèles de document
- ⊕ 06 Documentation de référence
- ⊕ 07 Statistiques BRI
- ⊕ 08 Mon dossier individuel professionnel



Macrostructure peu étendue
Logique basée partiellement sur une division par classes d'activité
Sémantique utilise peu d'abréviations

Figure 10. Schéma de classification qui obtient une proportion significativement plus élevée de documents repérés dès le premier essai (Influence de la structure, de la logique et de la sémantique sur le succès du repérage dès le premier essai)

CONCLUSION

Livrés à eux-mêmes, les employés ont des comportements classificatoires que l'on peut qualifier à la fois de créatifs et de réguliers. La créativité se manifeste par l'organisation des documents sur la base de critères de division variés et par une dénomination des classes très personnelle. La régularité se manifeste à travers plusieurs traits structurels, logiques et sémantiques observables dans plusieurs schémas de classification personnels. Par exemple, les employés conçoivent et utilisent des structures classificatoires beaucoup moins profondes que celles des schémas de classification institutionnels, ils s'accommodent d'un ordonnancement alphabétique plutôt que numérique des classes et sont davantage enclins à utiliser des classes aux noms abrégés. Le schéma de classification personnel «type» est généralement «déséquilibré» et sa complexité varie selon la nature du poste occupé. On observe également un degré de redondance structurelle et conceptuelle élevé, tant au niveau de la macrostructure que de la microstructure, ce qui est caractéristique d'un usage de classes très spécifiques. L'identification de ces traits communs suggère un certain écart entre les caractéristiques des schémas de classification personnels et les caractéristiques des schémas de classifications institutionnels. Ce décalage permet de mieux comprendre la réticence de certains employés à vouloir utiliser les schémas de classification institutionnels pour l'organisation des documents administratifs électroniques qui sont sous leur contrôle direct.

Malgré le non-respect des principes classificatoires observé dans un contexte d'organisation personnelle des documents, il s'est avéré possible pour une tierce personne de repérer des documents, dans un environnement contrôlé à l'aide d'un schéma de classification personnel, même si l'efficacité du repérage n'est pas optimale. Notre recherche a permis d'identifier les caractéristiques qui influencent le succès, le temps et le risque d'erreur au plan du repérage. En effet, l'analyse des résultats de la simulation du repérage de documents a permis de confirmer que les schémas de classification ne sont pas tous égaux sur le plan de l'efficacité du repérage d'un ensemble de documents ciblés. Nous avons pu observer que l'usage d'une sémantique explicite (par une définition des classes ou une faible proportion de noms de classes abrégés) s'est révélé particulièrement influent sur le succès du repérage, quelles que soient les caractéristiques structurelles et logiques d'un schéma de classification. Une macrostructure parcimonieuse sur le plan de l'étendue (i.e. moins d'une dizaine de classes principales) et un regroupement logique et cohérent des classes basé partiellement sur des classes d'activités apparaissent suffisantes pour accélérer le repérage de documents ciblés. Enfin, un schéma de classification dont la macrostructure est peu étendue, où la logique est basée partiellement sur une division par classes d'activités, et dont la sémantique utilise peu d'abréviations, minimise de manière significative le risque d'erreur au repérage.

La nature exploratoire de notre recherche oblige cependant à prendre des précautions quant à la généralisation statistique de ces résultats à d'autres contextes en raison de limites liées au mode d'échantillonnage des schémas de classification personnels, au mode de sélection des répondants et à la taille de l'échantillon des schémas de classification testés. La simulation du repérage de documents a été réalisée avec cinq schémas de classification possédant des caractéristiques structurelles, logiques et sémantiques particulières. Les conclusions de cette recherche de nature exploratoire sont valides uniquement pour les schémas de classification qui ont été inclus dans cette recherche. Il serait toutefois pertinent de comparer, dans des conditions expérimentales identiques, l'efficacité du repérage obtenu à l'aide d'un échantillon de schémas de classification possédant des caractéristiques structurelles, logiques et sémantiques différentes de celles déjà étudiées dans notre recherche. Afin de développer une connaissance plus généralisée (dans le sens de généralisation statistique) des caractéristiques des schémas de classification personnels des documents électroniques, une étude similaire pourrait être réalisée avec d'autres types d'employés (par exemple, des gestionnaires) ou élargie à d'autres milieux de travail. En effet, il semble opportun de poursuivre la recherche sur la classification des documents administratifs électroniques réalisée jour après jour par de nombreux employés en vue de coupler le point de vue de l'archiviste avec celui de l'individu au travail. En approfondissant l'étude de l'activité classificatoire personnelle réalisée dans les bureaux, il sera alors possible de proposer des schémas de classification susceptibles non seulement de convenir à l'acteur au travail, mais aussi d'être compréhensibles et donc exploitables par autrui.

