

Schémas de classification et repérage de documents électroniques : pratiques individuelles et application des principes classificatoires archivistiques

Résumé : Présentation des résultats de recherche doctorale. Examen des caractéristiques de 21 schémas de classification utilisés par des employés pour organiser et classer leurs documents électroniques. Vérification des différences entre divers schémas quant à leur efficacité pour le repérage de documents électroniques.

Abstract: Presentation of results of a doctoral research. Analysis of the characteristics of 21 classification structures used by employees to organize and file electronic documents. Differences between various classification structures as to their efficiency for retrieving e-documents.

1. Éléments contextuels

Le schéma de classification institutionnel des documents est un mode essentiel d'organisation et de repérage des documents administratifs. Ce schéma de classification des documents institutionnels 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, offrant ainsi des éléments de contexte qui facilitent le repérage et la compréhension des documents pour l'ensemble du personnel de l'organisme.

Pour répondre au besoin d'uniformité et de contrôle dans la gestion des documents électroniques, les archivistes et gestionnaires de documents encouragent les employés à utiliser le schéma de classification institutionnel, originalement conçu pour l'organisation des documents sur support papier, pour nommer et organiser les répertoires électroniques à l'intérieur desquels se trouvent classés les documents. Dans la pratique, un nombre important de documents électroniques ayant un caractère institutionnel et qui sont gérés par les employés eux-mêmes ne sont pas classés à l'aide du schéma de classification institutionnel mais plutôt à l'aide d'une structure classificatoire très personnelle (Gagnon-Arguin et Bannouri, 1998; Skupsky, 1999; Tough et Moss, 2003).

Plusieurs raisons sont mentionnées dans la littérature pour expliquer l'individualisme qui caractérise l'organisation des documents électroniques qui sont sous le contrôle direct de l'employé. Par exemple, on note (1) l'absence de motivation, de leadership et de ressources institutionnelles, (2) la quasi-absence de politiques, normes, méthodes et procédures de gestion intégrée des documents électroniques, (3) l'absence de formation et d'encadrement des employés, (4) la « complexité inhérente » à l'utilisation du schéma de classification institutionnel (Gagnon-Arguin et Bannouri, 1998, 30), (5) l'effort cognitif dans l'acte de classer et (6) le manque de temps (Malone, 1983; Sprehe, McClure et Zellner, 2002; Henderson, 2005).

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 amène à nous poser plusieurs questions : 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 à l'aide de ces schémas personnels ? Existe-t-il des différences entre les schémas de classification utilisés quant à leur capacité à repérer les documents ? 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 pour organiser leurs documents électroniques dans un contexte de travail sont peu nombreuses. Aucune étude empirique n'a été menée à ce jour afin de vérifier dans quelle mesure les schémas de classification personnels existants permettent, ou même facilitent, le repérage des documents électroniques par des tiers, dans le cadre d'un travail collaboratif par exemple, lorsqu'il s'agit de reconstituer un dossier ou lorsque des employés quittent l'organisation.

La recherche dont nous rapportons ici brièvement les résultats avait deux objectifs spécifiques. Un premier objectif consistait à décrire les caractéristiques de différents schémas de classification utilisés pour organiser les documents électroniques dans un contexte de gestion décentralisée des ressources informationnelles. Un deuxième objectif consistait à vérifier, dans un environnement contrôlé, les différences entre les schémas de classification quant à leur efficacité pour le repérage. La réalisation de ces deux objectifs visait à augmenter nos connaissances sur les structures classificatoires individuelles en contexte de travail et à mesurer empiriquement la capacité de tierces personnes à comprendre et à repérer des documents à l'aide de structures de classification et de navigation distinctes. Ce faisant, nous souhaitions identifier les caractéristiques d'une structure arborescente de répertoires qui permettaient 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.

Un protocole de recherche en trois étapes fut réalisé pour atteindre ces objectifs. Dans ce texte, nous résumons d'abord les résultats de notre analyse de schémas de classification personnels avant de présenter les caractéristiques de schémas de classification manipulés pour des fins d'expérimentation. En terminant, nous présentons les résultats de la simulation du repérage de documents électroniques ciblés à l'aide de cinq schémas de classification présentant des caractéristiques variées.

