

Indexation automatique de fonds bibliothéconomiques

Autor(en): **Clavel, Geneviève / Walther, Frédéric / Walther, Joëlle**

Objektyp: **Article**

Zeitschrift: **Arbido-R : Revue**

Band (Jahr): **8 (1993)**

Heft 1

PDF erstellt am: **30.06.2024**

Persistenter Link: <https://doi.org/10.5169/seals-771689>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern. Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden. Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

spraxis im Sinne einer Humanisierung zu verändern.»
(Stöckle, F.: Zum praktischen Umgang mit Oral History. Aus: Vorlaender, H.: Oral History. S. 156 ff.)

Anschrift der Autorinnen:

Else Bäumlin und Regula Fischer
Stadt- und Universitätsbibliothek Bern
Postfach
3000 Bern 7

Bibliographie

- HALLER, Michael. *Das Interview : ein Handbuch für Journalisten*. – München : Verlag Oelschläger, 1991. – (Reihe Praktischer Journalismus ; Bd. 6)
Das INTERVIEW : Formen, Technik, Auswertung / hrsg. von René König. – 10. Aufl. – Köln : Kiepenheuer und Witsch, 1976. – (Praktische Sozialforschung ; Bd. 1)
ORAL History : mündlich erfragte Geschichte : acht Beiträge / hrsg. von Herwart Vorlaender. – Göttingen : Vandenhoeck und Ruprecht, 1990. – (Kleine Vandenhoeck-Reihe ; 1552)

Indexation automatique de fonds bibliothéconomiques

Geneviève Clavel, Frédéric Walther, Joëlle Walther

Sur la base de communications d'un séminaire consacré aux questions linguistiques de l'indexation automatique, l'apport possible de celle-ci en indexation matières de fonds bibliothéconomiques est évalué sous différents aspects: son utilisation en entreprise, la comparaison entre celle-ci et l'indexation manuelle, un survol historique de la question. La contribution d'un esprit humain s'impose comme une réalité incontournable pour comprendre et restituer le contenu de documents, et on en arrive à considérer l'indexation automatique plutôt comme une assistance à l'indexation manuelle. Tout en souhaitant que le monde des bibliothèques tire profit des recherches et profite des expériences acquises par les industries de la langue, on constate que l'indexation assistée par ordinateur, lorsqu'elle est appliquée à des fonds bibliothéconomiques, se heurte fréquemment à de nombreuses limites.

Aufbauend auf den Akten eines Seminars, das den linguistischen Problemen der automatischen Erschließung gewidmet war, kann der Beitrag der automatischen Indizierung für bibliothekswissenschaftliche Bestände unter verschiedenen Gesichtspunkten evaluiert werden: Nutzung in einem Betrieb, Vergleich zwischen der automatisierten und der manuellen Erschließung, Erarbeitung eines geschichtlichen Überblicks zum Thema.

Es ist jedoch unumgänglich, dass die menschliche Intelligenz den Inhalt eines Dokuments versteht und wiedergibt, wodurch die automatische Indexierung eher als eine Unterstützung der manuellen Erschließung zu werten ist. Auch wenn der Wunsch besteht, dass das Bibliothekswesen Nutzen ziehe aus den Forschungen und von den Erfahrungen der Sprachwissenschaften profitiere, so zeigt sich, dass der vom Computer unterstützten Erschließung bei der Anwendung auf bibliothekswissenschaftliche Bestände Grenzen gesetzt sind.

Sulla base delle comunicazioni a un seminario dedicato alle questioni linguistiche dell'indicizzazione automatica, l'apporto possibile di quest'ultima per ciò che concerne l'indicizzazione delle materie dei fondi biblioteconomici è valutato sotto differenti aspetti: il suo uso nell'impresa, il paragone tra questa e l'indicizzazione manuale, una veduta storica della questione.