Remerciements

Nous tenons à remercier l'équipe de la Division des archives de l'Université de Montréal pour nous avoir gracieusement offert leur expertise et leur collaboration tout au long de ce projet. Pour le travail d'analyse et de vérification des données,

merci à l'équipe du Service de consultation en méthodes quantitatives de l'Université de Montréal. Nous remercions également Daniel J. Gonçalves, et Joaquim A. Jorge, professeurs à l'Instituto Superior Técnico de Lisbonne, pour nous avoir autorisées à utiliser leur logiciel pour la collecte des données. Enfin, mille mercis aux techniciennes à la coordination de travail de bureau, aux coordonnateurs et coordonnatrices de stages de l'Université de Montréal, ainsi qu'aux étudiants et aux diplômés de l'EBSI, de Psychologie et de Mathématiques qui nous ont accordé leur confiance et fait don de leur temps en acceptant volontairement de participer à cette recherche.

Sabine Mas

Stagiaire postdoctoral au laboratoire Technologie de la Coopération pour l'Innovation et le Changement Organisationnel (Tech-CICO) de l'Université de technologie de Troyes

Louise Gagnon-Arguin

Professeure associée à l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal

NOTES

1. Plusieurs études se sont intéressées à la façon dont les individus organisent leurs documents papier (Cole 1982; Malone 1983; Suchman et Wynn 1984; Lansdale 1988; Kwasnik 1989; Blomberg, Suchman et Trigg 1994; Whittaker et Hirschberg 2001), leurs courriels (Whittaker et Sidner 1996), leurs pages web (Abrams, Baecker et Chignell 1998; Jones, Dumais et Bruce 2002) et leurs documents électroniques (Carroll 1982; Kidd 1994; Barreau 1995; Barreau et Nardi 1995; Kaptelinin 1996; Chapman 1999; Boardman 2001; Ackerman *et al.* 2003; Henderson 2004).
2. De nombreux termes sont significatifs seulement lorsqu'ils sont combinés avec des termes plus larges dans les niveaux plus élevés de la hiérarchie (Lin et Chan 1999). Une fois pris hors de la hiérarchie, un terme peut ne pas toujours porter le contexte et, par conséquent, il peut ne pas être très efficace comme terme de repérage (Hertzum et Frøkjær 1996).
3. Par exemple, le contrôle des synonymes et homonymes.
4. L'ambiguïté des classes ne peut être toutefois totalement évitée selon Hertzum (2004).
5. Une communauté de discours est un groupe de personnes liées par une occupation ou des intérêts spéciaux et qui partagent un ensemble de connaissances, de possessions et de comportements (Bergquist et Ljungberg 1999).
6. Nous pensons que le repérage *manuel* pourrait également être qualifié de *sensoriel* puisque dans le cas de l'accès par navigation à un document particulier, l'œil parcourt les niveaux avec plus ou moins de rapidité afin d'y identifier des signes discriminants qui permettent de s'arrêter sur le choix d'un répertoire plutôt qu'un autre.
7. La communauté de chercheurs du domaine de recherche d'information s'intéresse à concevoir des systèmes automatiques d'indexation et de recherche de documents. Tout système proposé est testé expérimentalement afin de connaître son effet sur la réalité.
8. Au fur et à mesure que les répondants se présentaient à la simulation, ils furent assignés séquentiellement à un des 14 groupes, à l'occurrence de 5 répondants par groupe.
9. L'analyse de variance a permis d'analyser et de comparer, pour chaque mesure, les moyennes obtenues entre les schémas de classification.
10. Le test de Bonferroni est un calcul statistique complémentaire à l'analyse de variance qui a permis de déterminer si les différences observées sont statistiquement significatives, autrement dit de vérifier si les différences sont suffisamment grandes (quand $p < 0.05$) pour les attribuer simplement au hasard ou à une fluctuation de l'échantillonnage.