2. Examen des schémas de classification personnels

2.1 Modèle d'analyse

Nous avons conçu une grille d'analyse qui tient compte des caractéristiques multidimensionnelles de langages classificatoires particuliers. La littérature scientifique a servi de cadre de référence pour la conception de cette grille d'analyse. Plus précisément, nous nous sommes penchées sur la théorie de l'organisation des connaissances (Maniez 1987; Iyer 1995), sur les méthodes de travail et les résultats de recherches dans les domaines de la gestion personnelle de l'information (Jones, Dumais et Bruce, 2002; Gonçalves et Jorge, 2003; Boardman et Sasse, 2004; Henderson, 2005; 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, nombre de classes principales), logiques (par exemple, critère de division privilégié) et sémantiques (par exemple, stratégies d'abréviation observées dans la dénomination des classes) des schémas de classification personnels. On trouvera une description détaillée de cette grille d'analyse dans Hudon, Mas et Gazo (2005).

2.2 Procédure

Nous avons décrit 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 (11 schémas de classification personnels dont cinq partagés avec des collègues de travail) et des coordonnateurs de stage (10 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 à la fois
- (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.

La chercheuse principale a 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. La chercheuse a 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 créée 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 permet 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). Le rapport produit par le logiciel a été remis aux employés pour les rassurer sur la nature des données collectées, et sur le fait qu'aucune information sur le contenu même des documents n'avait été colligée.

2.3 Caractéristiques des schémas de classification observés

Chaque schéma de classification personnel fut examiné à partir de trois dimensions : sa structure, sa logique et sa sémantique. Nous présentons ici un résumé des caractéristiques observées pour chacune des dimensions.

2.3.1 Structure

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

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 macro-structure généralement étendue (nombreuses classes principales) et à une micro-structure 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.

2.3.2 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 comite 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.

2.3.3 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 soit partagé ou non partagé. 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).

2.4 Conclusion

En résumé, les schémas de classification observés se caractérisent sur le plan structurel par une macro-structure généralement étendue et une micro-structure plus étroite, une structure peu profonde, complexe et déséquilibrée. Les employés privilégient un regroupement des classes par thème et un ordonnancement alphabétique. 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, par des stratégies d'abréviation qui privilégient l'usage d'acronymes et par une redondance structurelle et conceptuelle à l'intérieur des noms de classes généralement élevée dès le premier niveau de la hiérarchie.

Le deuxième objectif de notre recherche consistait à vérifier les différences entre 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 permettent 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 structures classificatoires artificielles. 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 d'un coordonnateur de stage, que nous avons jugé représentatif des différentes caractéristiques observées dans les schémas de classification personnels (Schéma A). Les caractéristiques structurelles, logiques et sémantiques de ce schéma sont décrites dans le Tableau 1. Nous avons manipulé ce schéma personnel pour créer artificiellement quatre schémas de classification présentant des variations significatives sur le plan structurel, logique et sémantique.

3. Manipulation des schémas de classification

Pour créer artificiellement quatre structures, le schéma de classification de l'employé (Schéma A) a donné lieu à quatre manipulations différentes réalisées par la chercheuse. Les quatre 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 et transcrites par la chercheuse. 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.

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 (i.e., 90 % des classes sont redondantes) et (3) la proportion de classes abrégées (48 %). 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.

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, macro-structure peu étendue, division logique basée sur les activités de

l'employé) tout en conservant la sémantique du schéma original de l'employé. Cette manipulation consistait à utiliser le schéma de classification préalablement conçu et conforme aux exigences archivistiques (i.e., Schéma E) 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.

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, macro-structure peu étendue, critère de division basé sur les activités de l'employé, peu d'abréviations). Ce schéma de classification a été validé auprès d'un archiviste de la Division des archives de l'Université de Montréal.

Le Tableau 1 résume les principales caractéristiques structurelle, logique et sémantique du schéma de classification personnel retenu pour des fins de manipulation (Schéma A) ainsi que les caractéristiques des quatre schémas de classification artificiellement créés (Schémas B, C, D et E).

Tableau 1 Caractéristiques structurelles, logiques et sémantiques des schémas de classification A, B, C, D et E

	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

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.

4. Simulation du repérage d'un ensemble de documents

4.1 Procédure

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.

Un appel à participation a été lancé en avril 2006 par le biais des listes de courriel des étudiants et des listes professionnelles. La collecte des données a été réalisée entre la fin du mois d'avril et le début du mois de juillet 2006. Chaque individu qui répondait aux critères d'inclusion a été ensuite assigné aléatoirement¹ à l'un des 14 groupes afin de diminuer l'influence que des variables étrangères à la recherche pourraient avoir sur les variables étudiées.