Il contributo dello spirito umano si impone come una realtà ineluttabile per capire e restituire il contenuto dei documenti, e questo conduce a considerare l'indicizzazione automatica piuttosto come un aiuto all'indicizzazione manuale. Sperando che il mondo delle biblioteche tragga profitto dalle ricerche e dalle esperienze acquisite dagli studi linguistici, si prende atto che l'indicizzazione assistita dall'ordinatore incontra numerosi limiti quando è applicata a fondi bibliotecari.

1. Introduction

Le séminaire «*Documentation et linguistique*», dont le présent article reprend certaines communications, a été organisé par le Service de formation continue de l'Institut d'Etudes Politiques de Paris (IEP) en novembre 1991. Les quelque 100 séminaires qui y sont organisés chaque année s'adressent aux cadres exerçant des responsabilités dans la vie économique, sociale ou administrative et visent à répondre aux besoins de formation continue des entreprises, collectivités locales et administrations depuis 1974.

Ce séminaire de trois jours proposait une présentation de la problématique générale en indexation automatique de textes, un historique de la recherche dans ce domaine, une analyse théorique de la question et des comptes-rendus de travaux de recherche menés tant dans les milieux universitaires qu'industriels. Dans un second temps, trois grands systèmes d'indexation automatique, qui font appel à des techniques et à des outils de traitement automatique des langues naturelles, étaient présentés, accompagnés d'une démonstration pratique.

2. Approche linguistique

C'est en partant du constat qu'en indexation, l'objet analysé (le texte) et les descripteurs utilisés pour la représentation de celui-ci sont tous deux du ressort de la linguistique et que certains systèmes d'indexation automatique utilisent des techniques de traitement automatique du langage naturel, qu'une approche linguistique semble pertinente.

Nous ne parlerons ici que d'une des méthodes utilisées dans les systèmes d'indexation automatique, celle fondée sur le *modèle linguistique*, alors qu'il existe au moins trois autres approches possibles, le *modèle statistique*, le *modèle probabiliste* et le *modèle procédural*¹.

2.1 Lexique de la terminologie utilisée

Indexation assistée par ordinateur (IAO)

Un autre nom pour l'indexation automatique dans son deuxième sens (cf. ci-dessous), plus proche de la réalité puisque l'indexation automatique ne peut encore remplacer complètement l'intervention humaine.

Indexation automatique

1. En traitement informatique, constitution d'index des mots d'un texte.
2. Tentative de description, par un index, du contenu intellectuel d'un texte, avec l'aide de la linguistique automatique appliquée.

Langage naturel

En informatique, on parle de langage naturel dès lors qu'on considère un langage non seulement du point de

vue informatique, mais également linguistique. Ce langage naturel subit toujours une normalisation ou une modélisation, et n'est jamais la reproduction d'un langage parlé ou écrit, tel que l'utilise normalement un locuteur.

Traduction assistée par ordinateur (TAO)

Domaine du traitement automatique des langues qui a beaucoup de points communs avec l'indexation assistée par ordinateur.

Traitement automatique des langues (ou TAL)

Regroupe toutes les recherches et développements dans le domaine de la linguistique automatique, appliquée aux langues écrites et plus récemment parlées.

3. Problématique

3.1 La manipulation de données textuelles en entreprise

La notion de «texte» est elle-même polysémique, car elle désigne aussi bien l'*objet matériel* que l'*objet conceptuel*, lui-même décomposable en deux parties:

- le *contenu sémantique*, à savoir le thème traité (plus ou moins) objectivement par le texte
- la *signification*, c'est-à-dire l'interprétation qui en est faite en fonction du contexte de communication.

La manipulation de données textuelles peut prendre plusieurs formes:

- au niveau du support physique (archivage, récupération des données, normalisation, etc.);
- du point de vue conceptuel (consultation, hypertexte, extraction de séquences, extraction de phrases pour constituer des résumés ou de termes pour constituer une terminologie, assistance à la création et à la mise à jour d'index, etc.).

