

## CHAPITRE III

# Un panorama des applications de documentation et d'information mettant en œuvre des profils



## *Aperçu*

---

Pour bien situer la diffusion ciblée, et tirer parti des acquis et des enseignements issus de l'expérience, une exploration méthodique et critique des problématiques voisines est effectuée. Sont concernées les applications informatiques d'information et de documentation, qui font usage de profils ou offrent une forme de personnalisation. Concrètement, c'est l'occasion de cerner ce que recouvrent des termes tels que *filtrage, routing, diffusion sélective, agents intelligents, modèle de l'utilisateur, partage d'informations*. Quatre grandes familles d'applications ressortent, correspondant à quatre conceptions différentes des profils : (i) profil thématique et sélectif, pour la recherche continue (veille) sur un sujet, au service d'une personne ou d'un ensemble d'abonnés ; (ii) profil spécifiant, pour la différenciation des utilisateurs ; (iii) profil caractérisant et régulateur, qui sert de médiation dans un espace informationnel ; (iv) profil identificateur et discriminant, pour le repérage d'individus dans une structure.

---



## Table des matières du Chapitre III

<b>A. PRÉSENTATION .....</b>	<b>99</b>
<b>1. Approche choisie.....</b>	<b>99</b>
<b>2. Bannières de ralliement.....</b>	<b>99</b>
a) <i>Quelques données quantitatives.....</i>	<i>99</i>
b) <i>Le titre, comme reflet des thèmes porteurs .....</i>	<i>100</i>
c) <i>Thèmes stagnants et thèmes montants .....</i>	<i>100</i>
d) <i>Ce qu'il y a derrière les mots.....</i>	<i>100</i>
<b>B. TYPOLOGIE DES USAGES DE PROFILS .....</b>	<b>101</b>
<b>1. Profil thématique, sélectif : recherche continue (veille) sur un sujet, au service d'une personne ou d'un ensemble d'abonnés.....</b>	<b>101</b>
a) <i>DSI - Diffusion Sélective de l'Information (en anglais SDI).....</i>	<i>101</i>
b) <i>Routing, routage.....</i>	<i>102</i>
c) <i>Agents personnels et collecte sélective d'informations .....</i>	<i>105</i>
<b>2. Profil spécifiant : différenciation des utilisateurs.....</b>	<b>107</b>
a) <i>Mémoire, compte utilisateur .....</i>	<i>108</i>
b) <i>Modèle de l'utilisateur pour la contextualisation d'une requête .....</i>	<i>109</i>
c) <i>Service personnalisé : l'exemple de la composition d'un journal.....</i>	<i>111</i>
<b>3. Profil caractérisant et régulateur : médiation dans un espace informationnel ....</b>	<b>112</b>
a) <i>Partage d'information et vue personnalisée sur l'ensemble de l'information rendue accessible .....</i>	<i>112</i>
b) <i>Filtrage.....</i>	<i>114</i>
c) <i>Application de communication évoluée : vue personnalisée sur un ensemble d'informations en circulation .....</i>	<i>116</i>
<b>4. Profil identificateur et discriminant : repérage d'individus dans une structure..</b>	<b>117</b>
a) <i>Catégorisation : application à l'aiguillage d'information.....</i>	<i>117</i>
b) <i>Référentiel de l'activité : repérage d'interlocuteurs .....</i>	<i>118</i>
<b>C. ASPECTS CRITIQUES DE CES APPLICATIONS ET POINTS DE DISCUSSION .....</b>	<b>120</b>
<b>1. Echelle.....</b>	<b>120</b>
<b>2. Durée.....</b>	<b>120</b>
<b>3. Degré de réalisme.....</b>	<b>120</b>
<b>4. Spécificité du corpus.....</b>	<b>120</b>

<b>5. L'assimilation du profil à la personne .....</b>	<b>121</b>
<b>6. Les exigences de la mise en œuvre de l'apprentissage et le caractère adaptatif du profil .....</b>	<b>121</b>
<b>7. Gains réels .....</b>	<b>121</b>
<b>D. LA DIFFUSION CIBLÉE INNOVE : DIX POINTS POUR LA DIFFÉRENCIER DE LA DSI .....</b>	<b>123</b>
<b>1. Délimitation de l'espace des documents .....</b>	<b>123</b>
a) <i>Usage DSI : Espace documentaire clos.....</i>	<i>123</i>
b) <i>Usage diffusion ciblée : Espace documentaire ouvert.....</i>	<i>123</i>
<b>2. Délimitation de l'espace des profils.....</b>	<b>123</b>
a) <i>Usage DSI : Espace des profils indéfini .....</i>	<i>123</i>
b) <i>Usage diffusion ciblée : Espace clos des profils.....</i>	<i>123</i>
<b>3. Personne qui a l'initiative .....</b>	<b>123</b>
a) <i>Usage DSI : Initiative de la part du récepteur .....</i>	<i>123</i>
b) <i>Usage diffusion ciblée : Initiative de la part de l'expéditeur, et de la Direction.....</i>	<i>124</i>
<b>4. Déclenchement d'une utilisation .....</b>	<b>124</b>
a) <i>Usage DSI : Comportement actif.....</i>	<i>124</i>
b) <i>Usage diffusion ciblée : Attitude réactive.....</i>	<i>124</i>
<b>5. Présomption de qualité et de pertinence de l'information.....</b>	<b>124</b>
a) <i>Usage DSI : Information brute .....</i>	<i>124</i>
b) <i>Usage diffusion ciblée : Information sélectionnée .....</i>	<i>124</i>
<b>6. Degré de spécialisation de la recherche .....</b>	<b>124</b>
a) <i>Usage DSI : Requête focalisée.....</i>	<i>124</i>
b) <i>Usage diffusion ciblée : Profil général.....</i>	<i>125</i>
<b>7. Dynamique .....</b>	<b>125</b>
a) <i>Usage DSI : Représentation d'une attente entièrement actuelle, pour une durée éventuellement délimitée .....</i>	<i>125</i>
b) <i>Usage diffusion ciblée : Evolution intégratrice et chronologie interne du profil .....</i>	<i>125</i>
<b>8. Caractère (im)prévisible de l'information.....</b>	<b>125</b>
a) <i>Usage DSI : Résultats attendus.....</i>	<i>125</i>
b) <i>Usage diffusion ciblée : Résultats inattendus .....</i>	<i>125</i>
<b>9. Equilibre bruit / silence.....</b>	<b>125</b>
a) <i>Usage DSI : Ennemi numéro 1 = le silence.....</i>	<i>125</i>
b) <i>Usage diffusion ciblée : Ennemi numéro 1 = le bruit.....</i>	<i>125</i>
<b>10. Possibilités de reformulation .....</b>	<b>126</b>
a) <i>Usage DSI : Interactivité .....</i>	<i>126</i>
b) <i>Usage diffusion ciblée : Définition automatique à large échelle .....</i>	<i>126</i>

## A. PRÉSENTATION

### 1. Approche choisie

Face au volume croissant de l'information disponible et échangée sur les réseaux, et compte-tenu des puissances de calcul et de traitement démultipliées par les ordinateurs récents, la personnalisation des systèmes d'information est une piste d'investigation qui connaît un succès manifeste ces dernières années. L'état de l'art est alors une enquête nécessaire pour assimiler les expériences précédentes, tirer bénéfice des idées porteuses, prendre du recul pour analyser les facteurs d'achoppement.

Il s'agit donc ici de faire un état de l'art des *applications* (un autre bilan sera par ailleurs nécessaire, sur la prise en compte de la dimension textuelle dans les traitements automatisés, à savoir l'*approche* linguistique textuelle choisie dans cette thèse). Le domaine d'investigation doit être choisi suffisamment large pour inclure des variantes connexes, avec lesquelles la diffusion ciblée est parfois confondue ou vers lesquelles elle peut être amenée à évoluer. Soit donc la proposition suivante : tracer un panorama des applications qui traitent de l'information documentaire textuelle et mettent en œuvre des « profils ».

L'étude s'ancre sur un parcours des réalités très diverses que recouvrent les termes-phare mis en avant pour désigner les systèmes (*routing*, *filtrage*, etc.). Dans quels argumentaires s'inscrivent-ils ? Quels problèmes proposent-ils de résoudre, quels gains estiment-ils apporter ? Certains choix techniques sont expérimentés. A partir de quelle(s) source(s) d'information, et de quelle manière, sont constitués les profils ?

Puis vient le temps de faire le point des questions qui gravitent autour de l'utilisation de profils. Quelles difficultés apparaissent pour transposer telle ou telle proposition à notre contexte ? Quels aspects, insuffisamment explorés, nous obligent à renoncer aux solutions existantes ? Quels écueils peut-on d'ores et déjà repérer ?

### 2. Bannières de ralliement

#### a) Quelques données quantitatives

Une consultation des bases recensant les publications illustre d'une manière chiffrée la vague récente d'intérêt autour de ces questions<sup>1</sup>. Les chiffres ci-après correspondent au nombre de références enregistrées dans la base INSPEC, pour le domaine C72 (*information science and documentation*), et utilisant dans leur titre l'un ou l'autre des termes clé que nous avons repérés.

	depuis 1969	depuis 1991	1991	1992	1993	1994	1995	1996	1997	mi-1998
<b>SDI</b>	<b>172</b>	<b>9</b>	1	3	1	0	2	0	1	(1)
profile?	142	52	6	3	5	2	3	9	16	(8)
<b>personaliz*</b>	<b>46</b>	<b>33</b>	<b>0</b>	<b>4</b>	<b>2</b>	<b>0</b>	<b>2</b>	7	<b>11</b>	(7)
<b>routing</b>	<b>49</b>	<b>37</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>8</b>	<b>10</b>	<b>6</b>	<b>8</b>	(3)
<b>filter*</b>	<b>141</b>	<b>121</b>	<b>5</b>	<b>9</b>	<b>7</b>	<b>11</b>	<b>19</b>	<b>27</b>	<b>33</b>	(10)
user?	12 410	894	110	110	93	112	107	124	169	(69)
interest?	144	22	2	4	3	2	0	4	6	(1)

L'interrogation ayant été réalisée le 2 octobre 1998, les chiffres pour 1998 sont à considérer en sachant que l'enregistrement des références cataloguées au titre de l'année 1998 présente un décalage inévitable d'au moins plusieurs mois par rapport à la date de publication. L'année 1998 complète pourrait finalement compter le double de références par rapport au moment où les chiffres ci-dessus ont été obtenus.

<sup>1</sup> Je remercie Cyril DELESTRE (EDF-DER, Département SID), avec qui cette expérimentation a pu être réalisée.

### **b) Le titre, comme reflet des thèmes porteurs**

La restriction de la recherche aux termes présents dans le titre limite évidemment considérablement la sélection et explique les très faibles chiffres obtenus. Beaucoup d'autres documents ont pu avoir telle ou telle notion au cœur de leur argumentaire sans la mentionner sous la forme prévue dans le titre, parce que par exemple elle était déjà présente dans l'intitulé d'un thème de rattachement (colloque, workshop).

En revanche, la composition des titres est significative des termes que l'on veut afficher, et qui suscitent l'intérêt de la communauté scientifique.

### **c) Thèmes stagnants et thèmes montants**

Les chiffres les plus révélateurs ont été mis en gras, et les tendances se dégagent aisément :

- SDI (*Selective Dissemination of Information*) est confirmé comme un thème ancien, un « classique ».
- *Profile / profiles, personalize / personalized / personalization, routing, filter / filters / filtering*, sont autant de termes qui ont suscité un net regain d'intérêt récent.
- Les chiffres sont moins significatifs pour les termes *user(s)* et *interest(s)*, très généraux dans la problématique documentaire. On peut toutefois penser que leur sens évolue et que la manière d'envisager l'utilisateur est renouvelée par les modes de conception des systèmes et de leurs interfaces.

### **d) Ce qu'il y a derrière les mots**

L'engouement pour ces termes est sans doute renforcé par un effet de mode : quand un sujet est porteur, beaucoup s'en réclament. Aussi le domaine s'en trouve-t-il élargi : pour s'y retrouver, il faut distinguer les régions qui se constituent autour de problématiques finalement très différentes.

La typologie qui suit s'est appuyée sur les usages du vocabulaire, le choix des maître-mots, pour cerner les diverses approches : ceux qui décrivent leurs réalisations en termes de *routing* ne se placent pas dans la même perspective que les fidèles des *DSI*, par exemple.



## B. TYPOLOGIE DES USAGES DE PROFILS

### 1. Profil thématique, sélectif : recherche continue (veille) sur un sujet, au service d'une personne ou d'un ensemble d'abonnés

Le profil représente un thème d'intérêt, non pas une personne. En particulier, il peut être utilisé par plusieurs personnes (thème d'intérêt commun) ; et une personne peut utiliser plusieurs profils (déclinaison de plusieurs centres d'intérêt). C'est en quelque sorte une requête documentaire de longue durée, dont on peut recevoir les résultats en souscrivant à un abonnement.

L'analogie entre cette conception d'un profil de veille, qui filtre de l'information, et la recherche documentaire, est chose acquise depuis l'article de Belkin et Croft en 1992. Partageant la même problématique de fourniture d'une information pertinente à un utilisateur, le filtrage et la recherche documentaire se distinguent par trois points : le profil de filtrage représente des intérêts à *long terme* ; il s'applique à un *flux* de documents toujours nouveaux ; il procède *par élimination*.

La constitution des profils est conçue individuellement, même lorsqu'une décision stratégique globale est prise, d'imposer un « devoir d'information »<sup>2</sup> général au niveau de l'entreprise. En effet, on se focalise sur l'ajustement de chaque profil aux thèmes de travail de la personne, sans considérer comment ces profils se situent les uns par rapport aux autres (originalité, spécialité, connaissances partagées) ni la manière dont, ensemble, ils expriment une représentation de l'activité de l'entreprise. L'indépendance des profils les uns par rapport aux autres permet la création, la modification et la suppression de profils sans aucune répercussion sur les autres profils gérés.

