

Besoins et défis en gestion de l'information textuelle

Full-text Management: Requirements and Challenges

Retos y necesidades en la administración de la información textual

Richard Parent

Volume 38, numéro 2, avril-juin 1992

Analyse et gestion de l'information textuelle

URI : <https://id.erudit.org/iderudit/1028616ar>

DOI : <https://doi.org/10.7202/1028616ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

[Découvrir la revue](#)

Citer cet article

Parent, R. (1992). Besoins et défis en gestion de l'information textuelle.

Documentation et bibliothèques, 38(2), 111-114.

<https://doi.org/10.7202/1028616ar>

Résumé de l'article

Dans l'informatisation des bureaux, l'heure d'organiser la gestion des données textuelles est arrivée. Cet article aborde quelques aspects du défi à relever : comment caractériser les besoins au niveau de l'organisation et ceux des individus qui y travaillent ? Quel est le modèle général d'une application en repérage textuel ? Comment combiner l'approche textuelle avec l'approche des systèmes experts ? Quels sont les autres ingrédients d'un système intégré d'information administrative ? Et quelle sorte de changements organisationnels et socio-professionnels semblent souhaitables ?

Besoins et défis en gestion de l'information textuelle

Richard Parent

Direction générale des technologies de l'information
Ministère des Communications, Québec

Dans l'informatisation des bureaux, l'heure d'organiser la gestion des données textuelles est arrivée. Cet article aborde quelques aspects du défi à relever: comment caractériser les besoins au niveau de l'organisation et ceux des individus qui y travaillent? Quel est le modèle général d'une application en repérage textuel? Comment combiner l'approche textuelle avec l'approche des systèmes experts? Quels sont les autres ingrédients d'un système intégré d'information administrative? Et quelle sorte de changements organisationnels et socio-professionnels semblent souhaitables?

Full-text Management: Requirements and Challenges

With office automation in high gear, it would seem that the time has come to organise the management of full-text data. This article identifies the various aspects of the challenge: how to describe the organisation's needs as well as those of the individuals who work for the organisation? What is the general model applied to full-text retrieval? How is the textual approach reconciled with the approach of expert systems? What are the other constituent parts of an integrated administrative information system? What organisational and socio-professional changes appear desirable?

La gestion de l'information textuelle concerne la façon dont les textes sont produits, emmagasinés, indexés, édités, diffusés et consultés au sein d'une organisation. Des années d'utilisation de logiciels de traitement de texte ont permis la constitution de volumes importants de textes qui sont disponibles sur support magnétique. L'idée que cette situation se prête à une exploitation technologiquement assistée se heurte encore à la réticence de beaucoup d'informaticiens qui ont peine à mettre de côté leur paradigme de données codées, structurables en un monde fermé, au sens de la théorie générale des systèmes.

Dans l'administration publique, la gestion de l'information textuelle commence à peine à être organisée pour accroître la part qui est gérée par informatique plutôt que sur papier. Le support papier reste très largement prédominant dans la circulation et la consultation de l'information textuelle.

Retos y necesidades en la administración de la información textual

En la informatización de las oficinas ha llegado el momento de organizar la administración de datos textuales. Este artículo trata ciertos aspectos del reto que esto representa: ¿cómo caracterizar las necesidades a nivel de la administración y de los individuos que trabajan en ella? ¿Cuál es el modelo general de una aplicación de localización textual? ¿Cómo combinar el enfoque textual con el de los sistemas especializados? ¿Cuáles son los otros ingredientes de un sistema integrado de informatización administrativa? ¿Qué clase de cambios organizativos y socioprofesionales parecen deseables?

Il ne s'agit pas de parvenir au légendaire « bureau sans papier ». Les gens préfèrent généralement l'impression sur papier à l'affichage à l'écran pour toute lecture de plus de quelques pages. Des recherches ont d'ailleurs démontré que les lecteurs lisent considérablement plus vite un texte imprimé qu'un texte affiché sur écran lumineux¹. Pour la qualité de vie au travail, on veut donc en avoir du papier imprimé! L'ennui, c'est d'avoir à le classer, afin de pouvoir le retrouver: malgré bien des efforts de classement, on ne voit que piles sur le point de choir, planchers qui ploient, espaces de rangement visiblement trop étroits. Ces signes de surcharge d'information ont un coût élevé dans la mesure où, dans les faits, les employés ne s'y retrouvent que péniblement et incomplètement. Souvent, on ne prend même pas conscience de la situation, mais on choisit de ne pas chercher parce qu'on sait que ce serait trop long. Ces comportements nuisent à la qua-

lité et à la productivité et ils ne pourront changer qu'à condition que des outils viennent faciliter le repérage.