L'indexation automatique ne constitue qu'un des nombreux volets du traitement automatique de la langue (TAL): synthèse de textes ou de données factuelles, assistance à la rédaction, contrôle de la production écrite, traduction, assistance à la traduction (TAO). C'est d'ailleurs avec la TAO que l'indexation automatique des données a le plus de points en commun à résoudre: les problèmes de terminologie qui se posent dans les deux cas sont similaires.

¹ Pour plus de précisions, on se référera à l'article de Jacques Chaumier et Martine Déjean: «L'indexation assistée par ordinateur: principes et méthodes» *Documentaliste-Sciences de l'information*, 29, no 1, 1992.

En entreprise, les besoins sont liés à des impératifs de productivité et de rentabilité, qui prennent le pas sur la représentation fidèle des documents. L'accès aux données est donc primordial: idéalement, tout système de repérage des données devrait être à la fois exhaustif et précis; dans la réalité, ces deux critères sont le plus souvent antinomiques. D'ailleurs, il n'y a eu que très peu d'enquêtes d'évaluation des résultats concrets dans un contexte donné.

3.2 Caractérisation de l'indexation automatique par rapport à l'indexation manuelle

Les faiblesses de l'indexation manuelle humaine sont déjà bien connues des spécialistes. L'indexation intellectuelle déléguée à un spécialiste s'avère être une solution performante à condition que le nombre de documents à indexer ne soit pas trop élevé, que les documents à indexer ne soient pas sur support informatique, et que ni l'exhaustivité ni la rapidité du repérage ne soient essentielles.

Dans une grande organisation, le volume des textes à indexer et les modifications fréquentes que ceux-ci subissent rendent l'analyse humaine impraticable.

La subjectivité humaine représente le deuxième obstacle à un transfert impartial de l'information. Les représentations possibles d'un texte sont multiples et la lecture qu'en fait l'indexeur, tout comme la sélection des termes, sont nécessairement subjectives.

En entreprise, il est souvent essentiel, voire vital, de pouvoir répondre à des questions très fines et pointues, qui rendent indispensables la collecte, à partir des textes des documents, d'une infinité de détails. Ceux-ci n'ont parfois que peu de rapports avec le sujet du document en question, mais répondent à des besoins spécifiques de l'entreprise ou de l'employé chargé d'une mission ponctuelle. On peut même aller plus loin en affirmant que les motifs de consultation d'un document ne sont pas les mêmes pendant toute sa durée de vie.

Idéalement, il faudrait disposer d'un système d'analyse de la masse des documents qui offre les différentes approches nécessaires à l'efficacité de l'organisation dans laquelle ce système est en fonction. Dans la pratique, il est absolument irréaliste de continuer à croire qu'un indexeur humain peut remplir ces multiples tâches d'intermédiaire entre le texte et ses utilisateurs potentiels.

En disposant de plus en plus fréquemment de textes sur support informatique, on peut désormais envisager l'accès à l'information qui y est contenue suite à une indexation automatique.

Les termes servant à l'indexation automatique se définissent selon plusieurs caractéristiques: ils sont soit exclusivement unitermes, soit multitermes, choisis dans un thésaurus, dont on fait une mise à jour périodique – ce qui exclut la veille technologique –, ou alors

extraits du texte en langage naturel, ce qui impose malgré tout une gestion des synonymes.

Dans le traitement automatique de la langue existent une tradition statistique, basée sur le nombre d'occurrences et le taux de fréquence, et une méthode linguistique, basée sur la reconnaissance des unités significatives. Une combinaison des deux traitements statistique et linguistique devrait permettre de répondre aux besoins d'indexation automatique.

L'analyse des textes doit se faire à plusieurs niveaux: morphologique, lexicat, syntaxique, sémantique et pragmatique. Pour procéder à de telles analyses, le système doit être doté de modules de reconnaissance. En l'état actuel des recherches, les meilleurs prototypes opérationnels sont bloqués à l'analyse sémantique des phrases.