BIBLIOGRAPHIE

- ABRAMS, D., R. BAECKER et M. CHIGNELL. 1998. Information archiving with bookmarks : Personal web space construction and organization. In *Proceedings of the SIGCHI conference on Human factors in computing system.*, April 18-23, 1998, Los Angeles, California, United States. New York, ACM Press : 41-48.
- ADAR, Eytan, David KARGER et Lynn Andrea STEIN. 1999. Haystack : Per-user information environment. In *Proceedings of the 8th International Conference on Information and Knowledge Management*. November 2-6, 1999, Kansas City, Missouri, United States. [En ligne]. <http://www.hpl.hp.com/research/idl/people/eytan/p413-adar.pdf>
- ALVARADO, Christine *et al.* 2003. *Surviving the information explosion : How people find their electronic information*. Cambridge, Massachusetts, Massachusetts Institute of Technology, Artificial Intelligence Laboratory. [En ligne]. <ftp://publications.ai.mit.edu/ai.mit.edu/ai-publications/2003/AIM-2003-006.pdf>
- BÄLTER, O. 1997. Strategies for organising email. In *Proceedings of Human Computer Interaction on People and Computers XII*. London, United Kingdom. London, United Kingdom, Springer-Verlag : 21-38.
- BARREAU, D. K. 1995. Context as a factor in personal information management systems. *Journal of the American Society for Information Science* 46, 5 : 327-339.
- BARREAU, D. K., et B. A. NARDI. 1995. Finding and reminding : File organization from the desktop. *SIGCHI Bulletin* 27, 3 : 39-43.
- BATES, M. 1977. Factors affecting subject catalog search success. *Journal of the American Society for Information Science* 28, 3 : 161-169.
- BEAUPRÉ, L. 2004. Classification des documents électroniques : expérience à la Sûreté du Québec. *Actes du 33ème Congrès de l'Association des archivistes du Québec*. 27-29 mai 2004, Sainte-Adèle, Québec, Canada. Québec, Association des archivistes du Québec : 54-57.
- BERCHTOLD, Stefan., Alexandros BILIRIS et Euthimios PANAGOS. 1999. SaveMe : A system for archiving electronic documents using messaging groupware. In *Proceedings of the international joint conference on work activities coordination and collaboration*. February 22-25, 1999, San Francisco, California, United States : 167-176. [En ligne]. http://www.biliris.com/alex/pubs/papers/99_saveme_wacc99.pdf
- BERGERON, P. (1992). La gestion des archives électroniques : quelques questions-clés à considérer. *Archives (Revue de l'Association des archivistes du Québec)*, 23 (3), 51-70.
- BERGMAN, O., R. BEYTH-MAROM et R. NACHMIAS. 2003. The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology* 54, 9 : 872-878.
- BERGQUIST, M. et J. LUJUNGBERG. 1999. Genre in action : Negotiating Genres in Practice. In *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences*. Washington, DC, IEEE Computer Society. [En ligne]. <http://ieeexplore.ieee.org/iel5/6293/16782/00772655.pdf>