Dans l'administration publique québécoise, un regroupement de ministères et d'organismes a constitué dès 1989 un projet commun, DELTA, afin de se concerter sur le développement de méthodologies pour l'analyse et la gestion des textes et des connaissances. Les articles qui suivent, de Maurice Gingras et d'Yves Rochon, témoignent des réalisations qui participent à ce réseau d'échanges. Le présent article aborde des idées générales rattachées au projet DELTA. Il comporte cinq parties: caractérisation des besoins; modèle d'une application de repérage; combinaison des approches textuelles et cognitives; intégration au système d'information; perspectives d'avenir.

1. Roy Rada, *Hypertext - From Text to Expert-text*, London, McGraw Hill, 1991.

Caractérisation des besoins

Les besoins sont d'une variété quasi illimitée. Parmi les axes d'analyse utiles pour caractériser les besoins, on peut en mentionner deux qui sont d'application générale: les besoins organisationnels et les besoins individuels.

L'aspect statique ou dynamique des textes est important au plan opérationnel et soulève des problèmes de gestion en partie spécifiques à chacun. Du côté statique, on peut penser aux textes qui contribuent à l'encadrement: lois, règlements, directives, jurisprudence, manuels d'interprétation, répertoire de politiques de gestion, conventions collectives, etc. Du côté dynamique, on compte les nombreux actes administratifs en série qui comportent du texte: notes de service, lettres, comptes rendus, convocations, demandes, projets d'étude, mémoires, etc. et formant des flux textuels. Le respect de cette distinction devrait aider une organisation à mieux identifier ses besoins à l'égard de la gestion de l'information textuelle. Il faut prendre appui sur la politique, les procédures et les ressources actuellement associées à la gestion documentaire dans une organisation pour aborder de nouvelles préoccupations notamment:

- la production et l'édition de certains documents selon des formats prédéterminés;
- la constitution et la mise à jour de bases de données textuelles, leur diffusion, leur mode de consultation;
- l'organisation de la sécurité, l'archivage, les réseaux d'échanges de l'information textuelle.

Un deuxième axe important d'analyse est de viser à répondre à des besoins au niveau individuel en termes de:

- consultation: - lecture au complet, pour prendre connaissance, comprendre; - survol par repérage du contenu des textes; - exécution de relevés par l'analyse de texte pour faire des statistiques, choisir des extraits, annoter des modèles.
- production: - création, composition ou rédaction en fonction de modèles; - saisie, formatage, édition.

- ajout: - par annotation sur texte lu de catégories ou commentaires.

Modèle d'une application de repérage

La question du repérage est au cœur de la gestion de l'information textuelle. On peut modéliser une application de repérage au moyen d'un modèle conceptuel composé de trois éléments principaux: un corpus de textes, des utilisateurs, et un vocabulaire pour le repérage.

1. CORPUS:
L'ensemble de textes à interroger, composé d'un nombre plus ou moins élevé de documents.
2. UTILISATEURS:
Des profils d'intérêt pour des parties du corpus, et un degré de connaissance du vocabulaire employé dans les textes du corpus.
3. VOCABULAIRE:
Entre les utilisateurs et le corpus, il y a un élément intermédiaire, le vocabulaire, élément de médiation entre les concepts et expressions connus des utilisateurs et les mots employés dans les textes.

Le vocabulaire est donc vu comme médiateur entre les deux autres éléments du modèle. C'est dans ce vocabulaire que peuvent être formulées des requêtes pour repérer le (les) segment(s) de texte contenant l'information recherchée.

Des méthodologies nouvelles doivent être développées qui soient adaptées aux caractéristiques des données textuelles elles-mêmes ainsi qu'à l'activité mentale que suppose le repérage de l'information textuelle.

En pratique, il est encore difficile d'établir quels développements sont requis et dans quelles circonstances. Il faut toujours constituer le corpus mais l'ampleur des analyses nécessaires en vue de faciliter le repérage est l'objet d'interrogations intenses, et c'est ce questionnement qu'il faut éclairer. Il y a un besoin urgent d'étude comparative du rendement (coûts-bénéfices) de divers modes d'aide au repérage textuel.

En effet, si l'interrogation des mots du texte s'avérait efficace en toute

circonstance, il y aurait peu d'interventions à envisager puisque, avec tous les logiciels, les mots du texte sont organisés automatiquement en un index qui réfère chaque forme à ses occurrences dans le corpus de textes. Il n'y aurait pas de besoin pour le contrôle du vocabulaire, ou pour quelque forme d'indexation.

Cependant, en milieu administratif, certains textes sont si fréquemment consultés, selon quelques dizaines ou centaines de questions répertoriables, qu'il semble avantageux d'effectuer un certain traitement des données textuelles pour augmenter l'efficacité d'un très grand nombre de requêtes.