3.3 Problèmes linguistiques de l'indexation automatique

Niveau morphologique (isolation des termes)

Le premier problème à résoudre est l'*individualisation de chaque terme*. Les séparateurs de mots ne jouent pas toujours ce rôle et peuvent se révéler ambigus: l'espace blanc n'est pas séparateur dans les mots composés, de même que le tiret. Cette compétence linguistique est figée dans la mémoire du locuteur et ne peut être transmise à une machine que par l'intermédiaire d'un dictionnaire qui permet le contrôle des chaînes de caractères et le repérage des mots.

Le découpage en syntagmes nominaux, ou unités significatives, passe aussi par une comparaison avec les enregistrements contenus dans un dictionnaire, le plus souvent spécialisé dans le domaine concerné.

Ex.: //Comité d'experts// en //agronomie biologique//

Niveau lexical (polymorphisme)

Pour pallier les difficultés dues au polymorphisme de mots se rapportant au même concept, il est nécessaire d'en *supprimer les variantes combinatoires* telles que flexion, dérivation, conjugaison, etc., pour obtenir par réduction la forme canonique par lemmatisation. Les outils nécessaires à ce procédé de réduction sont des dictionnaires de correspondance entre forme fléchie (ou dérivée) et forme canonique et des règles d'établissement de correspondances.

La synonymie lexicale linguistique n'est pas la même que la synonymie utilisée en documentation, qui a une acception beaucoup plus large en incluant le concept de terme préférentiel.

3.4 Levée des ambiguïtés

Niveau syntaxique

Les mêmes questions se retrouvent au niveau de la phrase, et particulièrement avec l'utilisation, fréquente, de la paraphrase. Comment donner les moyens à la machine de reconnaître que la phrase

«utilisation des sulfates dans l'agriculture européenne»

contient le même renseignement que

«la CEE fait usage de sulfates en agriculture»

Transformer les syntagmes verbaux en syntagmes nominaux apporte une réponse à ce problème, mais seulement une réponse partielle.

«Elever des enfants»

ne peut simplement pas être transformé en

«élevage d'enfants».

Niveau sémantique

Toute langue naturelle est ambiguë, et cette ambiguïté empêche de choisir automatiquement une simplification à un terme ou à une expression dans tous les cas. L'impuissance des règles syntaxiques peut être illustrée par des phrases telles que

«la belle porte le voile»

«le pilote ferme la porte».

Niveau pragmatique

Pour résoudre les cas de polysémie intervient dans l'esprit humain intelligent le niveau de la connaissance du monde et du contexte. On touche donc là aux limites des systèmes de traitement automatique du langage naturel, et la levée des ambiguïtés restantes donne lieu à diverses tentatives:

- l'exclusion «a priori» de toute polysémie, en créant un champ sémantique nettement circonscrit dans un domaine très pointu, solution qui impose une importante charge de travail préalable;
- la «désambiguïsation» en fonction du contexte, qui n'est possible que si le domaine est très spécialisé;
- l'interactivité du système, qui «passe la main» à l'utilisateur en lui demandant de faire lui-même son choix.

3.5 Fonctionnement des systèmes de traitement automatique de la langue

La méthode générale est commune aux trois systèmes de TAL présentés, ALETH, SPIRIT et DARWIN:

1. On recherche la réponse par essais et erreurs successifs; on pose plusieurs hypothèses dont on retient la plus plausible.
2. L'analyse est décomposée en sous-problèmes successifs qui ne peuvent influencer, en cas d'impasse, les sous-systèmes déjà résolus. Ces sous-systèmes débouchent sur des modules transportables à d'autres applications.

Tous les trois privilégient l'approche linguistique qui permet une adaptation au type d'objet traité, le texte. Cependant, pour adapter le produit «d'indexeur» aux bases de données textuelles et au contexte ou domaines couverts par celles-ci, l'intervention humaine n'est jamais exclue. Bien au contraire, seul un esprit intelligent est capable de reconstituer les réseaux sémantiques nécessaires à l'implantation d'un système d'indexation automatique.

En effet, le traitement automatique de données en langage exclusivement naturel est impossible et passe toujours par des filtres artificiels, qui doivent être fabriqués selon des normes standard, en fonction non seulement du domaine concerné, mais aussi des utilisateurs potentiels.

