

**École Nationale Supérieure des Sciences
de l'Information et des Bibliothèques**

DEA

Sciences de l'Information et de la Communication

option : Systèmes d'Information Documentaire

MEMOIRE DE DEA

**REPRESENTATION DES STRUCTURES CONCEPTUELLES
DANS LE CADRE DE CARTES NAVIGATIONNELLES**

PHILIPPE VÉROT

RICHARD BOUCHÉ

1995

Résumé :

Dans le cadre du projet PORTULAN qui a pour objectif d'offrir de nouveaux services aux bibliothèques, ce travail soulève des interrogations concernant la conception de cartes navigationnelles. Le contenu et la présentation de ces cartes sont abordés, ainsi que certaines propriétés graphiques susceptibles d'être incorporées dans ces cartes.

Mots clés :

carte navigationnelle; structures conceptuelles; représentation graphique; notice bibliographique; thesaurus; cooccurrence; fisheye view

Abstract :

Within the project PORTULAN which have for objective to offer new services in libraries, this work raise questions about the conception of navigation maps. The content and the presentation of these maps were study, thus that some graphic proprieties likely incorporate in these maps.

Keywords :

navigation map; conceptual strutures; graphic representation; bibliographic record; thesaurus; co-occurrence; fisheye view

*Je tiens à remercier tout particulièrement
Monsieur Richard Bouché pour son aide
précieuse dans l'élaboration de ce mémoire.*

*Je remercie aussi Sylvie Lainé-Cruzel pour ses
lectures attentives et ses critiques constructives*

*Enfin, mes remerciements vont à l'ensemble du
personnel de l'ENSSIB et tout particulièrement
Jacqueline Billard de la bibliothèque, Marie-
Jo Malagola et David Sanchez du centre de
documentation pour leur gentillesse et leur
bonne humeur.*

Table des matières

- 1. INTRODUCTION.....6**
- 2. TYPES DE CONCEPTS A REPRESENTER8**
 - 2.1. REPRESENTATIVITE DES NOTICES BIBLIOGRAPHIQUES8
 - 2.1.1. *Structure et contenu des notices*.....9
 - 2.1.1.1. Structure physique d'une notice.....9
 - 2.1.1.2. Contenu d'une notice.....9
 - 2.1.2. *Le titre*.....12
 - 2.1.3. *Le mot clé*.....16
 - 2.2. RESUMES ET TABLES DES MATIERES17
 - 2.2.1. *Le résumé*17
 - 2.2.2. *La table des matières*.....17
 - 2.3. LE CONCEPT DE COOCCURRENCE19
 - 2.3.1. *Définition*19
 - 2.3.2. *Ajout de termes cooccurrent dans une requête*.....19
 - 2.3.3. *La méthode des mots associés*.....21
 - 2.3.3.1. La mesure du lien entre les mots clés.....21
 - 2.3.3.2. Construction de la carte stratégique22
 - 2.3.3.3. La structuration du domaine23
- 3. TYPES DE RELATIONS A REPRESENTER.....25**
 - 3.1. LES RELATIONS SEMANTIQUES.....25
 - 3.1.1. *Les relations d'équivalence*.....25
 - 3.1.2. *Les relations hiérarchiques*.....25
 - 3.1.3. *Les relations associatives*26
 - 3.1.4. *La relation de définition*26
 - 3.2. LE THESAURUS DYNAMIQUE (H. KIMOTO ET T. IWADERA).....27
 - 3.2.1. *Algorithme de génération des liens*.....28
 - 3.2.2. *Algorithme de calcul du poids d'un noeud*.....29
 - 3.2.3. *Algorithme de génération des mots clés associés*29
 - 3.2.4. *Résultats de l'expérimentation du thesaurus dynamique*.....31
- 4. ERGONOMIE DE LA REPRESENTATION DES STRUCTURES CONCEPTUELLES.....32**
 - 4.1. REPRESENTATION DES NOEUDS.....33
 - 4.1.1. *Représentation du texte à l'intérieur d'un noeud*.....33
 - 4.1.2. *Ergonomie des boites*.....34
 - 4.2. PLACEMENT DES NOEUDS35
 - 4.2.1. *La subordination*35
 - 4.2.2. *Les directions*35

4.2.3. *Un ordre préférentiel*..... 36

4.2.4. *Autres modèles*..... 36

4.3. STYLES DES ARCS..... 37

4.4. REPRESENTATION GRAPHIQUE DE THESAURUS..... 39

4.4.1. *Les schémas graphiques*..... 39

4.4.1.1. Les thesaurus à schémas fléchés (arrowgraphs)..... 39

4.4.1.2. Les schémas circulaires..... 42

4.4.1.3. Les représentations graphiques arborescentes..... 43

4.4.2. *Les tableaux graphiques (terminographs)*..... 44

4.4.3. *Les macro-schémas*..... 45

5. CARACTERISTIQUES GRAPHIQUES PARTICULIERES..... 46

5.1. AFFICHAGES SELECTIFS..... 46

5.2. SCALING : LE CONCEPT DE "FISHEYE VIEW"..... 48

5.3. LE PROCEDE VISUAL RECALL DE XEROX..... 51

6. CONCLUSION..... 52

7. BIBLIOGRAPHIE..... 54

1. Introduction

Ce mémoire se situe dans le cadre du projet européen PORTULAN qui a pour objectif d'offrir de nouveaux services dans les bibliothèques.