La chercheuse a reproduit sur un 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 par la chercheuse 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. Pour tester les hypothèses de recherche, des statistiques descriptives ont été calculées et l'analyse de variance (ANOVA) à mesures répétées a été réalisée en utilisant le logiciel d'analyse statistique en sciences sociales SPSS (« *Statistical package for social sciences* »). Le test de Bonferroni 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% ($p < 0.05$).

¹ 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.

4.2 Résultats de la simulation et discussion

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.

4.2.1 Schémas de classification et succès du repérage

Les résultats indiquent qu'une plus grande proportion de répondants (70%, n=49) retrouve tous les documents en utilisant le schéma de classification dont la structure est peu étendue, la logique est partiellement basée sur des classes d'activité et la sémantique utilise peu d'abréviations (Schéma E).

La Figure 1 décrit les différences dans la proportion moyenne de documents repérés à l'aide des divers schémas de classification. L'analyse des données empiriques indique que la proportion moyenne de documents repérés la plus élevée (91.07%) est obtenue à l'aide du schéma E alors que le schéma C est associé au taux de succès le plus faible (78.57%).

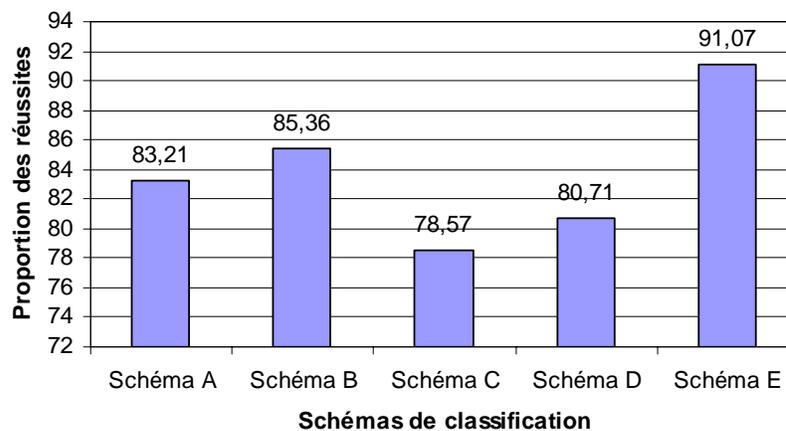


Figure 1 Proportion des documents retrouvés à l'aide des les schémas de classification.

L'analyse de variance à mesures répétées a permis de vérifier si ces différences sont significatives ou si la moyenne pour la proportion de documents repérés est la même pour chacun des schémas de classification. Les résultats nous autorisent à rejeter l'hypothèse nulle qu'en moyenne la proportion du succès pour les cinq schémas de classification est la même car la valeur p a été quantifiée très inférieure au seuil de signification de 5% ($p < 0.001$).

Les résultats révèlent que les répertoires électroniques organisés à l'aide d'un schéma de classification dont la structure est peu étendue, la logique est partiellement basée sur des classes d'activité et 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 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) ($p = 0.379$).

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.

Le schéma de classification le moins performant en terme de proportion de réussite au repérage est le schéma C puisqu'un tiers des répondants seulement (32.9%) a été capable de retrouver tous les documents en utilisant ce schéma. Nous rappelons que le schéma C a été artificiellement créé pour présenter des caractéristiques structurelles et sémantiques « extrêmes » (i.e., structure très étendue, absence de structure arborescente, forte redondance structurelle et conceptuelle, intitulés exagérément longs). Les répondants ont manifesté durant la simulation des signes de surprise (par exemple, « Oh, mon Dieu ! », « C'est quelque chose ! »), de découragement (i.e., « C'est long à lire et à chercher », « C'est l'enfer ! ») et parfois de frustration (i.e., de nombreux soupirs, beaucoup d'abandons) en utilisant le schéma C. La structure et les noms des classes ont été jugés « trop longs » par plusieurs répondants qui se sont dits désorientés. Les commentaires des répondants vont dans le sens des résultats des travaux de Prom (2004) qui constate que l'apparence visuelle, ainsi que des dénominations courtes, sont des éléments importants dans l'ergonomie d'un instrument de recherche en environnement électronique. La simplicité est une qualité théorique d'un schéma de classification relevée dans les écrits de Couture et Rousseau (1982), Roberge (2004), Molholt (1995), Héon (1999) et Zins (2002).