- BERLIN, L. M. *et al.* 1993. Where did you put it? Issues in the design and use of a group memory. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. April 24-29, 1993, Amsterdam, The Netherlands. Amsterdam, The Netherlands, IOS Press : 23-30.
- BLOMBERG, J., L. SUCHMAN et R. TRIGG. 1994. Reflections on a work-oriented design project. In *Proceedings of the Participatory Design Conference*. October 27-28, 1994, Chapel Hill, North Carolina, United States. New York, ACM Press : 99-109.
- BOARDMAN, Richard. 2001a. Category overlaps between hierarchies in user workspace. In *Proceedings of Interact'01, the 8th IFIP TC 13 International Conference on Human-Computer Interaction*. Tokyo, Japan. <http://www.iis.ee.ic.ac.uk/~rick/research/pubs/categories-interact2001.pdf>
- BOARDMAN, Richard. 2001b. Multiple hierarchies in user workspace. In *Proceedings of Conference on Human Factors in Computing Systems*. March 31-April 5, 2001, Seattle, Washington : 403-404. [En ligne]. <http://www.iis.ee.ic.ac.uk/~rick/research/pubs/workspace-chi2001.pdf>
- BOARDMAN, Richard, et M. Angela SASSE. 2004. Stuff goes into the computer and doesn't come out : A cross-tool study of personal information management. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. April 24-29, 2004, Vienna, Austria : 583-590. <http://www.iis.ee.ic.ac.uk/~rick/research/pubs/boardman-chi04.pdf>
- BORLUND, Pia et Peter INGWERSEN. 1998. Measures of relative relevance and ranked half-life : Performance indicators for interactive IR. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. August 24-28, 1998, Melbourne, Australia : 324-331. [En ligne]. <http://portal.acm.org/citation.cfm?id=291019>
- BOWKER, G. C., et S. L. STAR. 2000. *Sorting things out : Classification and its consequences*. Cambridge, Massachusetts. MIT Press.
- CANONNE, A. 1993. *Vocabulaire élémentaire des classifications*. Coll. «Bibliothèque du bibliothécaire». Liège, Belgique, Éditions du CÉFAL.
- CARROLL, J. M. 1982. Creative names for personal files in an interactive computing environment. *International Journal of Man-Machine Studies* 16 : 405-438.
- CHAMARD, J.-P. 2004. La classification des documents électroniques : expérience du ministère de la Justice du Québec. In *Actes du 33ème Congrès de l'Association des archivistes du Québec*. 27-29 mai 2004, Sainte-Adèle, Québec, Canada. Québec, Association des archivistes du Québec : 58-62.
- CHAN, L. M. 1995. Classification, present and future. *Cataloging et Classification Quarterly* 21, 2 : 5-17.
- CHAPMAN, H. 1999. *The file naming habits of personal computer users*. A Master's paper for the M.S. in I.S. degree, University of North Carolina at Chapel Hill.
- CHU, H. 2003. *Information representation and retrieval in the digital age*. Medford, New Jersey, American Society for Information Science and Technology by Information Today.

- CLEVERDON, C. 1967. The Cranfield tests on index language devices. *ASLIB Proceedings* 19,6 : 173-194.
- COCHRANE, P. A. 1985. Preparing for the use of classification in online computer cataloging systems and in online catalogs. *Information Technology and Libraries* 4 : 91-111.
- COLE, I. 1982. Human aspects of office filing : Implications for the electronic office. In *Proceedings of the Human Factors Society 26th Annual Meeting*. Santa Monica, California, United States. Santa Monica, California, Human Factors Society : 59-63.
- CONFÉRENCE DES RECTEURS ET DES PRINCIPAUX DES UNIVERSITÉS DU QUÉBEC (CRÉPUQ). 1994. *La gestion des archives informatiques*. Sainte-Foy, Presses de l'Université du Québec.
- CONFÉRENCE DES RECTEURS ET DES PRINCIPAUX DES UNIVERSITÉS DU QUÉBEC (CRÉPUQ). 2004. *La gestion des documents numériques des établissements universitaires du Québec : état de situation et planification stratégique*. [En ligne]. <http://www.crepuq.qc.ca/documents/arch/Rapport-GGDN.htm>
- COUTURE, Carol, et Jean-Yves ROUSSEAU. 1982. Les archives au XXème siècle : une réponse aux besoins de l'administration et de la recherche. Montréal, Université de Montréal, Secrétariat général.
- DHÉRENT, Catherine et collaborateurs. 2002. *Les archives électroniques : manuel pratique*. [Paris], Direction des archives de France. [En ligne]. <http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/DAFmanuel%20version%207.html>
- DUMAIS, Susan T., et Thomas K. LANDAUER. 1983. Using examples to describe categories. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. December 12-15, 1983, Boston, Massachusetts, United States : 112-115. [En ligne]. <http://portal.acm.org/citation.cfm?id=800045.801592>
- DUMAIS, Susan T. *et al.* 2003. Stuff I've seen : A system for personal information retrieval and re-use. In *Proceedings of the 26th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*. July 28-August 1, 2003, Toronto, Canada : 72-79. [En ligne]. <http://portal.acm.org/citation.cfm?id=860435.860451>
- FALLIS, D., et K. MATHIESEN. 2000. Consistency rules for classification schemes (or how to organize your beanie babies). In *Proceedings of the 6th International ISKO Conference*. July 10-13, 2000, Toronto, Canada. Würzburg, Germany, Ergon : 339-344.
- FARRELL, Stephen *et al.* 2002. Information programming for personal user interfaces. In *Proceedings of the 7th international conference on intelligent user interfaces*. January 13-16, 2002, San Francisco, California, United States : 190-191. [En ligne]. <http://portal.acm.org/citation.cfm?id=502751>
- FERRAIOLI, L. 2005. An exploratory study of metadata creation in a health care agency. *Cataloging et Classification Quarterly* 40, 3-4 : 75-102.