#### Repères

**BELKIN N.J., CROFT W.B. (1992)** - « Information filtering and information retrieval : Two sides of the same coin ? », *Communications of the ACM*, 35 (12), décembre 1992, pp. 29-38.

#### a) DSI - Diffusion Sélective de l'Information (en anglais SDI)

UN PROFIL EST une requête documentaire élaborée (mise au point avec un documentaliste), soumise périodiquement à des bases d'information vivantes. Cette équation de recherche peut éventuellement être modifiée (avec le documentaliste) pour rester à jour.

Le profil décrit un centre d'intérêt, un thème, pas une personne.

Une gestion de l'historique des interrogations permet de ne proposer que les documents nouvellement référencés, non encore présentés.

La mise en œuvre d'un système de DSI dans une organisation consiste à :

- se mettre en contact avec un ensemble de destinataires et leur proposer un service d'information régulière, de veille systématique
- définir les thèmes d'intérêt des destinataires et les retranscrire sous forme d'équation de recherche (une requête concerne un sujet et constitue un profil). Une manière de factoriser le travail est de gérer un ensemble de requêtes thématiques spécialisées standards, et de traduire les centres d'intérêt d'un destinataire au moyen de quelques-uns de ces thèmes prédéfinis.
- interroger de façon régulière (quotidienne, hebdomadaire, mensuelle,...) des sources d'information externes (presse, bases documentaires).

<sup>2</sup> L'abonnement à des profils de veille est le plus souvent une offre particulière dans une gamme de produits et services documentaires, un « plus » dont on se dote personnellement, par goût des nouvelles technologies de l'information ou par souci de se maintenir à l'affût de l'actualité. A l'instigation des supérieurs hiérarchiques et/ou dans un contexte stratégique où s'impose une surveillance maximale de l'environnement concurrentiel, cette offre d'information se mue en devoir d'information, au moins pour certaines fonctions centrales (fonctions de décision et de conseil notamment).

- signaler et/ou envoyer à chacun des abonnés les informations nouvelles correspondant à leur(s) profil(s).

Des réalisations sont ainsi proposées en exemple, à plusieurs titres. On s'accorde sur le fait que l'information a valeur stratégique pour l'entreprise : mieux chacun est informé, plus l'entreprise est puissante. Il s'agit donc bien à la fois de capter l'information (pour la thésauriser) et de la faire connaître avec à propos (pour qu'elle soit utilisable). L'entreprise, dans le monde actuel, ne peut pas se permettre d'ignorer son environnement technique et concurrentiel. Dans ce contexte, la DSI vise à recueillir une information récente, systématique, focalisée, donc *a priori* à forte valeur. Cela s'accompagne de surcroît d'un gain en terme de rentabilité, car les destinataires économisent du temps qu'ils utiliseraient pour rechercher l'information.

On est ici en fait devant un montage documentaire bien connu, si bien que dans la littérature publiée sur ce sujet (DSI) il est plutôt question de la mise en place d'un système (grille d'analyse pour la constitution d'un profil, identification des gains justifiant la démarche). Les techniques de repérage de l'information sont traditionnelles : c'est généralement le dépouillement de périodiques par un documentaliste (revue de presse), ou l'enregistrement d'une requête (booléenne) sur un serveur documentaire. Néanmoins, la DSI, s'inscrivant dans la durée, exige un niveau de qualité professionnelle : si l'équation de recherche n'est pas assez précise, l'abonné est noyé par un flux d'informations parasites ; à l'inverse, la veille suppose de ne manquer aucune des informations concernées, y compris celles qui ne font pas partie des types couramment recensés ; enfin, les bases, sources de l'information, conditionnent le tout : elles doivent être choisies en connaissance de cause, pour la qualité des renseignements fournis, leur accessibilité, leur couverture et leur complémentarité, leur adéquation au sujet et au niveau de spécialisation requis.

## Repères

### *Publications et communications*

**RAO M.M. (1993)** - « Selective Dissemination of Information (SDI) : a case study of Central Water and Power Research Station (CWPRS) », *Annals of Library Science and Documentation*, 40 (4), pp. 146-154.

Illustration d'une réalisation dans un Institut National. Le questionnaire à la base de la constitution des profils est en annexe.

**LE LOARER Pierre, NORMIER Etienne (1996)** - « Techniques linguistiques et statistiques pour sélectionner l'information pertinente », *Actes du Congrès IDT'96*, Paris, 21-23 mai 1996, pp. 115-120.

Utilisation de l'outil AlethIR pour (1) l'aide à la constitution de profils de veille, au Centre Commun de Recherche d'Aérospatiale, et (2) la mise en place de profils correspondants aux thématiques du fonds documentaire.

**BROADBENT Marianne, LOFGREN Hans (1993)** - « Information delivery : identifying priorities, performance, and value », *Information Processing & Management*, 29 (6), pp. 683-701.

Effectue une analyse des bénéfices rapportés aux coûts, de l'utilisation d'un système de veille (*Current Awareness Service*) basé sur des profils. Les coûts comprennent : le temps de définition des profils, le matériel et la maintenance ; parmi les gains on relève : des économies de temps et l'impact de l'information, utile, complète et d'actualité, sur la qualité et l'efficacité du travail. L'enquête a porté sur deux organismes et l'article est illustré avec les grilles de résultats chiffrés.

### *Autres acteurs, échos dans les médias :*

**TIP (Technical Interest Profile) à l'EPRI (Electric Power Research Institute)**

cf. : Plaquette de présentation en annexe au Compte-Rendu Express EDF-DER/IPN/SID N40/106, émis par François POURBAIX le 15 novembre 1994.

Un profil est constitué à partir de la sélection de rubriques prédéfinies : thèmes techniques, types de documents souhaités, fonction et rattachement professionnels.

## **b) Routing, routage**

UN PROFIL EST une requête documentaire durable destinée à filtrer un flux de documents, généralement non catalogués (texte intégral au lieu de mots-clés). Cette requête peut évoluer et

s'adapter en fonction des avis émis par l'abonné.

Le profil est ajusté à l'aide de calculs<sup>3</sup> en fonction de jugements émis sur la pertinence d'un ensemble de documents (ensemble arbitraire, conçu comme un échantillon représentatif, servant à calibrer la requête initiale ; ou ensemble des documents déjà soumis au destinataire du profil et (in)validés par celui-ci : on a alors un mécanisme de *feed-back*).

Ce rôle direct des calculs dans la constitution du profil fait qu'il n'existe finalement que dans un format interne ; il se présente par exemple comme une accumulation de termes pondérés, peu lisible pour l'utilisateur.

*Routing* est le terme consacré, dans les grandes conférences d'émanation américaine (telles que SIGIR ou TREC), pour désigner des applications de sélection de documents arrivant en flux. Le principe même du *routing*, c'est :

- de disposer de requêtes-profil : elles ont la forme de requêtes documentaires ; elles sont stables (maintenues dans le temps) et éventuellement évolutives ; elles sont considérées indépendamment les unes des autres.
- de considérer un flux de documents : les documents sont eux aussi considérés indépendamment les uns des autres ; chaque document est une unité, éventuellement scindable (en « passages ») ; il s'agit de documents primaires, en texte intégral.
- d'avoir un processus de décision, qui évalue la pertinence d'un document vis-à-vis d'un profil : il s'agit soit d'une valeur binaire (« envoi » vs « non envoi » du document), soit d'une valeur scalaire qui traduit un degré (relatif ou absolu) de pertinence, la décision d'envoi s'opérant à partir d'un seuil.

Techniquement, le *routing* est opposé à la recherche documentaire ponctuelle sur deux points.

Le premier est la *persistance* et la « mémoire » du profil. Le calcul du profil s'appuie sur son historique : comme le profil s'inscrit dans la durée, on peut considérer que l'on dispose d'un ensemble de documents qui lui ont déjà été attribués et qui ont été appréciés ou rejetés par l'abonné. A l'inverse, la recherche ponctuelle est toujours considérée comme neuve, inédite : on n'a pas de recul, d'expérience, sur les documents qui pourraient répondre à la requête. Cependant, l'initialisation du profil est une étape critique : soit on crée artificiellement un historique en se donnant un ensemble de documents supposé représentatif, soit l'ajustement du profil doit pouvoir se faire avec un historique dont le volume ne grossit que très lentement.

Le second point qui oppose le *routing* à la recherche ponctuelle est la prise en considération obligatoirement séquentielle des documents. Le profil est confronté à un *flux* de documents ; comme ceux-ci arrivent et sont traités de manière échelonnée, on ne dispose pas d'une vue d'ensemble de ce qui sera soumis au système. Dans le cas de la recherche documentaire ponctuelle, l'ensemble des documents de la base sont considérés simultanément, et les relations entre ces documents peuvent être prises en compte.

Une autre conséquence de l'arrivée échelonnée des documents est que l'abonné se voit présenter non pas l'ensemble des documents qu'il a pu sélectionner avec son profil, mais une fraction d'entre eux, ceux du moment. On n'a alors pas les mêmes possibilités d'analyse, de navigation, de structuration, de vue d'ensemble d'un domaine. En revanche la succession temporelle des envois peut, après un certain temps, fournir la base à des analyses de tendance et d'évolution. Des techniques pour cela sont mises au point par la discipline du *Data Mining* ; notamment, l'*analyse procustéenne* intègre une dimension chronologique<sup>4</sup>.

Le *routing* apparaît comme une version automatisée des DSI. Le déploiement à un grand nombre de profils et pour un important volume d'informations est facilement envisageable, sous réserves que le traitement soit conçu pour avoir de bonnes performances : des techniques d'optimisation sont étudiées à cet effet. Il faut aussi tenir compte des incidences périphériques et logistiques sur les services documentaires. Une forte augmentation du volume des commandes est à prévoir, si l'information diffusée consiste en signalements. Pour les documents sélectionnés, la

<sup>3</sup> Typiquement : ajout au profil de mots clés des documents validés ; renforcement ou atténuation de pondérations des mots-clés du profil en fonction de leur présence dans des documents appréciés ou rejetés.

<sup>4</sup> Voir par exemple le logiciel Tétralogie, développé par l'IRIT à Toulouse.

question se pose de leur catalogage rétrospectif, pour les conserver dans le fonds documentaire interne de l'entreprise.

Paradoxalement, cette automatisation peut générer pour les professionnels de l'information un important travail d'analyse documentaire. En effet, les documents repérés et retenus comme les plus intéressants sont archivés dans le fonds documentaire ; on assiste alors à un catalogage rétrospectif massif.

Le profil étant destiné à être actif à moyen-long terme, certains se préoccupent de son adaptation à l'évolution des centres d'intérêt de l'abonné. Les techniques classiques de *feedback* conviennent à une dynamique lente et progressive. Le cas d'une réorientation soudaine des préoccupations de l'abonné est également envisagé. C'est un cas limite, car alors on pourrait considérer que l'abonné décide lui-même de se doter d'un nouveau profil. Quoi qu'il en soit, que l'évolution soit superficielle ou profonde, la difficulté est de ne pas faire qu'ajouter des caractéristiques au profil, mais de garder une représentation synthétique des attentes de l'abonné, sans conserver ce qui n'est plus dans le champ de ses intérêts.

## Repères

### *Publications et communications*

**WYLE M.F., FREI H.P. (1989)** - « Retrieving Highly Dynamic, Widely Distributed Information », *12 th. International ACM SIGIR Conference on Research and Development in Information Retrieval*, June 1989, Cambridge MA, pp. 108-115.

« Our system involves a passive user who is periodically informed when new data matching his interest profile becomes available. The system differs from classical SDI or clipping services in the speed, medium, and information it provides. » (p.115)  
Développements sur la transmission via le réseau ; 5 abonnés (pour l'instant).

**OARD Douglas W. (1997)** - « Adaptive Filtering of Multilingual Document Streams », *Proceedings of RIAO'97 « Computer-Assisted Information Searching on Internet »*, 25-27 juin 1997, Montréal, pp. 233-254.

« As described by Belkin and Croft, information filtering systems seek to sift through large volumes of newly generated information, passing on the user only those which might be useful [Belkin, Croft 1992]. This is essentially the same concept that Luhn earlier called « Selective Dissemination of Information » (SDI) [...], but the term « information filtering » is now more commonly used when the information in question is arriving over a computer network. [...]

Many of the existing text filtering systems require that the user provide an explicit « profile » which specifies their information needs. What we call « adaptive » text filtering systems seek to minimize or eliminate this burden by learning the profile automatically. [...] Users are allowed to provide ratings for documents that they have examined [...] [, and we suppose that their] information need is fairly stable and specific, making it possible to acquire and exploit evidence about that need over time. » (pp.234-235)

**BELL Timothy A.H., MOFFAT Alistair (1996)** - « The Design of a High Performance Information Filtering System », *19 th. International ACM SIGIR Conference on Research and Development in Information Retrieval*, Zurich, 18-22 août 1996, pp. 12-20.

Cas d'un grand nombre de profils : algorithmes et techniques d'optimisation.

**CALLAN Jamie (1996)** - « Document Filtering with Inference Networks », *19 th. International ACM SIGIR Conference on Research and Development in Information Retrieval*, Zurich, 18-22 août 1996, pp. 262-269.

Compare l'ajustement progressif du profil (au fur et à mesure des envois) à la correction que l'on obtiendrait par *relevance feedback* sur l'ensemble des documents soumis. La convergence observée n'est pas très rapide.

**ALLAN James (1996)** - « Incremental Relevance Feedback for Information Filtering », *19 th. International ACM SIGIR Conference on Research and Development in Information Retrieval*, Zurich, 18-22 août 1996, pp. 270-278.

