
DIFFUSION CIBLÉE AUTOMATIQUE D'INFORMATIONS

CONCEPTION ET MISE EN ŒUVRE D'UNE LINGUISTIQUE TEXTUELLE POUR LA CARACTÉRISATION DES DESTINATAIRES ET DES DOCUMENTS

Bénédicte Bommier-Pincemin

CNRS-INaLF, équipe Sémantique des Textes
27 rue Damesme, F-75013, Paris, France
Mél : benie@club-internet.fr

EDF-DER,
Dépt. Systèmes d'Information et de Documentation
1, Av. Gal de Gaulle, F-92141 Clamart Cedex, France

1. Présentation

Le serveur DECID (Diffusion Electronique Ciblée d'Informations et de Documents), sur l'Intranet EDF, est utilisé pour trouver les personnes les plus concernées par un document ou les experts sur un sujet. La diffusion ciblée repose sur trois principes fondateurs :

- 1) l'automatisation et la robustesse des traitements,
- 2) le texte, aussi bien comme mode d'interrogation du système (les documents sont soumis tels quels, sans passer par des mots clés), que comme moyen de caractériser des intérêts et compétences des personnes (les profils sont calculés à partir d'un corpus de textes),
- 3) la base formée par les profils des destinataires potentiels : toutes les activités de l'organisme sont systématiquement prises en compte.

Dans le contexte de documents écrits à dominante scientifique et technique, nous définissons quatre facettes textuelles pour guider la conception des traitements :

- 1) la matière linguistique du texte,
- 2) son organisation interne, structurée, close et orientée,
- 3) l'intertextualité,
- 4) le rôle constitutif des lectures et la dynamique de l'interprétation. Ce modèle motive des innovations à toutes les étapes du traitement.

Un format de codage (DTD SGML) est construit, pour être à la fois robuste et général, mais aussi enregistrer des structurations significatives.

Pour la caractérisation des textes, nous proposons des unités descriptives, plus contextuelles que des mots-clés. Elles prennent en compte la détermination du local par le global et la formation des isotopies sémantiques. Ces unités descriptives sont adaptées à la recherche en texte intégral.

Une grille d'analyse est établie et appliquée à l'interprétation des formules de calcul et de mesures sur les textes. Enfin, nous introduisons de nouvelles formes d'interface pour adapter l'ergonomie Web à l'affichage de textes : vue simultanément globale et focalisée, points saillants, contextualisations multiples. Une représentation différentielle de la pertinence facilite le parcours des réponses du système et leur appropriation.

2. Mise en perspective dans le contexte des sciences cognitives

En somme, notre recherche rejoint les sciences cognitives sur plusieurs terrains.

L'expérience recueillie par la mise en œuvre de la diffusion ciblée en entreprise sur plusieurs années conduit à examiner la question de la circulation de l'information en lien avec l'organisation sociale à l'intérieur de l'entreprise, les relations interpersonnelles dans le cadre professionnel, les attitudes face à un nouvel outil de communication, les habitudes et attentes en termes d'ergonomie, la délicate question du rapport des personnes à leur profil... Notre problématique générale a également des liens évidents avec l'étude du *Syndrome de Saturation Cognitive* menée par un groupe de travail de l'ARC (<http://www.mines.u-nancy.fr/~arc/activites/arcindus.html>).

Au plan linguistique, le modèle proposé et les analyses mises en œuvre ne visent pas une plausibilité cognitive "forte", au sens où il s'agirait de mimer les mécanismes de compréhension de l'Homme. Néanmoins, nous nous attachons à respecter dans les traitements des principes sémantiques ayant des répercussions dans l'activité de lecture et de compréhension des textes : détermination du local par le global (la perception d'ensemble d'un texte joue sur la construction de son sens), existence d'effets sémantiques diffus (non localisables et réductibles au niveau du lexique).

Le codage structuré des textes est fondé sur la relation entre la perception visuelle d'une mise en page et les effets interprétatifs pour le lecteur.

Soulignons enfin que nous proposons une typologie des modèles de pertinence actuellement représentés dans les applications documentaires. Nous considérons que la machine ne peut se substituer à l'utilisateur / lecteur, ni déterminer à sa place l'intérêt d'une proposition. Son rôle est plutôt de présenter des propositions et d'accompagner l'utilisateur dans la construction de son interprétation personnelle.