- GAGNON-ARGUIN, Louise. 1999. La création. In COUTURE, Carol et collaborateurs. *Les fonctions de l'archivistique contemporaine*. Sainte-Foy, Presses de l'Université du Québec : 69-101.
- GAGNON-ARGUIN, Louise et Rabii BANNOURI. 1998. *La classification des documents institutionnels dans les ministères du gouvernement du Québec et le défi informatique*. Rapport de recherche. Montréal, Université de Montréal, École de bibliothéconomie et des sciences de l'information. [En ligne]. <http://www.msg.gouv.qc.ca/fr/publications/enligne/administration/ingenierie/classification.pdf>
- GONÇALVES, D. J. et J. A. JORGE 2003. An empirical study of personal document spaces. In *Proceedings DSV-IS'03, June 4-6, 2003*. Funchal, Madeira Island, Portugal : 46-60. [En ligne]. http://immi.inesc-id.pt/publication.php?publication_id=34
- HEDSTROM, M. 1996. Electronic records research : What have archivists learned from the mistakes of the past? *Archives and Museum Informatics* 10, 4 : 313-325.
- HENDERSON, S. 2003. *Information workspaces : Investigating the information behaviour of knowledge workers and its implications for the design of usable information workspaces*. PhD Dissertation Proposal, University of Auckland.
- HENDERSON, S. 2004. Personal digital document management. In *6th Asia-Pacific Conference on Computer-Human Interaction*. June 29-July 2, 2004, Rotorua, New Zealand. Berlin, Springer : 651-655.
- HÉON, Gilles. 1995. Les plans de classification en archivistique. *Archives* 27, 1 : 73-89.
- HÉON, Gilles. 1999. La classification. In COUTURE, Carol et collaborateurs. *Les fonctions de l'archivistique contemporaine*. Sainte-Foy, Presses de l'Université du Québec : 219-254.
- HERTZUM, Morten. 1993. Information retrieval in a work setting : A case study of the documentation part of chemists' work. In *Proceedings of the 16th IRIS. Information Systems Research Seminar in Scandinavia*. August 7-10, 1993, Copenhagen, Denmark : 786-798. [En ligne]. <http://akira.ruc.dk/~mhz/Research/Publ/IRIS1993.pdf>
- HERTZUM, Morten. 1999. Six roles of documents in professionals' work. In *Proceedings of the 6th European Conference on Computer Supported Cooperative Work*. September 12-16, 1999, Copenhagen, Denmark : 41-60. [En ligne]. <http://www.dcs.qmw.ac.uk/~mounia/PROJECTS/DocRoles.doc>
- HERTZUM, Morten. 2004. Small-scale classification schemes : A field study of requirements engineering. *Computer Supported Cooperative Work* 13, 1 : 35-61. http://akira.ruc.dk/~mhz/Research/Publ/JCSCW2004_preprint.pdf
- HERTZUM, Morten et Erik Frøkjær. 1996. Browsing and querying in online documentation : A study of user interfaces and the interaction process. *ACM Transactions on Computer-Human Interaction* 3, 2 : 136-161. http://akira.ruc.dk/~mhz/Research/Publ/TOCHI1996_preprint.pdf
- HILDRETH, C. R. 2000. Are Web-based OPACs more effective retrieval systems than their conventional predecessors? An experimental study. In *Proceedings of the*