Propose des mécanismes de renouvellement du profil, évitant la focalisation sur une sous-partie du thème d'intérêt (on mémorise les documents pertinents les plus différents les uns des autres), et permettant d'abandonner les critères de sélection périmés (on garde un nombre limité de mots-clés discriminants, dévalués avec le temps s'ils ne sont pas utilisés).

**LAM W., MUKHOPADHYAY S., MOSTAFA J., PALAKAL M. (1996)** - « Detection of Shifts in User Interests for Personalized Information Filtering », *19 th. International ACM SIGIR Conference on Research and Development in Information Retrieval*, Zurich, 18-22 août 1996, pp. 317-325.

L'évolution des intérêts de l'abonné peuvent conduire à la réinitialisation du profil sur une nouvelle définition.

Et toutes les participations à la tâche de *routing*, dans les éditions successives de la Conférence TREC. Les actes sont disponibles sur le site :

<http://www.nist.gov/nlpir>, ou <http://trec.nist.gov/>.

Une présentation en français :

LESPINASSE Karine (1997) - « TREC : une conférence pour l'évaluation des systèmes de recherche d'information », *Documentaliste - Sciences de l'information*, 34 (2), pp. 74-81.

#### *Systemes commercialisés :*

**PLS (Personal Library Software)**, de PLS, Inc., distribué par Stouls Nouveau Média S.A. (Champlan (91), France).

« Personal Agent allows you to submit, via the World Wide Web, persistent « intelligent queries » called agents. They're called persistent because these agents, without further specific directions from you, search full-text databases at periodic intervals. [...] To deploy an agent, you set up a profile for each area you're interested in. [...] Each Personal Agent regularly tries to improve itself by exploring word usage patterns in the incoming data stream »  
(extrait de la plaquette « New Personal Agent 1.0 helps the right information find you », reçue en mai 1997)

#### *Autres acteurs, échos dans les médias :*

**Daimler Benz Research**, à Ulm, Allemagne :

« The aim of information filtering is to select and rank relevant texts of large text collections according to the requirement of users and to forward these texts to the users. Personal filter should be automatically adaptable to the user's needs. »  
(extrait d'une offre d'emploi diffusée sur la liste électronique LN le 17 juin 1996)

### **c) Agents personnels et collecte sélective d'informations**

UN PROFIL EST une représentation des attentes de l'utilisateur en matière d'information. Un ou plusieurs « agents » porteurs de ce profil parcourent l'espace informationnel (avec une stratégie d'optimisation), éventuellement en collaborant entre eux et en se déployant ; ils sélectionnent des documents au fur et à mesure de leur recherche.

L'idée est d'avoir une population d'agents, qui démultiplie une procédure, ici une quête d'informations. Chaque agent étant autonome, il peut se comporter d'une manière particulière, définie par l'utilisateur : d'où l'association forte entre la notion d'agent et de personnalisation. On peut définir de nombreuses variétés d'agents, spécialisés pour diverses tâches : guidage et aide à la navigation dans un espace informationnel, présentation et mise en forme, indexation...

L'appellation même d'*agent* a fortement concouru à leur succès : la métaphore montre l'utilisateur comme un maître ayant une multitude de petits êtres rapides, discrets et « intelligents », purement dédiés à ses intérêts, et corvéables à merci. Un pouvoir vertigineux à portée de main... Il a fallu démystifier l'affaire. Informatiquement parlant, un agent est un (petit) programme, un module. Sa puissance effective est directement liée à l'algorithme qui dicte son comportement, et accessoirement à son implémentation.

Par principe, les agents obéissent à une logique de répartition maximale, de couverture d'un réseau. Dans l'optique d'une recherche d'informations, cette tactique diffusante et centrifuge doit être tempérée par une attention à la topologie du réseau en ce qui concerne la présence de zones plus intéressantes que d'autres<sup>5</sup>. Le parcours doit se rendre sensible aux pertinences locales. Le réseau présente en quelque sorte un relief vis-à-vis d'un sujet, et l'exploration doit en tirer parti plutôt que de le niveler. Il serait dommage de survoler certaines concentrations de pages autour du thème central pour mieux se disperser. Le gain maximal n'est pas d'être « passé un peu partout », mais de creuser les quelques points d'intérêt, de repérer puis de rendre compte des pôles rencontrés et de leur consistance. Les agents sont alors dotés d'une stratégie locale d'orientation (orientation suivant le contexte des liens par exemple) qui peut prendre le pas sur la tendance globale de dispersion. Et la population peut être inégalement renforcée, pour mieux parcourir les zones pressenties les plus pertinentes.

La notion d'agent a ouvert la voie à de nouveaux modes de traitement, à la fois modulaires et ouverts. Ainsi, le comportement d'un agent peut être en partie dicté par le contexte rencontré, et en

<sup>5</sup> La perspective interprétative n'est pas objective, au sens où elle se refuse à tout mettre sur un pied d'égalité. Elle distingue des points d'entrée, des moments critiques. Elle dénonce l'illusion, entretenue par le vocabulaire submergeant de *navigation* internet, de *flux* d'information ou de *débordement* cognitif, que l'on soit confronté à une réalité omniprésente et sans limites. En fait, chacun de nous est dans sa « niche écologique » (dirait François Rastier), et a ses repères, sa perception sélective de son environnement propre (*Umwelt*).

partie lié à un état interne qu'il a en propre : l'agent « s'adapte ». Ou encore, des mécanismes de communication et de coopération sont permis dans certaines réalisations, et sont plus puissants qu'un ensemble de tâches en parallèle. Une modélisation en termes d'agents conduit à envisager le système comme une population, avec sa démographie (clônages), son niveau d'interactivité (« sociabilité »), et tous ses processus d'équilibrage et d'adaptation (naissances / mortalité). Cela ne va pas sans susciter des recherches nouvelles en algorithmique et en architecture des systèmes.

Ces mécanismes qui gouvernent la dynamique de la population d'agents sont la voie pour passer d'une stratégie purement quantitative, de répartition maximale, à une stratégie plus fine de perception du « relief » du réseau par rapport à une attente particulière et de repérage des pertinences locales.

## Repères

### *Publications et communications*

**WIEDERHOLD Gio (1992)** - « Mediators in the architecture of future information systems », *Computer*, 25 (3), pp. 38-49.

Les « médiateurs » peuvent être vus comme les ancêtres des agents, dont ils ont déjà plusieurs caractéristiques.

**MANNINA Bruno, GIRAUD Eric, QUONIAM Luc, ROSTAING Hervé (1997)** - « AURESYS : An Intelligent Agent to collect strategic information from Internet », *Proceedings of RIAO'97 « Computer-Assisted Information Searching on Internet »*, 25-27 juin 1997, Montréal, pp. 658-660.

**ACKERMAN Mark, STARR Brian, PAZZANI Mike (1997)** - « The Do-I-Care Agent : Effective Social Discovery and Filtering on the Web », *Proceedings of RIAO'97 « Computer-Assisted Information Searching on Internet »*, 25-27 juin 1997, Montréal, pp. 17-31.

Les agents peuvent être mis « en cascade », l'agent en aval effectuant sa recherche sur la sélection effectuée par des agents en amont.

**MENCZER Filippo (1997)** - « ARACHNID : Adaptative Retrieval Agents Choosing Heuristic Neighborhoods for Information Discovery », Communication au Colloque *RIAO'97 « Computer-Assisted Information Searching on Internet »*, 25-27 juin 1997, Montréal.

Dans le contexte d'Internet, chaque agent s'oriente en fonction de la pertinence des liens, estimée par leur contexte d'ancrage. La population des agents se renouvelle avec un processus de sélection naturelle, pour démultiplier les agents qui ont été les plus satisfaisants.

**GUDIVADA Venkat N., TOLETY Siva Perraju (1997)** - « A Multiagent Architecture for Information Retrieval on the World-wide Web », *Proceedings of RIAO'97 « Computer-Assisted Information Searching on Internet »*, 25-27 juin 1997, Montréal, pp. 296-309.

Chaque agent représente un module de traitement : interface utilisateur, reconnaissance d'un terme ou d'un concept, apprentissage d'associations terme-concept, indexation, interface avec les moteurs de recherche, calcul de pertinence, coordination des actions.

**GILBERT Don (1997)** - « Intelligent Agents : The Right Information at the Right Time ».

<http://www.networking.ibm.com/iag/iaghome.html>.

Notamment décrit des scénarios d'utilisation d'agents et précise quelques caractéristiques des agents.

**FRANKLIN Stan, GRAESSER Art (1996)** - « Is it an Agent, or just a program ? : A Taxonomy for Autonomous Agents », *Proceedings of the 3<sup>rd</sup> International Workshop on Agent Theories, Architectures, and Languages*, Springer-Verlag.

Cite abondamment les définitions données par les concepteurs d'agents, inventorie les propriétés attribuées aux agents, et propose une classification arborescente.

**UBOIS Jeff (1996)** - « Agents of Change », Interview de Don NORMAN et Jim WHITE, *Internet World*, May 1996, pp. 60-76.

« I've always thought of agents as autonomous systems that are trying to satisfy some goal in the service of the user, but the problem is that almost any program can be thought of that way. » (Don Norman)

« One of the problems of agents is that they sound kind of humanistic, anthropomorphic. » (Don Norman)

« If you describe [an e-mail watcher system] as this wonderful thing to which you tell your preferences and it goes off and gives you a suggestion, now that's an agent. But if I simply say, it's just a vector match, it computes the vector on your preferences and puts it into the multidimensional vector space of all preferences of all people and finds the ones that are closest - that sounds mechanistic, right ? So where's the agent ? And it turns out both phrases core describing the same thing. » (Don Norman)

### Systèmes commercialisés :

#### TOPIC, de Verity Inc. :

« Les InfoAgents personnels, stade le plus évolué de la technologie actuelle, permettent le filtrage, l'analyse et la distribution des informations pertinentes et personnalisées afin de rendre à l'utilisateur le contrôle qu'il avait perdu sur l'accès et l'exploitation de l'information électronique.

Imaginez votre équipe personnelle d'agents intelligents de recherche qui travaillent pour vous, sélectionnant rapidement les documents au travers de larges volumes de données, trouvant l'information personnalisée et pertinente dont vous avez besoin. »

(extrait de la plaquette « Présentation de Verity et de ses produits - *What do you want to find today ?* », janvier 1996)

#### WiseWire, <http://www.wisewire.com>

« 1. What is WiseWire ?

WiseWire is a Web service that uses artificial intelligence to train a « Wire » to deliver high quality online information that can be tailored to a person's interests. WiseWire uses advanced neural network technology to filter all types of digital content. It then categorizes, rates, and delivers the best of Web pages, newswire articles, Usenet newsgroups, mailing lists, and other sources from the Internet to you. And unlike typical search engines that only search web sites on their hit list, over time WiseWire searches the entire Internet to bring you relevant content.

[...]

4. What is collaborative filtering ?

WiseWire uses feedback from others to help predict how valuable information will be to you. WiseWire leverages the feedback of the whole communities in a process called collaborative filtering. Collaborative filtering takes into account the fact that information rated highly by one person is likely to be useful to other people with similar interests. »

(extrait des *Frequently Asked Questions*, <http://www.wisewire.com/faqtext.html>)

#### Autonomy, <http://www.agentware.com/tech.htm>

« Dynamic self updating agent technology : Subject understanding continually evolves with experience. »

« Rather than the user having to go and track down information (the pull model) agents can find areas of interest and present them (the push model). »

« Autonomy Systems already produces end user client software : *Web Researcher* : An agent that performs automatic browsing ; *Press Agent* : An agent that edits your personal paper ; *Mail Agent* : Your agent reads and prioritizes your email before you even look at it ; *Guardian Agent* : Protects the young, a watchful agent ; *TV Listing Agent* : What to watch ? When ? Your agent will keep you informed ; *Browser Agent* : You like that - look at this ; *News Wire Agent* : High quality alerts on news wires for any subject area ; *Kennels* : Agents work on your behalf whilst you are offline ; *Agent Broadcasts* : Send messages to only the interested parties ; *Agent Dating* : A dating service that finds others with similar interests. »

#### IBM Intelligent Agents, <http://www.networking.ibm.com/iag/iaghome.html>

« Intelligent Agents form a new class of software which can act on a user's behalf. Analogous to a private secretary or concierge, agents can help you by finding and filtering information, personalizing it for you, negotiating for services, automating tedious tasks, and taking actions you delegate. »

« WBI (Web Browser Intelligence, pronounced « webby ») [...] can be used to [...] : annotate web pages with new bulletins, ads, or reminders ; track electronic purchases, and provide summaries & reminders ; filter or manage push information content ; develop expert « persona » while searching or browsing. [...] Personal History remembers sites you've visited, and lets you search them by keywords ; Traffic Lights help you avoid slow links ; Internet Watch agent alerts you when sites have changed ; Shortcuts help you traverse the web more productively ; Offline browsing lets you fetch sites onto your hard drive, then view them later ; Surfing is faster and more fun. »

« Agent Building Environment (ABE), alpha version, is a toolkit for software developers that makes it easy to build an application based on intelligent agents or to add agents to an existing application. In the current version, the intelligent agent watches for a certain condition, decides what to do based on the rules you've given it, and takes action as a result. »

### Autres acteurs, échos dans les médias :

Le numéro de juillet 1995 de *TechLink / Natural Language Processing* est consacré à la *Messaging Technology*, et réserve des paragraphes aux *Messaging Agents*, *Organizational Agents*, *Personal Agents*.

## 2. Profil spécifiant : différenciation des utilisateurs

Il est souvent question de « profil » pour désigner un ensemble de paramètres, dont les valeurs sont variables d'un individu à l'autre, et sont prises en compte pour mémoriser un historique d'utilisation, orienter les traitements, ou ajuster la présentation des résultats et les modes d'interaction. Alors que le profil thématique (décrit précédemment) s'intéresse au 'contenu' d'une recherche d'informations, le profil spécifiant concerne l'ajustement et la personnalisation d'un environnement de travail. Le réglage des paramètres est ainsi mémorisé, géré, et coordonné par le système, et épargné à l'utilisateur.