Avec des enquêtes auprès des utilisateurs, il devrait être possible de définir des profils d'utilisateurs, afin de pouvoir orienter l'indexation automatique selon le type de réponses attendues, et non plus seulement selon le contenu «objectivement restitué» d'un texte.

4. Survol historique et perspectives

Un survol de la progression des recherches en indexation automatique durant les années 60 et 70 permet de replacer les problèmes théoriques de l'indexation automatique dans une continuité historique: il y a une trentaine d'années, on appréhendait ce type de problèmes très globalement, en cumulant différentes approches (mathématique, linguistique, documentaire, informatique) dans une interdisciplinarité incontournable! Cette démarche, après avoir été quelque peu évincée pendant la dernière décennie, est en train de revenir sur le devant de la scène.

4.1 Vers un métalangage universel

Dans les années 60, la conception dominante postulait que l'indexation automatique passait par une traduction du langage naturel vers un métalangage tout aussi général, pouvant être défini par un *lexique*, des *relations sémantiques* entre les termes et une *syntaxe*.

Cette vision générale était héritée de la conception bibliothéconomique universaliste classique, et s'inscrivait dans le prolongement d'outils traditionnels reposant sur des langages universels, à l'exemple de la CDU (Classification décimale universelle). Ainsi, tout langage spécialisé ne constituait théoriquement qu'un sous-ensemble d'un langage plus général.

Dès les années 70, cependant, il a bien fallu admettre que cette analyse globale relevait de la fiction et qu'à des problèmes théoriques généraux, il n'existait aucune réponse générale, mais seulement des *solutions locales dans des contextes appliqués de domaines bien définis et circonscrits*, ce qui ne signifie pas restreints.

Ainsi, tout processus informatisé d'analyse et de recherche documentaire impliquait nécessairement la référence à un (méta)langage de représentation des connaissances dans le domaine visé, d'ailleurs impossible à faire découler d'un langage plus général. Quant à la correspondance entre le langage naturel et le métalangage, elle ne pouvait être automatisée par les seules méthodes de la linguistique, quelles qu'elles soient.

4.2 La fin de l'indexation automatique?

C'est à la même époque qu'est apparue la thèse selon laquelle la puissance des ordinateurs allait bientôt permettre de remplacer l'indexation par des recherches en langage naturel sur des documents en texte intégral. La popularité tenace de cette thèse ne s'est pas démentie jusqu'à nos jours, bien que les exemples ne manquent pas pour en prouver la vacuité: on peut désormais stocker toutes les pages d'un document sur support électronique, mais cela n'est d'aucune utilité si l'on en est réduit à un balayage séquentiel du document faute de critères de recherches performants. Les méthodes de réalisation de tables, index et autres concordances n'ont pas évolué en 30 ans et font partie du bagage commun, mais ce ne sont *ni le même produit ni le même objet qu'une indexation du contenu d'un document.*

4.3 Le retour de l'approche linguistique

Les méthodes d'indexation automatique peuvent être appliquées soit aux documents stockés, soit aux questions, soit encore aux deux, le choix de la stratégie optimale étant fonction du contexte. Deux tactiques linguistiques sont envisageables: la méthode «standard» commence par une analyse morpho-syntaxique, à laquelle elle ajoute un complément sémantique; l'autre méthode repose avant tout sur l'analyse sémantique du domaine visé, dans l'idée qu'«on peut comprendre un texte en langue étrangère spécialisé dans un domaine connu sans connaître la grammaire de cette langue». Quelques expériences d'analyse logico-sémantique ont été tentées dans les années 70, mais elles se sont révélées chères et purement locales (à titre d'exemple, il a fallu 36 personnes-années pour réaliser un réseau sémantique en physiologie basé sur 1000 résumés de textes). Les années 70 marquent donc la fin des recherches purement théoriques dans le domaine de l'indexation automatique, avec le sentiment d'avoir fait le «tour du problème». Cette protohistoire de l'indexation automatique tend à démontrer que le réseau sémantique reste un outil indispensable à une approche linguistique de l'indexation automatique. Ces réseaux sémantiques sont forcément spécialisés, même ceux qui se rapportent aux connaissances dites de sens commun. Les compétences de connaissance de toutes sortes pour construire et entretenir ces réseaux ne peuvent être qualifiées de linguistiques mais bien de spécialisées dans le domaine concerné.