- 6th International ISKO Conference*. July 10-13, 2000, Toronto, Canada. Würzburg, Germany, Ergon : 237-246.
- HUDON, Michèle. 2001. Structuration du savoir et organisation des collections dans les répertoires du Web. *Bulletin des bibliothèques de France* 46, 1 : 57-62.
- HUNTER, E. J. 2000. Do we still need classification? In MARCELLA, R., et MALTBY, A. (eds.). *The future of classification*. London, United Kingdom, Gower : 1-17.
- IYER, H. 1995. *Classificatory structures : Concepts, relations and representation*. Frankfurt/Main, Indeks. (Textbooks for knowledge organization ; v. 2).
- JACOB, E. K. 2004. Classification and categorization : A difference that makes a difference. *Library Trends* 52, 3 : 515-540.
- JONES, W. P. 2004. Finders, keepers? The present and future perfect in support of personal information management. *First Monday* 9. 3. [En ligne]. http://www.firstmonday.org/issues/issue9_3/jones/
- JONES, W. P. 2007. Personal information management. *Annual Review of Information Science and Technology* 41 : 453-504.
- JONES, W. P., et D. MAIER. 2003. Personal information management group report. In *National Science Foundation (NSF) Information and Data Management (IDM) Workshop*. September 14-16, 2003, Seattle, Washington, United States. [En ligne]. http://kftf.ischool.washington.edu/docs/Summary_of_PIM2003.pdf
- JONES, W. P. *et al.* 2005. Don't take my folders away! : Organizing personal information to get things done. In *CHI'05 extended abstracts on Human factors in computing systems*. April 2-7, 2005, Portland, Oregon, United States : 1505-1508. [En ligne]. <http://portal.acm.org/citation.cfm?id=1056952&dl=GUIDE&coll=GUIDE&CFID=48509855&CFTOKEN=8097759>
- JONES, W. P., S. T. DUMAIS et H. BRUCE. 2002. Once found, what then? : A study of «keeping» behaviours in the personal use of web information. In *Proceedings of the 65th ASIST Annual Meeting*. Philadelphia, United States : 391-402. [En ligne]. <http://kftf.ischool.washington.edu/docs/ASIST2002.pdf>
- KAPTELININ, Victor. 1996. Creating computer-based work environments : An empirical study of Macintosh users. In *Proceedings of the 1996 ACM SIGCPR/SIGMIS Conference on Computer personnel research*. April 1996, Denver, Colorado, United States : 360-366. [En ligne]. <http://portal.acm.org/citation.cfm?id=238921>
- KARJALAINEN, Anne *et al.* 2000. Genre-based metadata for enterprise document management. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*. January 4-7, 2000, Island of Maui, Hawaii. [En ligne]. <http://csdl.computer.org/comp/proceedings/hicss/2000/0493/03/04933013.pdf>
- KHOO, Christopher S. G. *et al.* 2007. How users organize electronic files on their workstations in the office environment : A preliminary study of personal information organization behaviour. *Information Research* 12, 2. [En ligne]. <http://InformationR.net/ir/12-2/paper293.html>
- KIDD, Alison. 1994. The marks are on the knowledge worker. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. April 24-28, 1994,

- Boston, Massachusetts, United States : 186-191. [En ligne]. <http://portal.acm.org/citation.cfm?id=191740>
- KRIKELAS, J. 1969. Subject searches using two catalogs : A comparative evaluation. *College and Research Libraries* 30, 6 : 506-517.
- KWASNIK, B. H. 1989. How a personal document's intended use or purpose affects its classification in an office. In *Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Cambridge, Massachusetts, United States : 207-210. [En ligne]. <http://portal.acm.org/citation.cfm?id=75334.75356>
- LANCASTER, F. W. 1968. *Information retrieval systems characteristics, testing, and evaluation*. New York, Wiley.
- LANGRIDGE, D. W. 1992. *Classification : Its kinds, elements, systems, and applications*. London, United Kingdom, Bowker-Saur.
- LANSDALE, M. 1988. The psychology of personal information management. *Applied Ergonomics* 19, 1 : 55-66.
- LANSDALE, M. et E. EDMONDS. 1992. Using memory for events in the design of personal filing systems. *International Journal of Man-Machine Studies* 36 : 97-126.
- LIN, X., et L. M. CHAN. 1999. Personalized knowledge organization and access for the Web. *Library et Information Science Research* 21, 2 : 153-172.
- LYTLE, R. H. 1980. Intellectual access to archives : II. Report of an experiment comparing provenance and content indexing methods of subject retrieval. *The American Archivist* 43 : 191-205.
- MALONE, T. W. 1983. How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems* 1, 1 : 99-112.
- MANIEZ, J. 1987. *Les langages documentaires et classificatoires : conception, construction et utilisation dans les systèmes documentaires*. Paris, Éditions d'Organisation.
- MANIEZ, J. 1991. A decade of research in classification. *International Classification* 18, 2 : 73-77.
- MARCELLA, R., et C. NEWTON. 1994. The functions of classification in the library. In MARCELLA, R. et C. NEWTON. *A new manual of classification*. Aldershot, England, Gower : 159-181.
- MAS, Sabine. 2007. *Schémas de classification et repérage des documents administratifs électroniques dans un contexte de gestion décentralisée des ressources informationnelles*. Thèse de doctorat, Université de Montréal. [En ligne]. <http://hdl.handle.net/1866/1433>
- METZGER, J.-P. et G. LALLICH-BOIDIN. 2004. Temps et documents numériques. *Document numérique* 8, 4 : 11-21.
- MOLHOLT, P. 1995. Qualities of classification schemes for the information superhighway. *Cataloging and Classification Quarterly* 21, 2 : 19-22.
- NARDI, Bonnie et Deborah BARREAU. 1997. «Finding and reminding» revisited : Appropriate metaphors for file organization at the desktop. *SIGCHI* 29, 1 : 76-78. [En ligne]. <http://sigchi.org/bulletin/1997.1/nardi.html>