Il y a plusieurs façons de concevoir de tels profils. Rich distingue trois dimensions :

- *court terme vs long terme* : certaines modélisations mémorisent et gèrent tout un historique des utilisations successives pour une même personne. D'autres ne constituent un profil que ponctuellement, pour une session d'utilisation ; par exemple, lorsque l'interaction est sous forme

de dialogue, un « modèle de l'interlocuteur » est peu à peu enrichi, le système enregistre au fur et à mesure les informations décrivant le besoin d'information et s'ajuste au mode de communication et au registre de son interlocuteur humain. Troisième cas, intermédiaire : le profil articule une description stable (dynamique lente : compétences cognitives, habitudes, situations déjà rencontrées) et une description conjoncturelle (domaine d'intérêt spécifique, momentané). Dans tous les cas, le profil joue le rôle d'une toile de fond, situant et précisant la requête, réduisant une part de l'implicite lié à la brièveté de celle-ci.

- *données explicites vs implicites* : dans le premier cas l'utilisateur se décrit, par exemple en cochant les traits qui lui correspondent dans un formulaire ou en répondant à un questionnaire. Dans le deuxième cas, le système induit les préférences de l'utilisateur à partir de l'interaction utilisateur-système, dans l'expression du besoin (au moment de la formulation de la requête) et/ou dans l'expression de la satisfaction (mécanisme de *relevance feedback* : l'utilisateur indique les documents qui lui conviennent *-feedback positif-* et ceux qui ne correspondent pas à ce qu'il cherche *-feedback négatif-*). Le système présente alors certaines capacités d'apprentissage et d'adaptation.
- *profil d'individu vs collectif* : l'accent est mis soit sur une description particulière propre à chaque utilisateur, soit sur l'adéquation du système à plusieurs catégories d'utilisateurs-type.

### Repères

**ALLEN Robert B. (1990)** - « User models : theory, method and practice », *International Journal of Man-Machine Studies*, 32, pp. 511-43.

Aborde la question de l'acceptation ou du rejet de la modélisation par les personnes. Nombreuses notations instructives sur les apports et les limites des modèles de l'utilisateur.

**RICH E. (1983)** - « Users are individuals : individualizing user models », *International Journal of Man-Machine Studies*, 18, pp. 199-214.

#### a) *Mémoire, compte utilisateur*

UN PROFIL EST une configuration du système ou de l'application informatique ; il définit un environnement de travail : aspect de l'interface, données enregistrées par l'utilisateur qu'il peut vouloir consulter à nouveau, enrichir, exploiter en propre.

Cette notion de compte-utilisateur existe de longue date en informatique, notamment dans le contexte de stations de travail en lien les unes avec les autres via un réseau. La confidentialité des environnements de travail de chacun est préservée par une procédure d'authentification : au profil est associé un mot de passe, qui en contrôle l'utilisation.

Un cas intéressant dans le contexte de notre étude est celui de postes de lecture, dans une grande bibliothèque publique par exemple. Chaque utilisateur peut insérer et conserver ses annotations, placer ses marque-page, se constituer une sélection de documents de travail (souvent appelée *panier*), conserver des résultats d'analyses. Ces informations sont personnelles et peuvent être retrouvées d'une session sur l'autre.

### Repères

#### *Publications et communications*

**BONHOMME P., BRUNESSEAU F., CRUZLARA-SILVA S., MAPELLI V., NITTI A., ROMARY L. (1997)** - *Le serveur SILFIDE : mise en œuvre, fonctionnalités, évolutivité*, Rapport final, AUPELF-UREF, CNRS, mars 1997, 89 pages.

SILFIDE (*Serveur Interactif pour la Langue Française, son Identité, sa Diffusion et son Etude*) a pour vocation de donner un accès centralisé, via Internet, à des ressources, corpus, et outils linguistiques.

Chaque utilisateur ou fournisseur est identifié. Un utilisateur peut sauvegarder des étapes d'analyse d'une session sur l'autre (notion de *panier*, cf. §5.5, p. 56).



CHAHUNEAU François, LÉCLUSE Christophe, STIEGLER Bernard, VIRBEL Jacques (1992) - « Prototyping the Ultimate Tool for Scholarly Qualitative Research on Texts », 8<sup>th</sup> Annual Conference of the University of Waterloo Center for the New Oxford English Dictionary and Text Research, 18-20 octobre 1992, 26 pages.

Présentation d'un prototype de Poste de Lecture Assistée par Ordinateur. Dans la réalisation décrite, l'utilisateur travaille sur une base personnelle, composée de documents importés du fonds documentaire et éventuellement d'autres sources : « this private database is seen as their own customized view of the central database » (§3.1). Le système ne gère donc pas parallèlement, et de façon centralisée, *n* profils utilisateurs, mémorisant *n* environnements par rapport aux mêmes ressources.

### ***b) Modèle de l'utilisateur pour la contextualisation d'une requête***

UN PROFIL EST une configuration des valeurs des paramètres d'un système (documentaire) en vue d'un comportement et d'une ergonomie adaptés à une des catégories d'utilisateurs. Le système dispose alors d'informations pour contextualiser la requête qui lui est soumise.

*Corrélaire* : on définit donc une typologie des utilisateurs, décrivant les caractéristiques utiles à connaître pour l'application, ainsi que la manière dont ces caractéristiques doivent être traduites et intervenir dans le traitement.

Concrètement, dans le domaine des systèmes documentaires, les applications s'efforcent de cerner des informations de trois ordres.

*Les caractéristiques personnelles de l'utilisateur* colorent ses recherches et se manifestent de façon relativement stable d'une session de recherche sur l'autre. Le *niveau d'études* de l'utilisateur est utilisé pour cerner ce qui lui est connu, ce qui peut être nouveau pour lui, sa capacité à comprendre et à tirer profit de tel ou tel document. Ses préférences générales quant au *type de document* (ouvrage de référence, article de vulgarisation, rubrique donnée dans un quotidien, forme détaillée ou synthétique...) paramètrent les propositions du système. Le tempérament de l'utilisateur joue sur le *temps* et l'*effort* qu'il est prêt à consentir pour atteindre une certaine qualité de résultats.

Deuxième type d'informations recueillies par le système, *le contexte particulier de la recherche* précise les attentes de l'utilisateur par rapport au thème de recherche formulé. La manière de s'exprimer de l'utilisateur, à savoir le vocabulaire employé, est classiquement utilisée comme un indice du degré de *généralité / spécificité* des informations souhaitées. Le système complète ce genre d'observations (limitées et heuristiques) par un questionnement, qui alourdit la session mais peut rendre plus clair le comportement du système. Il demande des renseignements sur la situation dans laquelle s'inscrit cette recherche d'information et l'*utilisation* escomptée des documents (état de l'art, résolution d'un problème ponctuel...). A cela est liée la *nature des informations* attendues (théorie ou pratique, prospective ou bilan, analyse ou synthèse, actualité ou historique, à diffusion large ou confidentielle, auteur ou collection reconnue). Dans le même temps, se précise le *volume* de propositions escompté, le degré d'*exhaustivité* de la recherche (un document contenant l'information suffit, ou bien recensement de l'ensemble des contributions à un sujet).

Troisième paramètre, l'interface doit s'adapter au *mode de relation de l'utilisateur à un système automatique*. On évalue sa *connaissance du système documentaire*, qui doit lui permettre de jauger les domaines couverts par la base, et qui détermine les modes d'accès (vocabulaire et syntaxe d'interrogation). D'autre part, si l'utilisateur a une *pratique* fréquente et professionnelle des recherches documentaires, l'emprise du guidage de la part du système doit être réduite, en faisant crédit à son habileté à élaborer des stratégies efficaces répondant à son besoin. Dans le cas inverse, l'utilisateur est moins autonome et le système prend davantage en charge la conduite de la recherche.

Le principe général de ce recueil d'informations est de disposer d'une prédiction concernant les intérêts de l'utilisateur, avant même l'expression d'une demande précise (la requête), afin d'être en mesure de contextualiser celle-ci.

Finalement, la donnée des caractéristiques factuelles sur lesquelles est construit le profil de l'utilisateur (âge, niveau,...) détermine de fait, de façon sous-jacente, une partition discriminante des utilisateurs : à chaque classe est associé un mode de comportement et de réponse du système. Une des difficultés est d'ailleurs d'obtenir des représentations stables et consistantes, de sorte qu'une variante minime dans le descriptif d'une personne n'engendre pas un saut qualitatif important d'une classe à une autre. Les classes instaurent une typologie des lecteurs et des usages, ni plus ni moins. Ni plus : on a affaire à une typologie particulière, liée aux circonstances de son utilisation ; d'autres typologies

seraient envisageables. Ni moins : cette typologie plaque un certain point de vue sur les pratiques documentaires. C'est un repère explicite mais réducteur (ancrage culturel, simplification) auquel rapporter la réalité<sup>6</sup>.

Cette représentation proposée par le système et dans le jeu de laquelle l'utilisateur doit rentrer, est plus de l'ordre d'une segmentation de la *clientèle (a priori)* que d'une adaptation personnelle à l'*utilisateur (a posteriori)*. Tout le problème de l'utilisation de profils semble donc se cristalliser sur l'identification des facteurs susceptibles d'aider à la prédiction des intérêts de l'utilisateur.

La difficulté pour caractériser les choix de lecture (facteurs sociaux, affectifs, relationnels) limitent le rôle et la portée des modèles d'utilisateurs. En général, le modèle opère une présélection grossière d'une partie de l'ensemble des documents, autrement dit délimite un sous-corpus, sur lequel est lancée la requête. On escompte un gain en précision et en efficacité puisqu'on a écarté tout un pan de la base *a priori* non pertinent, suivant en cela une stratégie documentaire éprouvée et efficace. Le profil n'a pas l'envergure d'une modélisation fine permettant de prédire l'ensemble des documents intéressant une personne<sup>7</sup>.

Mais la préoccupation de définir des profils aura été l'occasion de mettre en exergue que la perspective documentaire ne doit pas se focaliser sur une analyse fine du contenu du document, considéré indépendamment de ses lecteurs : car le sens d'un document ne lui est pas intrinsèque, il est construit par le lecteur. A son tour, il apparaît que le profil ne peut prétendre contenir les intérêts de l'utilisateur.

« It is intriguing to speculate on whether there is an inherent unpredictability in human behavior or whether with sufficient information all behavior could be predicted. » [Allen, 1990]

La nécessaire prise en compte des différences entre les utilisateurs, et du décalage entre, d'un côté, la manière dont la personne se représente sa problématique de recherche d'information, et de l'autre, la modélisation et le traitement mis en œuvre par le système, motive donc une observation méthodique des démarches des utilisateurs. Ces travaux font ressortir de nombreux facteurs intervenant dans l'appréciation de la pertinence<sup>8</sup>. Il s'agit aussi d'éviter l'écueil qui consiste à *dérivée* le modèle des indicateurs mesurables au lieu de *concevoir* la modélisation à partir de l'observation et l'analyse des personnes. La place d'une bonne connaissance des techniques est d'y trouver des voies opportunes de réalisations opérationnelles, et non de cantonner à leurs mécanismes la vision et l'analyse de la réalité.

## Repères

### *Publications et communications*

**INGWERSEN Peter (1994)** - « Polyrepresentation of Information Needs and Semantic Entities - Elements of a Cognitive Theory for Information Retrieval Interaction », *17 th. International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 1994, Dublin, Irlande, pp. 101-110.

Le paradoxe de la requête documentaire, c'est que l'on ne sait pas encore ce que l'on cherche. Par contre, on peut expliciter ce qui est sous-jacent et motive la recherche. L'espace cognitif complète l'espace informationnel de la base interrogée.

**BELKIN N.J., BROOKS H.M., DANIELS P.J. (1987)** - « Knowledge elicitation using discourse analysis », *International Journal of Man-Machine Studies*, 27, pp. 127-144.

Le modèle de l'utilisateur fait partie des connaissances mises en œuvre par l'intermédiaire (le documentaliste) qui assiste la personne dans sa recherche d'informations. Proposition d'architecture à partir de l'observation d'experts.

**DANIELS P.J. (1987)** - « Cognitive models in information retrieval : an evaluative review », *Journal of Documentation*, 42(4), pp. 272-304.

« Une excellente présentation des modèles utilisés et des difficultés rencontrées lors de leur application dans le domaine de la recherche d'informations », au dire de Yolla POLITY.

<sup>6</sup> Certaines modélisations tournent à la « caricature sociale » : l'âge, le sexe, la race, les diplômes conditionneraient alors les choix de lecture.

<sup>7</sup> Une telle conception serait d'ailleurs contestable, car cela reviendrait à définir une forme de pertinence *a priori*, cf. la discussion que nous faisons de la notion de pertinence.

<sup>8</sup> Ceci sera repris au chapitre traitant de la question de la pertinence.

**FOX Edward A. (1987)** - « Development of the CODER system : A testbed for Artificial Intelligence methods in Information retrieval », *Information Processing & Management* 23 (4), pp. 341-366.

Un exemple de mise en œuvre, qui intègre les conceptions classiques (présentation du système en cours de développement, en laboratoire), après avoir exposé un large panorama de l'état de l'art.

**CHEN Hsinchun (1992)** - « Knowledge-based document retrieval: framework and design », *Journal of Information Science* 18, pp. 293-314.

Présentation d'un prototype (Metacaf) incluant un module de représentation de l'utilisateur ; l'importance de ce module est soulignée et argumentée ; des éléments sont donnés sur son contenu.