Lorsqu'ils sont mis en relation avec l'étendue de la structure, ces résultats vont aussi dans le sens d'un regroupement des classes en vue d'offrir une structure dont l'étendue (i.e., nombre de classes à chaque niveau hiérarchique) doit éviter de dépasser les dimensions de l'écran. L'accès à un document précis est réalisé à partir d'une lecture rapide des classes et mobilise des facultés sensorielles propres à la lecture ou au « balayage » d'un texte. Le besoin de visualiser simultanément l'ensemble de classes sans avoir à faire défiler l'écran (« pas pratique du tout ») a été exprimé verbalement par les répondants durant la simulation. Ce besoin a été également relevé dans les écrits (Gonçalves & Jorge, 2003).

4.2.2 Schémas de classification et temps moyen du repérage

Les résultats indiquent qu'il est plus rapide de retrouver les documents en utilisant le schéma de classification dont la structure 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 moyen 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 2).

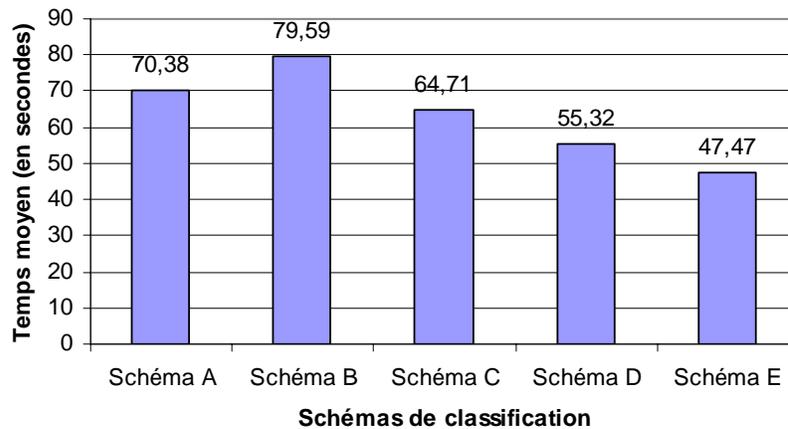


Figure 2 Temps moyen de réussite à l'aide des schémas de classification.

Ces résultats suggèrent qu'un schéma de classification dont la structure est peu étendue, la logique est partiellement basée sur des classes d'activité et 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).

Les pires résultats au niveau du temps moyen de réussite sont obtenus avec le schéma B (schéma personnel de l'employé dont les classes sont définies) (79.59 secondes) et le schéma A (schéma personnel de l'employé) (70.38 secondes). En ce qui concerne le schéma B, ces résultats ne sont guère surprenants car ils incluent le temps de lecture des définitions des classes effectuée par les répondants durant le repérage des documents électroniques. En ce qui concerne le schéma A, le faible score observé sur le plan de la rapidité du repérage pourrait s'expliquer par le nombre élevé de classes de premier niveau (i.e., 42 classes principales) et secondaires (i.e., plus d'une douzaine de classes par niveau), combiné à des subdivisions sur 1 à 2 niveaux qui obligent une lecture à la fois horizontale et verticale des classes. De plus, sans que ces présomptions aient été statistiquement vérifiées, il se pourrait que les critères de division logique qui sont privilégiés au premier (i.e., « Thème ») et au deuxième niveaux hiérarchiques (i.e., « Mélange ») accompagnés d'une forte redondance structurelle et conceptuelle (35.05% de l'ensemble de classes contre 29.7% pour le schéma D et 10.89% pour le schéma E) puissent rallonger la durée de la prise de décision dans l'identification des classes et sous-classes jugées pertinentes pour le repérage des documents cibles.

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

4.2.3 Schémas de classification et proportion de documents repérés dès le premier essai

Les résultats montrent que l'utilisation d'un schéma de classification dont la structure 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) permet de localiser dès le premier essai une proportion significativement plus élevée (79%) de documents (Voir Figure 3). Les résultats révèlent que la différence entre les temps moyen des réussites est significative entre le schéma E et tous les autres schémas testés (schéma A, B, C et D) ($p < 0.001$).