- O'DAY, Vicki. L. et Robin JEFFRIES. 1993. Orienteering in an information landscape : How information seekers get from here to there. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. April 24-29, 1993, Amsterdam, The Netherlands : 438-445. [En ligne]. <http://portal.acm.org/citation.cfm?id=169365>
- ORLIKOWSKI, W. et J. YATES. 1994. Genre repertoire : The structuring of communicative practices in organizations. *Administrative Sciences Quarterly* 33, 4 : 541-574.
- PATTERSON, G. et T. J. SPREHE. 2002. Principal challenges facing electronic records management in federal agencies today. *Government Information Quarterly* 19 : 307-315.
- PÉROTIN, Y. 1961. L'administration et les trois âges des archives. *Seine-et-Paris* 20 : 1-14.
- RAVASIO, Pamela, Sissel GUTTORMSEN SCHAR et Helmut KRUEGER. 2004. In pursuit of desktop evolution : User problems and practices with modern desktop systems. *ACM Transactions on Computer-Human Interaction* 11, 2 : 156-180. [En ligne]. <http://portal.acm.org/citation.cfm?id=1005363>
- RENAUD, J. 1975. Le traitement et la conservation des documents. *Archives* 75,1 :13-23.
- RIBEIRO, F. 1996. Subject indexing and authority control in archives : The need for subject indexing in archives and for an indexing policy using controlled language. *Journal of the Society of Archivists* 17, 1 : 27-54.
- RICHMOND, P. A. 1990. General theory of classification. In BENGSTON, B. et J.S HILL. *Classification of library materials : Current and future potential for providing access*. New York, NY : Neal Schuman, 16-26.
- ROBERGE, Michel. 1985. *La classification universelle des documents administratifs*. La Pocatière, Documentor.
- ROBERGE, Michel. 2002. *L'essentiel de la gestion documentaire*. Québec, GESTAR.
- ROBERGE, Michel. 2004. *L'essentiel de la gestion documentaire*. 2ème édition revue et augmentée. Québec, GESTAR.
- ROUSSEAU, Jean-Yves. 1980. Le choix d'un système de classement et son application. *Archives* 11, 1 : 3-19.
- SABOURIN, P. 2001. Constructing a function-based records classification system : Business activity structure classification system. *Archivaria* 51 : 137-154.
- SAFFADY, W. 2002. *Managing electronic records*. Prairie Village, Kansas, ARMA International.
- SALTON, G. 1989. *Automatic text processing : The transformation, analysis, and retrieval of information by computer*. Reading, Massachusetts, Addison-Wesley.
- SELLEN, A. et R. HARPER. 2001. *The myth of the paperless office*. Cambridge, Massachusetts, MIT Press.
- SENÉCAL, Sylvain. 1997-1998. La lecture et la description archivistique du document. *Archives* 29, 3-4 : 49-56.