**STADNYK Irene, KASS Robert (1992)** - « Modeling Users' Interests in Information Filters », *Communications of the ACM*, 35 (12), décembre 1992, pp. 49-50.

Facteurs qui complètent la requête (expression du thème recherché) pour mieux cerner les documents pertinents, par exemple l'usage prévu de l'information, la longueur du texte.

**RULQUIN V., DAVID A., THIERY O. (1995)** - « Personnalisation des réponses dans un Système de Recherche d'Information », Communication au Colloque sur *Les systèmes d'information élaborée*, 30 mai - 2 juin 1995, Ile-Rousse, Corse, *Revue Française de Bibliométrie* n°14, juillet 1995, pp. 499-500.

Enseignements tirés d'une application au prototype BIRDS, dans le domaine de l'enseignement assisté par ordinateur (équipe du CRIN à Nancy). S'intéresse entre autres au niveau de spécificité du vocabulaire.

**LAINÉ-CRUZEL Sylvie (1994)** - « Vers de nouveaux systèmes d'information prenant en compte le profil des utilisateurs », *Documentaliste - Sciences de l'information*, 31(3), pp. 143-147.

**LAINÉ-CRUZEL S., LAFOUGE T., LARDY J.-P., BEN ABDALLAH N. (1996)** - « Improving information retrieval by combining user profile and document segmentation », *Information Processing & Management* 32 (3), pp. 305-315.

**MICHEL C., LAFOUGE T. (1997)** - « Profil-doc : Un système personnalisé de requête à des bases de données en texte intégral », *Actes du Colloque sur Les systèmes d'information élaborée*, 12-16 mai 1997, Ile-Rousse, Corse, p. 37.

Le profil combine informations stables et contextuelles. Il se traduit par une équation booléenne de propriétés, sélectionnant la partie de la base qui sera interrogée.

**DUNNE Chris, VERBRUGGEN Renaat (1993)** - « Navigation in a Hypertext using dynamically constructed guided tours », *2èmes Journées Internationales sur L'interface des mondes réels & virtuels*, 22-26 mars 1993, Montpellier, pp. 263-272.

Modèle de l'utilisateur en fonction de son utilisation du système, servant à le guider dans une navigation hypertexte : enregistre la fiabilité (taux de satisfaction) et l'actualité (utilisation récente) de chaque critère de sélection d'une page.

**GUIZOL Jacques, RUBIO Ruben Gonzalez (1997)** - « Un système de recherche et de filtrage d'information multilingue », *Proceedings of RIAO'97 « Computer-Assisted Information Searching on Internet »*, 25-27 juin 1997, Montréal, pp. 773-782.

Le profil mémorise les documents retenus par l'utilisateur, et gère une liste de termes extraits de ces documents pour enrichir les requêtes suivantes.

### ***c) Service personnalisé : l'exemple de la composition d'un journal***

UN PROFIL EST un ensemble d'instructions (de sélection, de mise en page), pour la génération automatique d'un document à la convenance du lecteur, à partir d'un ensemble de données (articles de presse des différentes rubriques d'un quotidien par exemple).

Partant d'un même « matériau », un large ensemble d'articles par exemple, chaque lecteur se voit doté d'une édition spéciale, avec sa propre maquette et son propre choix d'articles. Cette édition met en valeur ses rubriques favorites, facilite l'accès aux informations que le lecteur veut repérer et suivre, propose un parcours de lecture de l'actualité orienté par ce qui préoccupe le lecteur.

Des questions déontologiques sont apparues : doit-on fournir au lecteur uniquement ce qui lui plaît ? n'a-t-on pas aussi le devoir de lui faire part des préoccupations partagées par tous, des questions en débat dans la société qui l'entoure ? Un juste équilibre est à trouver, pour une personnalisation ouverte, non égocentrée.

## Repères

### *Publications et communications*

**KAMBA Tomonari, BHARAT Krishna (1996)** - « An interactive, personalized, newspaper on the WWW », *Proceedings of the SPIE*, vol.2667, pp. 290-301.

Une réalisation (*Krakatoa Chronicle*) et les enseignements, techniques et sociaux, tirés de l'expérience.

**MALAVAL Catherine (1995)** - « La presse à l'heure du multimédia : l'exemple américain. A new frontier for American newspapers », *Actes de la Conférence IDT Marchés et Industries de l'Information*, 13-15 juin 1995, Paris, pp. 237-240.

Une première étape est la présentation sur support électronique de journaux existants ; l'interactivité et la personnalisation ne connaissent encore que très peu de réalisations.

### *Autres acteurs, échos dans les médias :*

Journal *Le Monde*, samedi 29 avril 1995, p.25 : « Le Wall Street Journal crée une édition personnalisée sur écran », et encart « les archives du *Monde* en ligne »

Choix par le lecteur de 10 sujets, sous la forme de noms d'entreprises ou de titres de rubriques du *Wall Street Journal* ; également, sélection de cours de la Bourse, et informations générales (données à tous les abonnés). Les informations sont enrichies en temps réel.

« Un service permettant de donner à ses lecteurs des informations « profilées », du type de celui du *Wall Street Journal*, est actuellement à l'étude. Dans une première étape, *Le Monde* proposera un service d'informations économiques réalisé en partenariat avec Reuter. »

## 3. Profil caractérisant et régulateur : médiation dans un espace informationnel

Le paradigme ici décrit est celui d'un très vaste ensemble de documents, sur lequel se greffent une multiplicité de points de vue, ceux des utilisateurs. Par rapport aux deux familles de profils précédentes (profil thématique, profil spécifiant), l'accent est mis sur le partage d'une ressource commune, l'existence d'un terrain d'échange et de circulation de l'information, autour de laquelle les personnes peuvent entrer en communication.

Le volume important de l'ensemble de documents peut découler de plusieurs facteurs. Il peut s'agir du regroupement dans un seul grand fonds de multiples petites bases d'information, dispersées et partiellement redondantes. Ce peut être aussi les flux qui circulent sur les « autoroutes de l'information ».

Du côté des personnes, les pratiques professionnelles balisent les usages de l'information. Chacun pioche dans une base commune ce qui lui est utile, chacun y contribue dans son secteur, et ne doit pas être perturbé par le reste. En effet, le volume démesuré et la diversité des documents rassemblés sont tels qu'il n'est ni possible, ni souhaitable, de présenter l'ensemble directement à celui qui consulte la base. La médiation par un profil doit aider à éviter le *Cognitive Overflow Syndrom*, surcharge informationnelle de la personne qui perd alors ses repères et son efficacité.

## Repères

### *Autres acteurs, échos dans les médias :*

Un groupe de travail de l'*Association pour la Recherche Cognitive*, en France, étudie le COS (*Cognitive Overflow Syndrom*, Syndrome de Saturation Cognitive), dans les situations de travail de bureau.  
<http://www.mines.u-nancy.fr/~arc/index.html>

### ***a) Partage d'information et vue personnalisée sur l'ensemble de l'information rendue accessible***

UN PROFIL EST un ensemble de caractéristiques d'une personne ou d'un groupe (équipe, fonction) focalisant les interactions possibles sur ce qui est pertinent (par exemple, des opérations de recherche dans le fonds et les archives, ou bien des opérations de signalement de

nouveautés, avec dans chaque cas la définition d'un point de vue sur la base).

On trouve cette notion dans les systèmes de GED (*Gestion Electronique des Documents*), et dans certains aspects du *Groupware* (*collecticiels* pour le travail collectif) et du *Workflow* (circuit des documents : rédaction, relecture, évolution des versions, validation, conservation, diffusion...).

Un fonds regroupe toutes les informations liées à une collectivité et les gère de façon centralisée. Cependant, chacun a une vision différente de ces informations. En fonction de son poste ou de ses responsabilités par exemple, la nature et le sujet des documents qui le concernent varient ; des degrés ou des zones de confidentialité contrôlent et limitent la divulgation des données sensibles et protègent les informations privées.

Du fait de la division du travail et de la spécialisation des activités, le partage d'informations n'est pas aussi massif que ce à quoi l'on pourrait s'attendre. Et quand le besoin (ponctuel) d'élargir son champ de connaissances et d'action se fait sentir, le réflexe n'est pas toujours acquis, d'avoir l'initiative de consulter la base (sans être gêné par des barrières d'accessibilité, dans le meilleur des cas).

Enfin, la désignation souvent adoptée de *mémoire d'entreprise* appelle d'elle-même la question des mécanismes d'oubli, celle d'un relief entre des éléments saillants ou latents, etc.

## Repères

### *Publications et communications*

**ISRAEL Rachel, DE BRITO Gabrielle (1998)** - « Technique et mémoire collective », *Actes du Séminaire Technique & Cognition – Finitude, Situation et Inscription corporelle*, Compiègne, 19-23 janvier 1998, pp. 133- 136.

La mémoire humaine ne se comporte pas comme un stockage accumulatif d'informations. Il s'y joue notamment des processus léthogènes, *i.e.* qui génèrent l'oubli.

Le modèle dominant pour les mémoires d'entreprise devrait faire place à des dispositifs de sélection, d'oubli, d'appropriation, d'inscription multi points de vue, et reconnaître le rôle des interactions sociales (représentations mutuelles, processus de convergence et de stabilisation).

**NGUYEN Nam (1997)** - « Les usages performants des Nouvelles Technologies de la Communication », *Ateliers de la Recherche EDF*, 6 novembre 1997, Clamart.

Témoignage d'utilisation d'un logiciel de *groupware*.

**NUWEMDAM F., BALLAY J.-F. (1993)** - *DIADEME : Etat de l'art et analyse des technologies informatiques pour la gestion des connaissances*, Collection des Notes Internes de la Direction des Etudes et Recherches d'EDF, 94NO00019, décembre 1993, Clamart : EDF-DER, 93 pages.

Explore hypertexte, systèmes experts et bases de données, et la bureautique communicante.

**POITOU Jean-Pierre, BALLAY Jean-François, SAINTIVE Brigitte (1997)** - *La gestion collective des connaissances : « Savoir s'y prendre »*, Note interne EDF-DER, HN-51/97/010, juillet 1997, accessibilité libre, Clamart : EDF-DER, 29 pages.

Retranscription de l'enregistrement d'un exposé de Jean-Pierre Poitou (Centre de Recherche en Psychologie de Provence), suivi d'une intervention de Jean-François Ballay (EDF-DER, responsable du projet DIADEME) et d'un débat avec la salle. Apporte une réflexion de fond et un retour d'expérience sur la mise en œuvre d'un système de gestion collective des connaissances, fondé sur l'archivage et la mise à disposition d'un fonds documentaire local de référence, pour un domaine spécialisé.

Sur la division du travail, le besoin effectif de partage des connaissances, et l'initiative de consulter la base, voir p. 26.

### *Systèmes commercialisés :*

**Documentum**, de Documentum Inc. :

« we recognize that a « one-size-fits-all » approach doesn't make it in today's business environment. You need to be able to deploy full document management functionality to users whose roles require it, while delivering more tailored solutions to others based on their jobs or roles. » (extrait de la plaquette *The Right Product for the Right User*)

**Fulcrum**, de Fulcrum Technologies Inc. (une présentation est fournie dans les *Proceedings of RIAO'97 « Computer-Assisted Information Searching on Internet »*, 25-27 juin 1997, Montréal, pp. 687-690) :

« The Knowledge Map's shared knowledge folder capability lets your users focus on the content of information that is relevant to their own projects, without concerning themselves with its physical location or access. Your users can create personalized views of the various information sources by creating personal knowledge folders that match their own areas of responsibility. »

## b) Filtrage

UN PROFIL EST une représentation des attentes d'une personne utilisée pour évaluer *a priori* l'intérêt de documents arrivant en flux indifférencié, par un canal qu'il n'est pas possible ou souhaitable de fermer (l'exemple par excellence est la messagerie électronique). Les documents concernés sont généralement de l'information périssable et non cataloguée (texte intégral, pas de mots-clés). Le profil est une représentation interne, calculée (peu lisible, délicate à interpréter), éventuellement enrichie en fonction des avis émis par l'utilisateur.

A la base de l'idée de filtrage, est le consensus sur un contexte alarmant : on redoute d'être noyé sous un flux d'informations. La notion de filtrage séduit au premier abord, tout particulièrement vis-à-vis d'une information fourmillante et fragmentaire<sup>9</sup>. Sont donc typiquement visés les documents éphémères, hôtes des forums ou de messageries.

Trivialement, le principe du filtrage ne fait que retourner, en négatif, un principe de sélection. L'accent est mis sur l'idée que l'on élimine de l'information (inutile...), que l'on réduit le volume d'informations à traiter, ce qui est porteur dans le contexte précité d'attitude défensive devant une surcharge d'informations.

En ce qui concerne la réduction du volume d'informations, le filtrage se situe en aval alors que la diffusion ciblée opère en amont. En évitant une diffusion générale aveugle (*broadcast*), la diffusion ciblée supprime le besoin de filtrage, désencombre les circuits d'information, et économise, par un traitement au niveau de l'émetteur, *n* traitements au niveau des destinataires.

Dans le cas d'une messagerie, l'élimination pure et simple de courriers, non supervisée (faite automatiquement par le système), est incontestablement dangereuse. En effet, le propriétaire de la boîte aux lettres est destinataire explicite des courriers qu'il reçoit ; or il est extrêmement délicat de fixer des critères infailibles pour faire la part entre les messages qui doivent être présentés et ceux qui ne mériteraient aucune attention. De nombreux paramètres rentrent en ligne de compte, articulant non seulement la teneur du message mais aussi son auteur.

Comme l'élimination par erreur par le système d'un message peut être très grave, la prudence voudrait que le destinataire puisse contrôler les messages que le système propose à l'élimination. Le gain en temps apporté par le système de filtrage est alors douteux.