Cette étude comparative devrait viser à mettre en lumière les facteurs de coût et d'efficacité du développement de trois moyens d'aide au repérage dans le texte, soit: le vocabulaire libre (interrogation par n'importe quel mot du lexique), le vocabulaire contrôlé (interrogation via une liste de termes valués du lexique) et l'indexation des textes (projection sur les segments du texte des éléments du vocabulaire contrôlé).

Les facteurs de coûts, tels le nombre de pages ou de textes, l'homogénéité de ceux-ci, le nombre d'utilisateurs, le temps passé au repérage, etc., et les facteurs d'efficacité, telle la réduction du bruit et du silence documentaires, devraient pouvoir être incorporés dans une équation globale permettant d'évaluer plus facilement, selon les circonstances, la rentabilité des diverses solutions à l'aide au repérage.

Combinaison des approches textuelles et cognitives

Le problème commun des organisations est que les connaissances circulent mal ou peu dans la plupart des organisations. Le défi est de créer des mécanismes permettant le partage de ces connaissances. Pour y arriver, il faut d'abord identifier où elles se trouvent, que ce soit dans des documents, des systèmes informatiques ou dans la tête de spécialistes, et en quoi elles consistent au juste. Ensuite, et c'est là le noeud du problème, il faut concevoir comment on peut les enregistrer, les organiser, les capitaliser et les rendre disponibles et aisément accessibles.

C'est la particularité du projet DELTA de miser sur la complémentarité des méthodologies applicables à l'analyse, à l'indexation et au repérage des textes d'une part, et au développement des systèmes experts d'autre part. L'espoir se fonde sur la dialectique langage et connaissance, particulièrement en ce qui a trait à la terminologie et aux vocabulaires de domaine.

- Sauf pour les « systèmes temps réel » qui font exclusivement de la communication de machine à machine, toute expertise comporte un niveau fondamental de terminologie. Cette terminologie sert à décrire les connaissances, exprimer des faits, présenter des conclusions, etc. et elle peut être gérée à des degrés variables par le mécanisme des messages dans les progiciels de construction de systèmes experts : ces messages, généralement du texte en français, supportent un dialogue personne-machine. (Ces textes sont mis à jour dans le système par simple copie automatisée ou référence directe à un support d'emmagasinage externe au système expert).

- Pour certains systèmes, l'information contenue dans les textes est trop volumineuse, complexe, floue ou ouverte pour être formalisée de façon à permettre un dialogue-à-la-système-expert, plus aisé pour l'utilisateur, mais à un coût de développement onéreux. Le défi est alors de concevoir un accès conceptuel facile et rapide à l'information, d'une façon bien adaptée au contexte de tâches. Des connaissances pointues et/ou générales peuvent être montées sous moteur d'inférence et permettre un dialogue dans ce contexte, mais la réponse du système consiste en segments de texte qu'il faut pouvoir interpréter en contexte.

À mesure que le texte prend de l'importance dans les applications, le dialogue-à-la-système-expert perd de l'importance dans les interfaces et dans l'usage au profit de la navigation dans un thésaurus. La structuration conceptuelle d'un système est ici encore fournie à l'utilisateur par ce moyen, mais sous un mode de plus en plus exigeant quant à la charge cognitive qu'il doit assumer dans l'interaction.

L'importance des textes est d'ailleurs soulignée par Paul Harmon² à propos

des applications « *help desk* ». Il décrit cette classe d'applications comme recourant aux techniques de bases de données, de repérage textuel et de dialogue de diagnostic (ou système expert).

- Nous venons de voir qu'au niveau des applications, les textes peuvent occuper une place importante en combinaison avec l'approche des systèmes experts. Les liens entre l'information textuelle et les connaissances se manifestent d'une autre façon : à l'étape d'acquisition/modélisation initiale des connaissances, en effet, des méthodes d'analyse de texte permettent d'accroître la productivité dans la transformation de la forme information textuelle à des formes de représentation de connaissances aptes à soutenir l'inférence (par héritage, relation de cause à effet).

Cette forme d'analyse est pertinente en fonction de l'existence de textes significatifs pour un domaine de connaissances. L'administration publique, sur un plan pragmatique, est organisée à base de textes : c'est dans l'essence même d'une bureaucratie (au sens non péjoratif) de refléter textuellement sa structure et ses principes d'organisation et d'action. De sorte que le potentiel d'application des systèmes experts y passe largement par les textes, principal dépôt des connaissances à défaut de pouvoir lire dans le cerveau de milliers de personnes.

Dans une méthodologie de développement de systèmes experts, il semble recommandable d'accorder une place très visible à l'exploitation méthodique de l'information textuelle pour le développement d'applications en milieu administratif. On pourrait appeler cela l'exploitation des sources textuelles d'expertise, en complément de l'apport d'experts sélectionnés.