- SENÉCAL, Sylvain. 1999. *La notion de classification et l'utilisation d'un plan de classification pour la gestion des documents dans un environnement électronique*. Montréal, Hydro-Québec.
- SKUPSKY, D. S. 1999. Applying records retention to electronic records. *The Information Management Journal* 33, 3 : 28-35.
- SPREHE, T. J., C. R. McCLURE et P. ZELLNER. 2002. The role of situational factors in managing U.S. federal recordkeeping. *Government Information Quarterly* 19 : 289-305.
- SUCHMAN, L. et E. WYNN. 1984. Procedures and problems in the office. *Office Technology and People* 2 : 134-154.
- SVENONIUS, E. 1983. Use of classification in online retrieval. *Library Resources and Technical Services* 27, 1 : 76-80.
- TOMS, Elaine G. et D. GRANT CAMPBELL. 1999. Utilizing information «shape» as an interface metaphor based on genre. In *Proceedings of the 27th Annual Conference of the Canadian Association for Information Science*. June 9-11, 1999, Université de Sherbrooke, Canada : 370-386. [En ligne]. http://www.cais-accsi.ca/proceedings/1999/Toms_1999.pdf
- VAN DER WALT, M. S. 1998. The structures of classification schemes used in Internet search engines. In *Proceedings of the 5th International ISKO Conference*. August 25-29, 1998, Lille, France. Würzburg, Germany, Ergon : 379-387.
- WHITTAKER, Steve et Julia HIRSCHBERG. 2001. The character, value and management of personal paper archives. *ACM Transactions on Computer Human Interaction* 8, 2 : 150-170. <http://portal.acm.org/citation.cfm?id=376932>
- WHITTAKER, Steve et Candace SIDNER. 1996. Email overload : Exploring personal information management of email. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. April 13-18, 1996, Vancouver, British Columbia, Canada : 276-283. [En ligne]. <http://dis.shef.ac.uk/stevewhittaker/emlch96.pdf>
- WHITTAKER, S., L. TERVEEN et B. NARDI. 2000. Let's stop pushing the envelope et start addressing it : A reference task agenda for HCI. *Human Computer Interaction* 15 : 75-106.
- ZELLER, J.-D. 2004. Documents numériques : à la recherche d'une typologie perdue. *Document numérique* 8, 2 : 101-116.
- ZINS, C. 2002. Models for classifying Internet resources. *Knowledge Organisation* 29, 1 : 20-28.

ANNEXE 1

Principales caractéristiques structurelle, logique et sémantique des cinq schémas de classification testés

	Schéma personnel d'un employé	Schémas de classification artificiellement créés			
	Schéma A	Schéma B	Schéma C	Schéma D	Schéma E
Structure					
Étendue	Macro-structure étendue (42 classes)	Macro-structure étendue (42 classes)	Macro-structure très étendue (192 classes)	Macro-structure peu étendue (8 classes)	Macro- structure peu étendue (8 classes)
Profondeur	Peu profonde (1.56 niveaux)	Peu profonde (1.56 niveaux)	n/a	Moyennement profonde (2.62 niveaux)	Moyennement profonde (2.62 niveaux)
Complexité	192 classes	192 classes	192 classes	202 classes	202 classes
Équilibre	Peu équilibrée (Écart- type=6.90)	Peu équilibrée (Écart-type=6.90)	n/a	Équilibrée (Écart- type=2.32)	Équilibrée (Écart- type=2.32)
Logique					
Premier critère de division logique	Thème	Thème	Mélange	Type de contenu	Activité
Deuxième critère de division logique	Mélange	Mélange	n/a	Type de contenu	Type de contenu
Ordre de succession des classes	Alphabétique	Alphabétique	Alphabétique	Numérique	Numérique
Sémantique					
Nombre moyen de caractères dans l'intitulé des classes	19.12	19.12	37.69	19.33	20.47
Proportion des classes abrégées	32 %	32 %	48 %	28 %	1 %
Principales stratégies d'abréviation	Acronyme (10% des classes)	Acronyme (10% des classes)	Mélange (17% des classes)	Code (7%) et Mélange (7%)	Acronyme (1% des classes)
Proportion des classes redondantes	35.05 %	35.05 %	90 %	29.70 %	10.89 %
Définition des classes	Non	Oui	Non	Non	Non