En définitive, quand on a affaire à des canaux d'information personnels, le filtrage est un procédé trop brutal. D'ailleurs, dans la réalité du travail de bureau, l'élimination prend la forme aimable et progressive d'un changement de statut : le document est mis sur la pile en attente, attente qui n'est très vraisemblablement qu'un sursis. La pile joue le rôle d'une zone transitoire. La coupure binaire qui tranche entre les documents sélectionnés et les documents éliminés s'apparente à une conception booléenne de la pertinence, conception dont le caractère réducteur est connu. D'autres prétraitements sont envisageables, comme des classements ou toute autre forme de hiérarchisation des messages (cf. paragraphe suivant, sur les applications de communication évoluées).

La déficience que l'on peut trouver à l'objectif du filtrage de l'information ne doit pas faire rejeter en bloc les travaux qui s'en réclament. De nombreuses compétences se sont mobilisées autour de cette problématique du filtrage, et certaines contributions se distinguent par des idées intéressantes pour la conception de profils et l'ergonomie de l'interface.

Il faut signaler qu'un autre concept de filtrage d'information existe. Il s'agit de la sélection de phrases dans un texte, ou du repérage de certaines informations. Le processus est guidé non pas par un profil, mais soit par des indices d'importance générale (mise en valeur d'un élément par le rédacteur),

<sup>9</sup> Le groupe de travail sur le *Cognitive Overflow Syndrom* (Charles LENAY, Saadi, LAHLOU, etc.) voit dans la *fragmentation* croissante de l'information un des facteurs de saturation cognitive. Le traitement cognitif de multiples petits documents indépendants n'est pas équivalent à celui d'un seul document, plus complexe mais plus construit. Dans le premier cas, chaque petit document oblige le lecteur à (re)construire un contexte approprié. La gestion de ces sauts contextuels rencontre rapidement les limites des capacités humaines.

soit par une grille de lecture fixe et prédéterminée, qui spécifie ce qui est relevé dans les textes. Du fait de l'absence de profils ou de personnalisation, ces travaux sortent du champ de cette étude<sup>10</sup>.

## Repères

### *Publications et communications*

**LOCH Shoshana, TERRY Doug (1992)** - « Information filtering », *Communications of the ACM*, 35 (12), décembre 1992, pp. 26-28.

Introduction à un numéro spécial sur ce thème ; les auteurs citent Peter Denning, dans une Lettre sur l'*Electronic Junk* : « The visibility of personal computers, individual workstations, and local area networks has focused most of the attention on *generating* information - the process of producing documents and disseminating them. It is now time to focus more attention on *receiving* information - the process of controlling and filtering information that reaches the persons who must use it. »

**FOLTZ Peter W., DUMAIS Susan T. (1992)** - « Personalized Information Delivery : An Analysis of Information Filtering Methods », *Communications of the ACM*, 35 (12), décembre 1992, pp. 51-60.

Expérimentation de méthodes connues dans le domaine de la recherche d'information (VSM, LSI...) pour la sélection de documents (seuillage).

**SHETH Beerud, MAES Pattie (1993)** - « Evolving Agents for Personalized Information Filtering », *Proceedings of the 9<sup>th</sup> IEEE Conference on Artificial Intelligence for Applications*, Orlando, 1-5 March 1993, pp. 345-352.

Prototype appliqué aux *newsgroups* (forum électronique). L'utilisation d'algorithmes génétiques permet l'évolution des profils, et leur renouvellement continu : cela évite de s'enfermer autour de quelques mots-clés, auteurs ou sources connus.

**SMEATON Alan F., KELLEDY Fergus, O'DONNELL Ruairi, QUIGLEY Ian, RICHARDSON Ray, TOWNSEND Emmet (1995)** - « Information retrieval applications : using linguistic resources or using language processing », *Actes des 15èmes Journées Internationales I.A. Génie Linguistique*, 27-30 juin 1995, Montpellier, pp. 65-73.

Filtrage de *news* (projet BORGES). Utilisation de *WordNet* pour enrichir le profil.

**SORENSEN Humphrey, O'RIORDAN Adrian (1995)** - « A learning personalized information filter », *Actes des 15èmes Journées Internationales I.A. Génie Linguistique*, 27-30 juin 1995, Montpellier, pp. 75-85.

**SORENSEN Humphrey, McELIGOT Michael (1995)** - « Representation, adaptation and optimisation in an electronic information filter », *Actes des 15èmes Journées Internationales I.A. Génie Linguistique*, 27-30 juin 1995, Montpellier, pp. 87-95.

Application au filtrage des *news* (messages dans les forums électroniques). La technique utilisée (graphe reprenant tous les liens syntagmatiques) ne permet que de traiter des textes très courts. L'organisation des profils les uns par rapport aux autres est envisagée, afin que les profils s'enrichissent mutuellement tout en respectant leur confidentialité.

**AMATI Gianni, CRESTANI Fabio, UBALDINI Flavio, DE NARDIS Stefano (1997)** - « Probabilistic learning for information filtering », *Proceedings of RIAO'97 « Computer-Assisted Information Searching on Internet »*, 25-27 juin 1997, Montréal, pp. 513-530.

Profils pour la sélection des *news* (messages dans les forums électroniques) ; une même personne peut utiliser plusieurs filtres, un par centre d'intérêt.

### *Systèmes commercialisés :*

**InfoScan**, de Machina Sapiens Inc. (une présentation est fournie dans les *Proceedings of RIAO'97 « Computer-Assisted Information Searching on Internet »*, 25-27 juin 1997, Montréal, pp. 678-679) :

« A personal Assistant that reads and filters your Internet messages » (plaquette *Questions and Answers*)

« InfoScan est un outil de filtrage de l'information en provenance d'Internet. L'utilisateur définit ses champs d'intérêt à l'aide de mots-clés regroupés par thèmes. Les filtres ainsi créés servent à évaluer la pertinence de chaque document. [...]

Le logiciel présente les documents les plus pertinents en utilisant la métaphore du radar, ce qui permet de saisir en un coup d'œil la valeur relative des documents trouvés. »

<sup>10</sup> Le passage de (courts) textes au crible de grilles de lecture est l'objet des célèbres conférences MUC (*Message Understanding Conference*). En France, la journée de l'ATALA du 24 mai 1997 avait pour thème : *Résumé et filtrage automatique de textes* ; y étaient présentées plusieurs expériences de résumé automatique par sélection de phrases, sans grille qui précise a priori l'information à extraire. Sur une initiative franco-tunisienne, des conférences sur le filtrage rassemblent les principaux acteurs de ces approches par extraction des « phrases importantes » : on peut citer la *Rencontre Internationale sur l'Extraction, le Filtrage et le Résumé Automatique*, qui s'est tenue à Sfax, Tunisie, du 11 au 14 novembre 1998.

*Autres acteurs, échos dans les médias :*

Appel d'offre du **Ministère** de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche et du Ministère de l'Industrie, des Postes et des Télécommunications (diffusé sur la liste électronique LN-FR, le 22 avril 1997), intitulé : *Recherche et filtrage de l'information sur les réseaux*.

« L'accès rapide à l'information pertinente, et en particulier à l'information électronique accessible sur les réseaux, est une nécessité pour l'ensemble des secteurs professionnels. Qu'il s'agisse des entreprises du secteur privé ou public, des organismes de recherche et des administrations, l'information joue un rôle croissant dans le processus d'innovation et de compétitivité.

Alors que les nouveaux réseaux de télécommunication à haut débit permettent une diffusion plus rapide de l'information, l'offre de produits informationnels est telle qu'il devient difficile sinon impossible d'identifier l'information pertinente en un temps acceptable pour l'utilisateur.

[...]

Deux catégories d'outils sont concernés : la catégorie des moteurs de recherche et la catégorie des outils appelés 'agents intelligents'.

[...]

Sans devoir répondre nécessairement à l'ensemble des critères suivants, les outils de recherche et de filtrage devront notamment :

- améliorer la phase de formulation de la requête ainsi que la lisibilité et la compréhension des informations retrouvées grâce à une ergonomie efficace ;
- aider l'utilisateur dans la formulation de sa requête (assistance dans la stratégie d'interrogation) ;
- assister l'utilisateur dans la définition de son profil de besoins ;
- relancer automatiquement la requête sans intervention de l'utilisateur ;
- prendre en compte l'hétérogénéité des sources d'informations (serveurs et documents) ;
- prendre en compte la nature multilingue de l'information accessible sur le réseau Internet. Le projet devra nécessairement traiter les requêtes formulées en langue française ;
- orienter les informations retrouvées en fonction des profils d'utilisateurs ;
- classer les informations retrouvées par ordre de pertinence. »

**c) Application de communication évoluée : vue personnalisée sur un ensemble d'informations en circulation**

UN PROFIL EST la représentation des centres d'intérêt, des habitudes et des pratiques d'un destinataire individuel ou collectif, utilisée pour gérer le risque de surcharge informationnelle liée à l'ouverture à un large espace d'échange d'informations.

Le volume peut être géré selon plusieurs modalités : filtrage, organisation en thèmes, hiérarchisation en niveaux de priorité (urgence, importance).

La première application des systèmes de communication évolués est bien sûr la messagerie électronique. Les modalités d'usage sont revues et corrigées : pour l'expéditeur, aide à la rédaction et à la diffusion ciblée ; pour le destinataire, aide au dépouillement de sa boîte aux lettres, au traitement des messages (réponse, réexpédition, impression, enregistrement, extraction de données), et archivage des messages.

L'expérience de la mise en place de systèmes de partage d'informations fait surtout ressortir les points sensibles : mobilisation de l'ensemble des acteurs d'un organisme, tendance naturelle à la rétention d'informations. Les informations mise en commun sont un bien public, pour lequel les personnes sont moins prêtes à s'investir que pour elles-mêmes. D'autre part, l'information est liée au pouvoir et à l'argent : dans un contexte économique peu stable, l'employé voudra se rendre indispensable, par exemple en étant le seul à détenir certaines informations. En outre, pour bien fonctionner, ces systèmes imposent une certaine structuration et standardisation des documents. Cela peut demander un surplus d'efforts, qui n'est bien accepté que si le système apporte des avantages sensibles : accès facilité aux informations dont on a besoin ; reconnaissance explicite, nominative, manifestée d'une façon ou d'une autre par les collègues et la hiérarchie. Les rapports sociaux peuvent avoir leur rôle : on ne veut pas perdre sa réputation ou son crédit en étant accusé de refus de coopération. Les facteurs incitatifs se situent donc du côté de gains personnels directs, récompense sociale, intellectuelle, ou matérielle.



## Repères

### Publications et communications

**TERRY Douglas B. (1990)** - « 7 Steps to a better mail system », *Note Interne Xerox*, Palo Alto Research Center, CSL-90-12, 12 pages.

Aux listes de diffusion préétablies, préférer un système d'archivage (garantissant l'accessibilité à long terme) et de recherche d'informations dans les messages (ciblage en fonction d'une requête exprimée) : le destinataire a l'initiative au lieu de subir un flot de messages. Réalisation d'un prototype (TAPESTRY), 100 utilisateurs, 2000 messages par jour.

**MALONE Thomas W., GRANT Kenneth R., TURBAK Franklyn A., BROBST Stephen A., COHEN Michael D. (1987)** - « Intelligent Information-sharing systems », *Communications of the ACM*, 30 (5), mai 1987, pp. 390-402.

Ajout d'options à une messagerie standard : édition de messages structurés (formulaire), règles de filtrage, envoi pour diffusion ciblée (« To : anyone »). Testé pendant un an par cinq utilisateurs, vient juste d'être mis en place à l'échelle d'un centre de recherche. Quelques problèmes potentiels sont abordés : l'élimination aveugle de messages, qu'il faut sans doute limiter aux messages émis par des listes (par opposition aux messages personnels) ; questions liées à la sécurité et à la confidentialité.

**LUTZ Ernst, v. KLEIST-RETZOW Hans, HOERNIG Karl (1990)** - « MAFIA - An active Mail-Filter-Agent for an Intelligent Document Processing Support », *SIBOIS Bulletin*, 11 (4), December 1990, pp. 16-32.

L'utilisateur peut indiquer des règles s'appliquant aux messages qu'il reçoit : sur une condition (présence de certains mots, type du message reconnu par le système, valeur d'un champ...), différentes actions peuvent être déclenchées, par exemple : « sort message in one or several folder(s) ; direct notification and delivery of urgent message ; reject, print and delete message with or without notifications ; forward messages ; fill in forms with values of the relevant content ; automatic answering service with prepared texts ; initiation of office procedures ; storage in the MULTOS filing and retrieval system. »

**FLEISCHER Robert (1991)** - « Total document control : a text-retrieval perspective », *Proceedings of the Institute of Information Scientists 1990 Text Retrieval Conference*, pp. 78-88.

La visée est de prendre en charge tous les documents, formels et informels, d'un organisme. On favorise la production de documents semi-structurés, qui sont ensuite mieux gérés par le système. Différents projets sont présentés : *Office Filter* (filtrage, diffusion ciblée) ; *Answer Garden* (aiguillage de questions : rapport de bugs), *Mercury* (bibliothèque électronique), *Electronic Information Serveur* (DSI sur le fonds, sur les news ; vue sélective sur l'agenda des événements).

**BARUA Anitesh, RAVINDRAN Suryanarayanan, WHINSTON Andrew B. (1997)** - « Effective intra-organizational information exchange », *Journal of Information Science* 23 (3), pp. 239-248.

Aborde en particulier les résistances individuelles au partage d'informations, et les facteurs incitatifs.

**De SAINT GUILHEM Xavier (1996)** - « Du partage d'informations à l'intelligence collective », *Centraliens* n°475, avril 1996, p. 24.