Intégration au système d'information

Les tâches reliées aux textes s'inscrivent dans des enchaînements, administratifs notamment. Dans l'espace électronique, le poids des choses est changé : la complétion de formulaires peut être assistée, l'échange par réseau électronique va de soi pour le routage et l'approbation. Les besoins

de repérage de l'information textuelle sont les plus criants, mais il ne faut pas perdre de vue les autres besoins correspondant aux procédures et tâches routinières de saisie, préparation, transmission et classement.

Comment faire pour ne pas simplement ajouter une couche de coûts sur des façons de faire pensées dans un autre contexte et devenues inadéquates? Il faut être en mesure de réviser la façon dont l'ensemble devrait fonctionner, sans toutefois s'empêcher de commencer par des aspects où l'offre technologique converge avec des besoins perçus dans une nouvelle perspective de fonctionnement.

En pratique, la gestion de l'information textuelle et la modélisation des connaissances doivent être intégrées avec les données structurées ou codées de l'informatique conventionnelle, ainsi qu'avec divers axes d'ajout de valeur en contexte administratif. Un tel essai de mise en relation est fait dans la figure qui présente les ingrédients d'un Système d'information administrative.

Perspectives d'avenir

L'ampleur des changements organisationnels requis pour une intégration, telle que proposée dans la figure, d'îlots existants, constitue un grand défi. Un dialogue plus constructif des utilisateurs et des informaticiens est hautement souhaitable. Mais il ne s'agit que du plus immédiat des conflits socio-professionnels et organisationnels qui ne manqueront pas de se produire lors du déploiement de nouvelles formes de gestion de l'information textuelle. Les exigences d'une productivité accrue dans l'administration publique comme ailleurs poussent irrésistiblement sur cette voie.

Les inévitables arbitrages peuvent être facilités par une participation étendue des personnes concernées au processus de changement. Cette participation à une activité organisée requiert un minimum de formation : c'est le but poursuivi par la diffusion

2. Paul Harmon, « Expert Systems for Help Desk », *Intelligence Software Strategies*, vol. VIII, no. 9 (Septembre 1991), 4-16.

SIA : Système d'information administrative : les ingrédients



- Échange de données informatisées (ÉDI) (SGML)
- Formulaires informatisés
- Courrier électronique
- Routage et approbation électroniques
- Soutien au travail d'équipe
- Accès, sécurité, confidentialité
- Emmagasiner et échange

SGBD textuels
en indexé séquentiel

DONNÉES
TEXTUELLES

- Traitement de texte, éditique
- Analyse de texte
- Indexation et repérage
- Hypertexte

- Vocabulaire, terminologie
- Thésaurus
- Dictionnaire

SGBD relationnels
ou orientés objets

DONNÉES
CODÉES

- Outils de développement
- Outils de modélisation
- Génie logiciel

- Modèle conceptuel référentiel
- Encyclopédie (France : Dicoforme)

Représentation et inférence

CONNAISSANCES

prochaine d'un *Guide d'aide au contrôle du vocabulaire et à l'indexation* produit par les auteurs d'un autre article du présent numéro, soit Suzanne Bertrand-Gastaldy et Gracia Pagola.

Les informaticiens dominent le soutien technique au développement d'applications conventionnelles aux données codées. Cependant, avec les nouveaux projets de type textuel et cognitif, ils doivent laisser place, dans le soutien au développement, aux spécialistes en sciences de l'information, terminologues, lexicographes, ergonomes, cognitiens pour oeuvrer avec des spécialistes des domaines d'application concernés.

À mesure que les développements se multiplieront dans de multiples domaines ainsi que dans divers aspects d'un même domaine, il serait intéressant d'essayer d'organiser une architecture lexicale à l'échelle de l'administration publique. Parmi ces grands éléments, on peut distinguer :

- un niveau sémantique général comprenant le vocabulaire général de la langue française et les terminologies publiées dans des dictionnaires et répertoires ;
- un niveau pragmatique comprenant un tronc commun de vocabulaire dans l'administration publique et des vocabulaires spécifiques aux domaines divers qui sont les objets de cette administration ;
- le cas spécial des noms propres (toponymes, patronymes, sigles, etc.).

Comme c'est le cas avec le génie logiciel, le but d'une telle architecture lexicale serait de permettre à la multiplicité de développements d'applications d'emprunter au maximum aux bases de données lexicales déjà construites en les adaptant et en les spécifiant, de façon à accroître la productivité³.

3. Pour cet article, consulter également : Thomas A. Stewart, « Brain Power », *Fortune* (June 3, 1991), 44-60.