Il est toutefois certain que l'analyse linguistique doit aller de pair avec cet apport provenant de la «connaissance du contexte»; c'est pour cette raison qu'actuellement, les problèmes théoriques de l'indexation automatique sont étudiés dans le cadre de projets appliqués à des domaines particuliers et clairement circonscrits au sein de l'univers général des connaissances.

5. Conclusions

Dans tous les systèmes présentés lors de ce séminaire, les apports linguistiques tout comme les outils cognitifs, représentent toujours des *solutions locales*, certes parfois transportables, mais sous réserve d'adaptations aux contextes divers dans lesquels ils sont en application. En premier lieu, la disparité des *types de documents*, même s'ils sont tous sur support informatique, tout comme celle des *domaines de connaissances couverts*, sans oublier la *diversité des langues utilisées*, sont autant de barrières à une indexation intégralement automatique. Une réponse possible serait de créer des corpus de documents homogènes, mais là déjà, l'intervention humaine s'avère nécessaire. C'est également le cas lorsqu'il s'agit de restituer des notions implicites, face auxquelles un automate se retrouve très démuné.

De ce fait, l'intervention manuelle humaine ne peut être totalement supprimée, même dans la situation idéale d'un corpus de textes écrits dans une même langue, avec un même niveau de spécialisation et traitant un même champ de connaissances délimité. L'homme intervient forcément en amont ou en aval de l'indexation par machine. Il serait plus judicieux de parler d'indexation assistée par ordinateur et de se servir de celle-ci pour un premier filtrage et dégrossissage, apportant ainsi une bonne réponse à la masse de documents à traiter.

L'interprétation d'un texte, et donc aussi son indexation fait appel à des éléments cognitifs qu'actuellement seul un *esprit humain* peut gérer et utiliser à bon escient.

Grâce à cette constatation évidente, on reconnaît peu à peu *l'apport du travail intellectuel de la documentation*, qui ne peut être réduit à une intervention automatique. Jusqu'à très récemment, la modélisation de la tâche d'indexation n'était pas basée sur le travail des indexeurs. Pourtant, comme pour toutes les autres professions dont on a voulu informatiser les tâches, c'est une étape obligée de demander au professionnel de la documentation d'expliquer ses actions, dont celle entre autres de reformuler des textes pour en rendre le contenu accessible à l'utilisateur.

Seuls certains champs de recherche ou de documentation très pointus, nettement circonscrits et fortement réglementés, et dont on possède une masse considérable de documents, peuvent être rendus accessibles par l'indexation automatique à un meilleur rendement que ce qu'on obtiendrait avec une indexation manuelle.

Dans tous les autres cas, l'indexation automatique doit être considérée pour ce qu'elle est: une *assistance à l'indexation humaine*.

5.1 L'indexation automatique et l'indexation des bibliothèques

Des besoins et des objectifs identiques

D'un point de vue général, il ne peut être que bénéfique que le monde bibliothéconomique, traditionnellement plutôt centré sur lui-même, s'ouvre à la recherche et aux pratiques qui sont développées et qui ont cours dans des domaines proches mais trop souvent méconnus.

Depuis quelques années, les industries de la langue prennent leur essor et trouvent un créneau porteur tant dans les milieux administratifs internationaux que dans le domaine industriel. Tous semblent avoir un urgent besoin de se comprendre, comme de pouvoir retrouver, en le comprenant, ce qui a été dit et écrit.