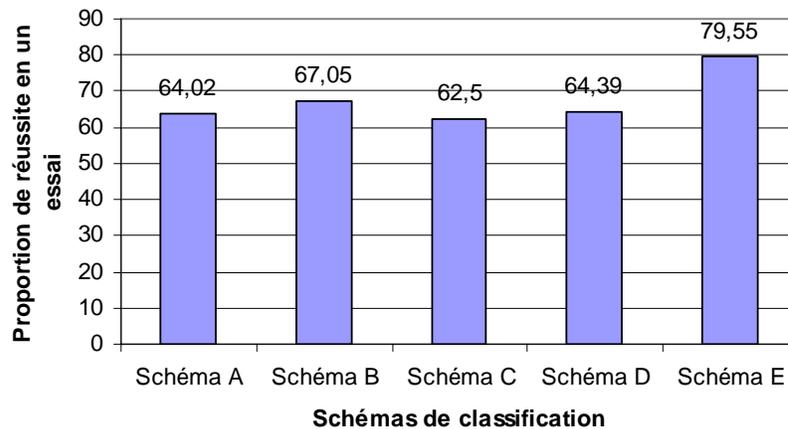


Figure 3 Proportion de documents retrouvés dès le premier essai à l'aide des schémas de classification.

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., macro-structure 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.

Le score le plus faible est obtenu avec le schéma C (62.5% des documents sont retrouvés en un seul essai). Ce schéma est essentiellement caractérisé par l'absence de structure arborescente, l'utilisation d'énoncés extrêmement longs et une plus forte proportion de classes abrégées et redondantes. Ces caractéristiques pourraient générer de la confusion (« on s'y perd ») et parfois de l'impatience (« on s'arrête aux premiers [répertoires] qui semblent pertinents »), sentiments qui pourraient encourager une réponse souvent erronée ou hâtive, d'où un nombre d'essais requis plus élevé.

Les résultats relatifs au faible taux de réussite dès le premier essai obtenu à l'aide du schéma D (i.e., le schéma de classification personnel dont la structure est peu étendue et la logique est partiellement basée sur des classes d'activité) (64.39%) sont relativement surprenants puisqu'ils sont identiques aux résultats obtenus à l'aide du schéma A (i.e., le schéma de classification personnel de l'employé) (64.02%). À l'inverse, les résultats

relatifs au taux de réussite dès le premier essai obtenu à l'aide du schéma E (i.e., le schéma de classification dont la structure est peu étendue, la logique est partiellement basée sur des classes d'activité et la sémantique utilise peu d'abréviations) sont significativement différents des résultats obtenus à l'aide du schéma D. Autrement dit, la dimension sémantique qui distingue le schéma E du schéma D joue un rôle déterminant sur le plan de la probabilité de retrouver une plus grande proportion de documents dès le premier essai.

5. Conclusion

Notre recherche visait à mieux connaître les caractéristiques d'un échantillon de schémas de classification utilisés par deux catégories d'employés pour l'organisation des documents électroniques. Même si nous croyons avec Jones (2007) que la gestion personnelle de l'information réalisée par un individu est unique, les résultats de notre recherche révèlent plusieurs caractéristiques structurelles, logiques et sémantiques communes à une majorité de schémas de classification personnels observés.

Notre recherche avait également comme objectif de vérifier la capacité d'un groupe de répondants à repérer un ensemble de documents à l'aide de cinq schémas de classification distincts aux plans structurel, logique et sémantique. Ce faisant, nous voulions découvrir les schémas de classification qui semblent particulièrement bien fonctionner sur le plan du repérage. Notre recherche a permis d'observer des différences significatives entre des schémas de classification quant à leur efficacité au repérage. Étonnamment, les résultats ont révélé qu'une sémantique personnelle rendue explicite par des définitions de classes permettrait de repérer les documents quel que soit le niveau de normalisation de la logique et de la structure. En ce qui concerne la rapidité du repérage, l'utilisation d'une sémantique personnelle (énoncés plus courts et abrégés) ou formelle (énoncés plus longs, peu abrégés) ne paraît pas affecter de manière significative la rapidité du repérage à l'aide des schémas ayant une structure et une logique normalisées (par exemple, macro-structure peu étendue, logique partiellement basée sur des classes d'activité). Enfin, un schéma de classification dont la structure est peu étendue, la logique est partiellement basée sur des classes d'activité et la sémantique utilise peu d'abréviations minimise de manière significative le risque d'erreur au repérage (i.e., proportion de documents plus élevée dès le premier essai).

Le besoin demeure de poursuivre la recherche dans le domaine de la classification personnelle des documents électroniques. En raison de la taille de l'échantillon et de la méthode d'analyse statistique retenue, nous avons pu attirer l'attention sur les variables les plus influentes sur le repérage soit les caractéristiques logiques, structurelles et sémantiques des schémas de classification personnels. Étant donné l'interrelation et l'influence mutuelle possible de ces différentes caractéristiques, les calculs de variance ne permettent pas d'identifier avec précision quels indicateurs de mesure des dimensions identifiées affectent positivement l'efficacité du repérage. Une telle précision serait souhaitable pour améliorer la conception de schémas de classification personnels qui soient compréhensibles par autrui dans un contexte de travail. Une approche corrélacionnelle pourrait être envisagée à condition de disposer d'un échantillon de schémas de classification de plus grande taille et qui comporterait un ensemble de caractéristiques structurelles, logiques et sémantiques très variées.