Expérience de CarnaudMetalbox. « Il s'est [...] produit un effet 'boule de neige'. L'application n'est pas un outil obligatoire, mais un système donnant / donnant : l'utilisateur doit y trouver plus d'informations qu'il n'en met. Au-delà d'un seuil critique, il n'y a plus besoin d'être particulièrement motivé pour utiliser ce système quotidiennement. »

## 4. Profil identificateur et discriminant : repérage d'individus dans une structure

L'ensemble des profils est pris comme base d'information. Ce cas se détache donc bien des trois premiers (profil thématique, profil spécifiant, profil caractérisant et régulateur) car les profils ne sont plus des médiateurs dans l'accès à l'information, ils constituent eux-mêmes l'information, au cœur du système de recherche.

Il convient d'abord d'insister sur le fait que le profil n'est pas la modélisation d'une personne, mais l'angle sous lequel l'application considère celle-ci : par exemple, sa fonction dans l'organisme, ses compétences acquises, son domaine d'activité actuel

La qualité de la base repose sur son exhaustivité, ou du moins sa représentativité, compte tenu de la communauté de personnes qu'elle est censée couvrir. Les profils ne sont donc pas considérés indépendamment les uns des autres comme dans la plupart des applications vues précédemment. Leur positionnement relatif intervient dans leur définition, les points communs comme les spécialités de chacun sont significatives.

### a) Catégorisation : application à l'aiguillage d'information

UN PROFIL EST une catégorie-cible, dans une structure définie de catégories-cibles (référentiel) ; un mécanisme d'appariement associe à tout document soit une seule catégorie, soit une liste de

propositions avec des indications.

L'aiguillage de documents est une tâche complexe quand les destinataires potentiels sont nombreux et changeants. Dans le cas particulier d'une grande entreprise très structurée comme EDF, un *organigramme* détaillé et tenu à jour est l'aide la plus naturelle<sup>11</sup>. Comme l'organigramme dessine une représentation des domaines d'activité et des fonctions, il a deux limites : (i) son niveau de détail est souvent insuffisant (la description s'arrête généralement aux équipes, sans aller jusqu'aux personnes) ; (ii) le découpage qu'il propose n'est pas universel, et correspond à un point de vue particulier ; aussi certains thèmes sont « transverses » et ne trouvent pas de positionnement évident dans cette représentation. Typiquement, des secteurs émergents sont en décalage par rapport à une organisation qui, pour servir de repère, doit garder une relative stabilité. Il est bien connu aussi que les structurations (en groupes, en hiérarchies) rencontrent des individus atypiques, c'est-à-dire qui ne rentrent pas dans leur grille, et qui sont rattachés arbitrairement : ils sont alors « noyés », occultés par la représentation.

L'utilisation d'une base de profils est complémentaire à celle d'un organigramme, en ce qu'elle permet un décloisonnement. Le repérage d'un destinataire peut être fait directement, ce qui permet de gagner en rapidité, d'avoir une meilleure vision de la diffusion de l'information, et de ménager les éventuels intermédiaires qui auraient dû sinon relayer l'information.

## Repères

### *Publications et communications*

**BONNET Alain, PIGAMO Frédéric (1995)** - « Sylex-Icat : Vers un outil générique de filtrage et de catégorisation conceptuelle de textes », *Actes des 15èmes Journées Internationales I.A. Génie Linguistique*, 27-30 juin 1995, Montpellier, pp. 97-106.

Les documents sont confrontés à une structure (arbre, graphe) : pour chaque document on sélectionne un ou plusieurs éléments de la structure. Les liens entre les éléments de la structure interviennent : renforcement mutuel des caractéristiques communes aux éléments voisins (hypothèse de transitivité), et mise en valeur des caractéristiques originales (aspects discriminants). Applications : routage automatique de courriers quand le destinataire est mal défini (la structure est l'organigramme des destinataires potentiels), codage d'actes médicaux par des libellés standards (la structure est alors le référentiel des libellés), indexation thématique d'articles de presse (la structure est un thesaurus).

**FERTON Pascal (1997)** - « Routage personnalisé d'information à Radio-France », Conférence *Documation*, Paris, 18-19 février 1997.

Système rédactionnel des radios locales. Versement en temps réel des dépêches AFP dans le système. Diffusion sur profil aux journalistes, rédaction des scripts en Intranet et accès en Intranet à l'ensemble de l'information produite.

### **b) Référentiel de l'activité : repérage d'interlocuteurs**

UN PROFIL EST une caractérisation de l'activité présente (centres d'intérêts actuels) ou passée (compétence acquise, expertise) d'une personne ou d'une équipe, permettant de l'indiquer comme contact pertinent sur tel ou tel sujet, au sein de l'organisme (communauté scientifique, entreprise, centre de recherche, institution, administration, etc.).

Le profil acquiert ici une certaine épaisseur, en prenant en compte l'historique et la diversité des activités de la personne. Chaque profil peut avoir plus ou moins de points communs avec les autres profils ; cependant, ce sont les particularités propres à chaque profil qui sont mises en valeur, relativement à l'ensemble.

La difficulté que l'on perçoit en filigrane dans les réalisations évoquées est le déploiement et la maintenance du système. Pour avoir une base d'interlocuteurs représentative, il est essentiel de n'avoir pas que les profils des personnes spontanément motivées par un tel système. Il s'agit ensuite d'avoir des descriptions valides et actualisées. Se baser sur des documents descriptifs de l'activité

<sup>11</sup> Disposer d'un organigramme est une situation d'exception, presque à contre-courant de l'individualisme et de la forte mobilité ambiante. Ainsi, la communauté des internautes présents sur le Web mettrait en difficulté une description de ce type.

émis régulièrement (ou sur toute autre trace existante et captable de l'activité), semble à cet égard la meilleure solution : elle a été adoptée dans plusieurs réalisations.

Enfin, il apparaît bien que ces systèmes ne sont qu'un relais, un coup de pouce, qui amplifie mais ne remplace pas les relations interpersonnelles et le bouche à oreille. L'apport pourra être d'étendre les relations vers des secteurs éloignés, où l'on n'a pas de contact initial ; d'aider le nouveau venu à trouver les compétences dont il a besoin.

## Repères

### *Publications et communications*

**HARVEY Clare F., SMITH Peter, LUND Peter, BELL Graeme (1994)** - « The Word-of-Mouth Emulator - Representing expertise and interest in a research and development environment », *Proceedings of the 18<sup>th</sup> International Online Information Meeting*, London, 6 décembre 1994, pp. 423-436.

Chaque personne peut définir elle-même son profil ; le système propose des évolutions en fonction des contributions répétées à un même projet (documents entrés dans le système). Chacun garde le contrôle de la représentation donnée par son profil. Le système n'a pas d'autre prétention que de renforcer ce qui se fait déjà : le bouche-à-oreille. Prototype expérimental.

**STREETER Lynn A., LOCHBAUM Karen E. (1988)** - « Who knows : a system based on automatic representation of semantic structure », *Proceedings of RIAO'88 « User-Oriented Content-Based Text and Image Handling »*, Cambridge MA, 21-24 mars 1988, pp. 380-388.

Les profils décrivent des équipes (cinq à dix personnes) et sont constitués à partir des descriptifs des projets annuels (un à deux documents par profil, de 270 mots en moyenne). La technique d'appariement profil-requête (recherche d'une compétence) est LSI (100 dimensions pour 7100 termes).

**STA Jean-David (1993)** - « Information filtering : a tool for communication between researchers », Communication au Congrès *INTERACT'CHI « Bridges between worlds »*, Amsterdam, 24-29 avril 1993.

Présentation d'une expérience de mise en relation des Départements du centre de recherche EDF sur la base d'un calcul de similarité entre les descriptifs d'activité (« Qui-Fait-Quoi ? »).

**STA Jean-David (1993)** - *Le KIFEQUOI 1993 : expérience de diffusion ciblée des textes d'actions de recherche*, Note interne EDF-DER, HN-46/93/29, mars 1993, accessibilité libre, Clamart : EDF-DER, 23 pages.

Technique de calcul de similarité entre les descriptifs d'activité (outil ADOC), analyse des résultats obtenus, avis des chefs de Département destinataires (validation de chaque proposition et réponse à une enquête).

**FRENCH James C. (1994)** - « DIRE : An approach to improving informal scientific communication », *Information and Decision Technologies* 19, pp. 527-541.

Il s'agit de réunir une masse critique de profils de documents et de personnes. Ensuite, quatre modes opératoires sont envisagés :

DU : documents finding users (selective dissemination of information or passive searching)

UD : users finding documents (retrospective or active searching)

UU : users finding users (une classification génère des « collègues » virtuels ; attention au respect de l'intimité)

DD : documents finding documents (classification, organisation d'archives, spécialisation des sites pour réduire l'espace de recherche).

### *Autres acteurs, échos dans les médias :*

Le « **Qui-Fait-Quoi ?** », à la Direction des Etudes et Recherches d'EDF (cf. les publications de Jean-David STA indiquées ci-dessus).

Le « Qui-Fait-Quoi ? » est un bulletin papier, édité annuellement, qui indique à chaque équipe les équipes qui ont les projets d'activité les plus proches pour l'année à venir. Cela favorise la coordination des équipes à l'échelle du Centre de Recherche (2600 agents).

**CEASCOPE**, au CEA :

Descriptions des compétences des personnes et des activités des services (application de SPIRIT) (cité dans une présentation de SPIRIT, en septembre 1994).

**Telelab**, par le Ministère de l'Éducation Nationale, de la Recherche et de la Technologie :

<http://www.mesr.fr/telelab/telelab.cfm>

Catalogue 7 000 laboratoires de recherche.

**Best Europe**, projet auquel participe le CNRS :

Un questionnaire de sept pages doit permettre de décrire chaque chercheur, et d'alimenter une base de données sur les savoir-faire et compétences en matière de recherche. Informations recueillies : coordonnées professionnelles, cursus (diplômes, postes successifs), compétences linguistiques, thèmes de recherche principaux et annexes, actuels et envisagés, enjeux industriels, brevets, publications, appartenance aux corps professionnels. L'ensemble est renseigné en texte libre ; des mots-clés anglais sont demandés.

## C. ASPECTS CRITIQUES DE CES APPLICATIONS ET POINTS DE DISCUSSION

### 1. Echelle

Si mise en œuvre il y a, le nombre de profils et le volume de documents traités (fréquence, flux) peuvent être déjà des indicateurs de la viabilité du système. Un petit nombre de profils peut traduire une conception maladroite : charge de travail trop grande pour élaborer les profils, mauvaise perception du système... Et sur le plan technique, on rejoint la question de la *scalability*, à savoir la capacité du système à être étendu à de grands volumes de données, sans voir ses performances ou la qualité des traitements se dégrader, voire s'effondrer.

### 2. Durée

Les profils correspondent à des usages dans la durée. Deux phénomènes d'usure peuvent apparaître.

Le premier est lié au profil lui-même et à la technique de représentation. Si par exemple, le système enrichit régulièrement le profil (par *feedback*), sans par ailleurs le rééquilibrer (cohésion d'ensemble, dévalorisation des éléments inactifs ou peu efficaces), le cumul des caractéristiques finit par émousser la précision des attentes représentées, en se dispersant, ou à l'inverse il se focalise toujours plus sur un aspect qui occulte les autres.

Le second phénomène d'usure est la lassitude de l'utilisateur. De nombreux facteurs ont pu concourir à son insatisfaction : mobilisation excessive pour entretenir le profil, gains personnels insuffisants. Cela amène à un diagnostic des principes de fonctionnement du système, et de l'environnement créé autour de sa mise en place : l'échec ne tient pas nécessairement à des raisons techniques.

### 3. Degré de réalisme

Le formalisme ou le modèle choisi pour la représentation de l'utilisateur, compte-tenu de l'application, est quelquefois discutable : c'est d'autant plus facilement le cas que la modélisation est complexe, et cherche davantage à représenter la personne elle-même plutôt que de bien cerner les facteurs directement utiles dans le traitement.

La crédibilité du système tient aussi aux utilisateurs qui l'expérimentent : a-t-on des utilisateurs « tout-venant », ou bien quelques utilisateurs motivés (des étudiants dédommagés, les concepteurs, quelques collègues sympathisants ou curieux d'une nouvelle technologie) ? Ce n'est pas tant le nombre des utilisateurs qui compte ici, que leur représentativité des utilisateurs dans un déploiement de l'application. L'implication et la mobilisation demandées aux utilisateurs peut être trop exigeante ; la persévérance, la patience et la tolérance que peuvent montrer les plus motivés n'est pas généralisable.

### 4. Spécificité du corpus

Certaines réalisations sont très dépendantes de la nature des documents qui sont confrontés aux profils, et ne sont donc pas aisément transposables à d'autres situations.

On a ainsi beaucoup de travaux autour d'un corpus de *news* (forum électronique) : ces annonces sont des documents plutôt courts, et les techniques qui conviennent pour les *news* ne sont plus nécessairement bonnes pour des textes plus développés. La même remarque s'applique aux dépêches et aux courriers de la messagerie électronique.

Les messages éphémères constituent des corpus faciles à rassembler. C'est de l'information déjà sous forme électronique et le volume ne manque pas. Tout un chacun a affaire à ce genre d'information, et peut se sentir concerné par ses possibilités de traitement. En revanche, la priorité de développer des outils de filtrage pour l'information éphémère est discutable. Les pratiques courantes

de communication ont forgé des habitudes de rédaction et de lecture efficaces. En un coup d'œil, le destinataire *prend connaissance* de l'information. Là où les outils peuvent apporter davantage de valeur ajoutée, c'est sur les textes qui seraient *étudiés*, et qui demandent une réelle attention. L'outil prépare et accompagne le travail sur le texte en faisant valoir les apports spécifiques d'un document, en projetant un éclairage inattendu, en donnant des représentations synthétiques ou focalisées, en suggérant des rapprochements, etc.