L'indexation alphabétique des matières vise le même but: il est donc fondamental de rassembler nos forces et de se tenir au courant de ce qui se fait dans des domaines ayant les mêmes préoccupations. On peut ainsi échanger des expériences et profiter de celles des autres en s'évitant de répéter les mêmes recherches (et erreurs) coûteuses.

Le contact avec les milieux d'entreprise permet de mieux comprendre quels sont les impératifs du marché et les exigences d'un public qui base celles-ci sur des contingences de marketing, de rendement, d'offre et de demande. Cette ouverture nous donne la possibilité d'améliorer notre perception des besoins des usagers et, par effet de contraste, de prendre davantage conscience que certaines fonctions des bibliothèques/services publics diffèrent de celles du secteur privé, puisqu'elles n'ont pas comme finalité première de «coller» aux lois du marché.

Par des échanges et des discussions entre milieux qui n'ont que peu d'occasions de se côtoyer, il devient possible de comprendre le pourquoi des malentendus, qui ne sont pas toujours perçus comme tels et peuvent de ce fait bloquer tout dialogue productif. L'emploi de terminologie identique, mais avec des sens différents, est révélateur des difficultés de compréhension dont il faut avoir pris conscience pour pouvoir les surmonter. Les exemples de «indexation» ou «langage naturel» sont probants (cf. lexique de la terminologie utilisée, § 2.1.)

Les limites

En écho à ces besoins, les limites inhérentes à tout système d'indexation assistée par ordinateur se dessinent clairement. Première limite, l'indexation automatique n'est conçue que pour le *traitement d'information textuelle*, alors que les bibliothèques et centres de documentation doivent aussi gérer des informations de type

iconographique, sonore, etc. Deuxièmement, les textes à traiter doivent être disponibles sur un *support lisible par machine*, ce qui actuellement est loin d'être le cas pour les fonds documentaires des bibliothèques. La *longueur des textes* est également soumise à des limitations draconiennes, ce qui n'est guère acceptable pour une bibliothèque de conservation comme de recherche. A ces différentes restrictions, il faut ajouter celle du domaine de connaissance. En effet, les quelques systèmes d'indexation automatique actuellement opérationnels ne le sont que dans un cadre extrêmement délimité, pour une application très précise. Ils ne se prêtent donc pas au traitement du savoir interdisciplinaire, voire même parfois encyclopédique, auquel se trouvent souvent confrontées les bibliothèques, et ce d'autant moins dans la perspective d'un travail en réseau bibliothéconomique. Ces différentes contraintes et limites montrent assez nettement l'impossibilité d'envisager à moyen terme une indexation automatique des documents dans les bibliothèques universitaires.

Reste que plusieurs idées mériteraient d'être retenues en vue d'améliorer les systèmes de gestion bibliothéconomique que nous connaissons actuellement. Jusqu'à présent, le travail d'analyse documentaire a surtout porté sur le contenu des documents à indexer, et beaucoup moins sur les demandes concrètes des utilisateurs: il y aurait là pour nous un important effort à effectuer pour permettre une *meilleure adéquation entre ce que demandent les usagers et ce qu'ils reçoivent en retour*. De même, l'idée d'une *présentation des résultats dans l'ordre décroissant de pertinence* semble très prometteuse par rapport au traitement «manichéen» de l'information qu'implique la logique booléenne. Au niveau pratique, certains des modules composant les systèmes d'indexation automatique pourraient s'avérer d'un apport tout à fait intéressant pour les systèmes de gestion bibliothéconomique que nous connaissons actuellement. Le traitement phonétique de la demande de l'utilisateur, par exemple, pourrait augmenter sensiblement le taux de réponse pour les noms propres, aussi bien dans les recherches par auteurs que par matières.

Nous avons constaté que, dans des recherches de pointe sur le traitement de l'écrit, les systèmes les plus performants s'appuient à un moment ou à un autre sur un ensemble de connaissances prédéfini, un métalangage introduit dans le système comme préalable au traitement automatique. Cela se fait le plus souvent au moyen de modules qui ressemblent fort à nos *traditionnels outils documentaires* tels que thésaurus, dictionnaires et autres listes de mots.