Bibliographie

Boardman, R. et Sasse, M. A. 2004. Stuff goes into the computer and doesn't come out : A cross-tool study of personal information management. *CHI 2004, April 24-29, Vienna, Austria* : 583-590.

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.

Couture, C. et Rousseau, J-Y. 1982. *Les archives au XXe siècle : une réponse aux besoins de l'administration et de la recherche*. Montréal : Université de Montréal, Secrétariat général.

Gagnon-Arguin, L. et Bannouri, R. 1998. *La classification des documents institutionnels dans les ministères du gouvernement du Québec et le défi informatique : rapport de recherche*. Montréal : École de bibliothéconomie et des sciences de l'information.

Gonçalves, D. J. et Jorge, J. A. 2003. An empirical study of personal document spaces. *Proceedings DSV-IS'03, 4-6 June 2003, Funchal, Madeira Island, Portugal*. http://immi.inesc-id.pt/publication.php?publication_id=34 (Page consultée le 10 avril 2007).

Henderson, S. 2005. Genre, task, topic and time : facets of personal digital document management. In *Proceedings of the 6th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: making CHI natural, Auckland, New Zealand* : 75-82. <http://staff.business.auckland.ac.nz/staffpages/shenderson/docs/CHINZ2005.pdf> (Page consultée le 10 avril 2007).

Héon, G. 1999. La classification. In : Couture, C., & collaborateurs. *Les fonctions de l'archivistique contemporaine* : 219-254. Sainte-Foy, QC : Presses de l'Université du Québec.

Hudon, M. 2001. Structuration du savoir et organisation des collections dans les répertoires du Web. *Bulletin des Bibliothèques de France*, 46, 1 : 57-62.

Hudon, M., Mas, S. et Gazo, D. 2005. Structure, logic, and semantics in ad hoc classification schemes applied to web-based libraries in the field of education, *Canadian Journal of Information and Library Science/Revue canadienne des sciences de l'information et de bibliothéconomie (CJILS/RCSIB)*, 29, 3 : 265-288.

Hunter, E. 2000. Do we still need classification ? In : Marcella, R., & Maltby, A. (eds.). *The future of classification*. London : Gower : 1-17.

Iyer, H. 1995. *Classificatory structures : Concepts, relations and representation*. Frankfurt, Germany: Indeks.

Jones, W. P., Dumais, S. T. et Bruce, H. 2002. Once found, what then ? : A study of "keeping" behaviors in the personal use of web information. In *Information, connections and community. Proceedings of the 65th ASIST Annual Meeting, Wyndham Franklin Plaza Hotel, Philadelphia, PA.* : 391-402.

Jones, W. P. 2007. Personal information management. *Annual Review of Information Science and Technology*, 41 : 453-504.

- Khoo, C. *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. <http://InformationR.net/ir/12-2/paper293.html> (Page consultée le 10 avril 2007).
- 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 : Éd. d'Organisation.
- Molholt, P. 1995. Qualities of classification schemes for the information superhighway. *Cataloging & Classification Quarterly*, 21, 2 : 19-22.
- Prom, C. J. 2004. User interactions with electronic findings aids in a controlled setting. *The American Archivist*, 67, 2 : 234-268.
- Roberge, M. 2004. *L'essentiel de la gestion documentaire*. Québec : Éditions GESTAR.
- Skupsky, D. S. 1999. Applying records retention to electronic records. *The Information Management Journal*, 33, 3 : 28-35.
- Sprehe, T. J., McClure, C. R., et Zellner, P. 2002. The role of situational factors in managing U.S. federal recordkeeping. *Government Information Quarterly*, 19 : 289-305.
- Tough, A., et Moss, M. 2003. Metadata, controlled vocabulary and directories : Electronic document management and standard for records management. *Records Management Journal*, 13, 1 : 24-31.
- Van Der Walt, M. 1998. The structure of classification schemes used in Internet search engines. In *Structures and relations in knowledge organization : Proceedings of the Fifth International ISKO Conference, 25-29 August 1998, Lille, France* : 379-287.
- Zins, C. 2002. Models for classifying Internet resources. *Knowledge Organisation*, 29, 1 : 20-28.