D'autres travaux se situent explicitement dans un contexte où les documents sont structurés, ou au moins où certains champs sont repérés. La reprise de tels systèmes ne peut être directe : soit on cherche une structuration appropriée pour le nouveau corpus et l'on voit si elle est techniquement, économiquement et socialement possible, soit l'on étudie si le système peut se passer de l'utilisation de la structure en gardant une qualité de traitement acceptable.

## 5. L'assimilation du profil à la personne

Le terme de « profil » est fortement connoté, il est perçu comme beaucoup plus centré sur l'individu (portrait) que sur le système (configuration, instanciation). D'où confusion sur l'ambition de ces modèles : les profils ne sont pas une collection de portraits fidèles de utilisateurs, il est clair qu'on ne saurait prétendre à une description complète et parfaitement juste de la personne. La modélisation est au contraire tout à fait relative à l'application : les variables descriptives et les valeurs que ces variables peuvent prendre ont été choisies parce que jugées significatives et opératoires pour les traitements réalisés.

Cette confusion, quelquefois entretenue (pour mettre en avant le fait que l'on considère personnellement chaque utilisateur) ne fait pourtant pas la publicité du système, car elle touche des points sensibles. La constitution du profil est perçue comme un processus inquisiteur, qui porte atteinte à la *vie privée* : l'utilisateur peut ne pas vouloir confier ses intérêts et préférences, nominativement, à un système, qui les enregistre comme données. Si de plus le mode de constitution et d'utilisation des profils est complexe et échappe à l'utilisateur, cela est ressenti comme une atteinte à sa liberté personnelle : il est essentiel de garder le *contrôle* sur son propre profil. Incidemment, l'idée que ces profils sont représentatifs des personnes introduit un *effet miroir* pas toujours agréable. Le calcul renvoie froidement l'image que le système possède des préférences de l'utilisateur ; celui-ci peut ne pas l'apprécier et vouloir la désavouer.

Enfin, on n'accepte pas que le seul fait d'être associé à un certain profil, parfois grossier, interdise l'accès à des documents, accessibles à d'autres. Le profil est perçu comme un classement. Il « parque » la personne qui en dépend dans un certain domaine d'information, et pose des barrières sournoises, mystérieuses ou infantilisantes.

## 6. Les exigences de la mise en œuvre de l'apprentissage et le caractère adaptatif du profil

Le volume des interactions conditionne la qualité de la description. Un volume de données minimal critique est nécessaire pour qu'il y ait apprentissage. Inversement, le système ne doit pas non plus « dégénérer » avec l'accumulation progressive d'informations.

D'autre part, on peut vouloir ponctuellement être caractérisé très différemment de son profil habituel ; peut-on s'affranchir de ce contexte, sans non plus le perdre ou l'infléchir par des considérations étrangères ? Cette question serait celle de l'*inertie* du profil.

## 7. Gains réels

L'engouement autour de la notion de profil masque quelquefois une réalité décevante. Le cas le plus simple est celui du profil qui n'est qu'une nouvelle façon de désigner une entité qui existait déjà. Il est fréquent qu'un profil ne soit ni plus ni moins qu'une requête documentaire. Ce qui porte plus à conséquence, c'est le profil qui enlève, réduit, plutôt qu'il n'enrichit. S'il réduit le champ balayé à une zone trop étroite, cela va à l'encontre d'une saine curiosité, d'une certaine ouverture d'esprit, de la connaissance utile de son environnement. Le profil se comporte alors comme des

œillères. Tout ceci est affaire de dosage : le profil, en contextualisant le traitement, peut à la fois orienter et enrichir celui-ci.

La visée d'un gain d'ensemble ne peut éluder le gain individuel perçu par les utilisateurs. Les témoignages de mise en œuvre soulignent l'importance d'un « retour sur investissement » sensible ; l'utilisateur n'adoptera le système que s'il y trouve son compte directement et personnellement.

Les propositions, projets, prototypes foisonnent. Quelles en sont les retombées ? Le suivi attentif de la littérature du domaine et du marché des outils commercialisés laisse perplexé : la plupart des présentations, même les plus prometteuses, restent sans lendemains. Dans les cas où la valeur technique et la pertinence stratégique ne sont pas en cause, cela pose la question de l'engagement et de la mobilisation des différents acteurs dans la durée.

## D. LA DIFFUSION CIBLÉE INNOVE : DIX POINTS POUR LA DIFFÉRENCIER DE LA DSI

La Diffusion Ciblée est souvent confondue, à tort, avec la veille documentaire sur une base au moyen d'un profil, dite DSI (*Diffusion Sélective de l'Information*). Or, leur usage s'oppose sur beaucoup de points. La conception du système DECID a été l'occasion de dégager les traits contrastés saillants, éléments de réponse aux demandes de clarification.

### 1. Délimitation de l'espace des documents

#### *a) Usage DSI : Espace documentaire clos*

La base de recherche des documents est définie à l'avance, par exemple : INSPEC, Pascal, fonds documentaire interne... Et en effet, la requête est adaptée aux identifiants propres à la base (code de classement donnant le domaine, types de mots-clefs, etc.)

#### *b) Usage diffusion ciblée : Espace documentaire ouvert*

Théoriquement, tout document apporté doit pouvoir être confronté à la base des profils : l'espace des documents à cibler est ouvert. En pratique cependant, on pourra juger opportun techniquement de se limiter à des types de documents étudiés (ex.: *abstract* d'article scientifique, CV, corpus stratégiques pour l'entreprise).

### 2. Délimitation de l'espace des profils

#### *a) Usage DSI : Espace des profils indéfini*

Le profil est individuel. Il n'est pas défini dans un contexte d'autres profils, par rapport à un ensemble de profils.

Dans la pratique, le nombre des profils est généralement réduit. Dans le cas d'une entreprise, il est bien moindre que l'effectif de ses chercheurs.

#### *b) Usage diffusion ciblée : Espace clos des profils*

Tout profil fait partie de la base des profils (par exemple l'ensemble des chercheurs du centre) et prend sa valeur relativement à elle (cf. sémantique différentielle). Il importe que soit mise en valeur la spécificité de chaque destinataire, pour que l'envoi soit effectivement ciblé.

Le fait que l'espace soit clos n'empêche pas le nombre des profils d'être élevé. Ce nombre est normalement supérieur à la centaine, et il peut tout à fait dépasser le millier.

### 3. Personne qui a l'initiative

#### *a) Usage DSI : Initiative de la part du récepteur*

Le chercheur définit son profil, en général avec l'aide d'un professionnel de la documentation. Il demande que lui soit régulièrement communiqué le signalement des nouveautés de la base correspondant à son profil.

***b) Usage diffusion ciblée : Initiative de la part de l'expéditeur, et de la Direction***

Les profils de l'ensemble des chercheurs qui peuvent / doivent pouvoir être contactés sont construits automatiquement. La personne qui souhaite diffuser un document adresse une requête au Service de Diffusion Ciblée pour obtenir une proposition de liste de diffusion.

Du côté des destinataires, on pourrait parler de « demande latente ».

Un troisième acteur s'introduit, qui n'est ni l'expéditeur, ni le destinataire. Pour mettre en place un service de diffusion ciblée et constituer la base des destinataires, il faut l'intervention d'un instigateur et commanditaire, qui perçoit et confirme l'intérêt global du service pour l'entreprise. La diffusion ciblée suppose donc également une initiative déterminante de la part de la Direction.

**4. Déclenchement d'une utilisation**

***a) Usage DSI : Comportement actif***

Grâce à son profil, le chercheur opère une scrutation active d'une base, il veut capturer certaines informations.

***b) Usage diffusion ciblée : Attitude réactive***

La recherche de destinataires vient souvent en réaction à un stimulus externe : l'expéditeur répercute un document, voire utilise l'outil pour ses propres documents à la demande de ses supérieurs (procédure).

**5. Présomption de qualité et de pertinence de l'information**

***a) Usage DSI : Information brute***

Tout document présent dans la base est susceptible d'être retrouvé. Rien ne donne caution sur sa valeur, sinon le fait qu'il a été entré dans la base –et le choix de telle ou telle base est déjà une étape importante dans la conception de la DSI. Mais les bases sont de qualité inégale, et la ou les base(s) choisie(s) peuvent aussi bien être fiables et bien représentatives du domaine concerné, que le débordant très largement, fâcheusement partiales, ou mal complémentaires. Autrement dit, la présélection de l'information ne se joue qu'à l'échelle d'une base entière.

***b) Usage diffusion ciblée : Information sélectionnée***

Un expéditeur a choisi d'effectuer une diffusion ciblée, il estime donc que le document est digne d'intérêt et susceptible de concerner une partie de l'ensemble des destinataires potentiels. Chaque document a donc initialement une présomption d'intérêt et de qualité vis-à-vis de l'ensemble des destinataires.

Il ne serait pas exclu qu'une base puisse être jugée globalement d'intérêt général. Ce peut être le cas de certaines parties du fonds documentaire interne, par exemple les dernières notes internes reçues, au fur et à mesure de leur parution.

**6. Degré de spécialisation de la recherche**

***a) Usage DSI : Requête focalisée***

Grand soin est pris, d'abord pour cerner le besoin documentaire du chercheur, puis pour formuler au mieux pour le système la question qu'il souhaite résoudre.



### ***b) Usage diffusion ciblée : Profil général***

Le profil ne correspond pas à une demande précise, mais donne un ordre d'idée des activités sur lesquelles le chercheur peut entrer en relation et être contacté (compétences, domaines d'intérêt...). Cependant, plutôt que de mêler indistinctement les différentes facettes de l'activité du chercheur, le profil doit les articuler en préservant l'autonomie relative de chacune. Bien sûr, il n'y a pas non plus d'obligation de retenir toutes les facettes d'activité.

## **7. Dynamique**

### ***a) Usage DSI : Représentation d'une attente entièrement actuelle, pour une durée éventuellement délimitée***

Une question est posée, correspondant à un intérêt particulier du moment. Cette question est statique ; elle peut éventuellement être soit ajustée (correction de dysfonctionnements), soit révisée (versions successives, l'une remplaçant l'autre), soit abandonnée ou complètement renouvelée. L'évolution se présente comme une suite discrète d'états dans la formulation de la requête.

### ***b) Usage diffusion ciblée : Evolution intégratrice et chronologie interne du profil***

Le profil caractérise un destinataire (au plan de ses activités professionnelles), et non pas un point particulier précis de ses travaux. Cette caractérisation doit suivre l'évolution continue des activités et pôles d'intérêt du chercheur, du moins de façon transparente pour l'utilisateur (non-intervention de celui-ci, mise à jour en fonction des nouveaux documents). D'où d'ailleurs dans ce cas une analyse à mener sur les phénomènes de mémoire et d'oubli : comment prendre en compte les acquis et la trajectoire de la personne ? comment rendre inopérants les aspects devenus mineurs et obsolètes ?

## **8. Caractère (im)prévisible de l'information**

### ***a) Usage DSI : Résultats attendus***

Les réponses sont fournies au moment convenu, et correspondent à l'idée que l'on s'en fait *a priori* (plus ou moins, selon la connaissance antécédente de la base et la précision de la requête).

### ***b) Usage diffusion ciblée : Résultats inattendus***

Les profils ne sont pas ajustés à une demande explicitée de document, ils sont en quelque sorte moins subjectifs. Pour l'expéditeur, le document à diffuser est caractérisé par (une partie de) son texte : on constate l'impact, les rapprochements effectués, *a posteriori*. Pour le destinataire, le système apparaît en mesure de renouveler sa problématique en répondant à des questions qu'il ne se posait pas, qu'elles n'aient pas été formalisées voire même qu'elles soient à peine conscientes.

## **9. Equilibre bruit / silence**

### ***a) Usage DSI : Ennemi numéro 1 = le silence***

Le chercheur entend être averti de tous les documents correspondants à sa requête ; il fait une veille systématique sur une base. Il a une exigence de sécurité, quant à la capacité du système à ne pas laisser échapper un document concernant sa demande.

### ***b) Usage diffusion ciblée : Ennemi numéro 1 = le bruit***

Le volumineux courrier déjà en circulation dans l'entreprise fait que l'envoi (voire le signalement) de documents non pertinents est très mal accueilli. En outre, la diffusion ciblée concerne

des documents qui sinon n'auraient pas circulé (ou du moins pas à la même échelle) : il y a toujours un gain malgré l'existence possible de destinataires intéressés mais non repérés par le système.

Ce qui précède est le point de vue du destinataire. L'expéditeur, qui cherche les personnes à qui envoyer un document, peut dans certains cas tenir à trouver toutes les personnes concernées, et donc vouloir avant tout éviter le silence. Or on sait qu'une minimisation du silence documentaire (*i.e.* des personnes concernées mais oubliées) se paie par une augmentation du bruit (*i.e.* de propositions non pertinentes). La solution se trouve dans une démarche interactive. Le système recense aussi exhaustivement que possible les destinataires. Il présente les résultats de la recherche de façon suffisamment claire, organisée et expliquée, pour que l'expéditeur puisse les parcourir efficacement, et ne perde quasiment pas de temps avec les propositions non pertinentes (bruit).

## **10. Possibilités de reformulation**

### ***a) Usage DSI : Interactivité***

Le chercheur est généralement motivé pour ajuster son profil et prendre en compte les résultats déjà observés : intégration d'un feed-back. Cela lui demande un investissement non négligeable : les techniques à base d'apprentissage par exemple requièrent un nombre conséquent d'essais, et d'avis exprimés en retour.

### ***b) Usage diffusion ciblée : Définition automatique à large échelle***

La définition d'une base de profils à grande échelle et homogène rend très coûteuse une procédure de reformulation individuelle des profils. Si ajustements il y a, ils sont du ressort de l'expéditeur. Pour un ciblage précis, des itérations test permettent d'affiner la caractérisation du document et de calibrer les paramètres du ciblage.