Tant pour constituer ces outils que pour les utiliser à bon escient dans un contexte cognitif et culturel, l'*intervention humaine* reste indispensable: l'*apport intel-*

lectuel du spécialiste documentaire se révèle non reproductible automatiquement, et constitue donc une véritable valeur ajoutée au document qu'on souhaite pouvoir retrouver et consulter.

Adresse des auteurs:

Genevieve Clavel
Réseau des bibliothèques utilisant SIBIL
Case postale 478
1000 Lausanne 17

Frédéric Walther
Université de Genève, Service de coordination des bibliothèques
1211 Genève 4

Joëlle Walther
Réseau des bibliothèques romandes et tessinoises
Case postale 461
1000 Lausanne 17

Conserver – Consommer – Eliminer

Vers une politique de conservation différenciée *

Andrea Giovannini

Dans les bibliothèques, on constate depuis plusieurs années une détérioration rapide des papiers imprimés à partir de 1850. Les recherches en matière de conservation des livres sont menées dans quatre directions: désacidification de masse, recherche sur la production actuelle des livres, sur leurs conditions d'utilisation et sur les diverses méthodes de conservation. L'auteur met l'accent sur l'organisation d'une politique de conservation au sein d'une institution donnée, laquelle s'effectue sur la base de quatre constats: conditions de conservation, état de conservation des fonds, fréquence et conditions de consultation des documents. Ensuite, admettant l'idée que tous les documents d'une bibliothèque ne doivent pas être conservés de façon identique, l'auteur propose une structure formelle permettant de distinguer des groupes de documents ayant un niveau de conservation similaire: il s'agit du concept de «conservation différenciée». L'article se termine donc par la description d'une expérience concrète en la matière, à savoir celle qui est en cours à la Bibliothèque cantonale et universitaire de Fribourg, expérience qui ne permet toutefois pas de résoudre tous les problèmes de conservation.

In den Bibliotheken stellt man seit einigen Jahren den raschen Zerfall der Druckpapiere ab 1850 fest. Die Forschungen nach Konservierungsmaterialien für das Buch zielen in vier Richtungen: Massenentsäuerung, Forschungen auf dem Gebiet der gegenwärtigen Buchproduktion sowie auf die Erforschung der Gebrauchsbedingungen und der verschiedenen Konservierungsmethoden. Der Autor legt den Akzent auf die Organisation innerhalb einer gegebenen Institution, welche auf der Basis von vier Tatbeständen arbeitet: Konservierungsbedingungen, Zustand der Konservierung der Bestände, Frequenz und Bedingungen der Benutzung der Dokumente. Die Idee aufnehmend, dass nicht alle Dokumente einer Bibliothek gleichartig konserviert werden müssen, schlägt der Autor ein Schema vor, welches erlaubt, Gruppen von Dokumenten auseinanderzuhalten, die unter ähnlichen Bedingungen zu konservieren sind: es handelt sich also um ein Konzept der differenzierten Konservierung. Der Artikel schliesst mit der Beschreibung eines Experiments auf diesem Gebiet, das an der Kantons- und Universitätsbibliothek Freiburg durchgeführt wird, ein Experiment, das nicht in jedem Fall erlaubt, alle Probleme zu lösen.

Nelle biblioteche si constata da diversi anni un rapido deterioramento degli scritti stampati a partire dal 1850. Le ricerche sulla conservazione dei libri si svolgono in quattro direzioni: deacidificazione di massa, ricerca sull'attuale produzione dei libri, sulle loro condizioni d'uso e sui diversi metodi di conservazione. L'autore pone l'accento sull'organizzazione di una politica di conservazione in seno ad un'istituzione esistente, che si effettua sulla base di quattro osservazioni: condizioni di conservazione, stato di conservazione dei fondi, frequenza e condizioni di consultazione dei

* Conférence prononcée le 27 novembre 1992 à la Bibliothèque cantonale et universitaire de Fribourg