Une abondante littérature montre que l'accès aux catalogues des bibliothèques n'est pas satisfaisant et restreint la valorisation des ressources documentaires. D'un autre côté, les bibliothèques offrent un nombre grandissant de services qui diffèrent de la communication de documents habituelle, comme les CD-ROM, les bases de données, les accès à des documents digitalisés, etc. De plus, l'identification de ressources satisfaisantes par l'utilisateur lui-même est souvent difficile.

L'objectif du projet PORTULAN d'offrir de nouveaux services dans les bibliothèques peut être envisagé sous trois aspects :

- ♦ Faciliter l'identification des documents pertinents grâce à une interface permettant une navigation hypertextuelle dans le catalogue. Cette représentation hypertextuelle, sous une forme cartographique, est obtenue grâce aux catalogues existant dans les bibliothèques. Des noeuds sont alors représentés à l'écran, liés entre eux par des relations et une navigation de noeud en noeud est possible. Le catalogue de la bibliothèque est généralisé aux nouveaux types de documents comme les bases de données, les documents numérisés, etc.
- ♦ Faciliter l'identification de documents intéressants par l'intermédiaire de groupes de conférences proposées aux utilisateurs (bibliothécaires ou utilisateurs finaux). Les débats reconnus par les groupes de conférences sont alors considérés comme des documents et sont intégrés au catalogue général. Ils sont alors signalés sur la carte navigationnelle en relation avec les noeuds qui les indexent. La participation des bibliothécaires à ces débats leur permet de mieux connaître l'utilisateur final et d'améliorer les fonctions d'*open learning* de la bibliothèque.
- ♦ Faciliter l'accès en réseau, depuis la bibliothèque, au catalogue d'une autre bibliothèque partenaire possédant cette même interface. Ceci pour deux raisons : l'utilisateur se trouve devant une interface dont il a l'habitude et l'obstacle de la langue est moins important du fait que l'interprétation des noeuds de la carte est induit par son environnement sémantique.

L'objectif de ce travail est de soulever certaines interrogations qui serviront à l'analyse lors de la conception des cartes navigationnelles, puis de présenter des idées, des travaux et des outils existant qui répondent à certaines des interrogations pour les incorporer dans les cartes navigationnelles.

La construction des cartes navigationnelles fait intervenir différentes sources d'information, comme le catalogue d'une bibliothèque, c'est à dire l'ensemble des notices bibliographiques qui servent à décrire les documents de cette bibliothèque, mais aussi les thésaurus, les bases de données ou les textes numérisés.

Il nous a donc paru intéressant de savoir quelles données parmi toutes ses sources d'information seraient à même de décrire, avec pertinence, les documents stockés dans ces sources. Ainsi, quels types de concept et quels types de relation doivent être représentés dans les cartes, sous forme de noeuds et d'arcs, pour permettre une recherche efficace?

Une fois les éléments pertinents sélectionnés, il faut nous intéresser à la façon de les représenter dans la carte navigationnelle, et pour cela étudier les possibilités de représentations graphiques qui nous sont offertes. Ainsi, la représentation des noeuds et des arcs de même que leur placement dans la carte doivent répondre à des contraintes ergonomiques; ces contraintes ayant pour but de créer un affichage clair et structuré.

Enfin, certaines fonctionnalités graphiques doivent être incorporées dans les cartes navigationnelles afin de permettre une meilleure utilisation de ces cartes. Différentes caractéristiques graphiques seront étudiées pour voir si elles sont compatibles avec la représentation des cartes.

2. Types de concepts à représenter

Les cartes navigationnelles intègrent des concepts provenant de sources d'information différentes. Une sélection doit être effectuée pour présenter les concepts décrivant le contenu des documents avec la meilleure pertinence possible. Nous étudierons donc successivement différents concepts provenant de diverses sources d'informations : éléments des notices bibliographiques, résumés contenus dans les bases de données, tables des matières contenus dans les textes numérisés et cooccurrences obtenues par l'intermédiaire d'un thesaurus.

2.1. Représentativité des notices bibliographiques

Les notices bibliographiques peuvent être définies comme "la somme de tous les éléments qui peuvent être utilisés pour décrire, identifier ou rechercher n'importe quelle entité bibliographique (tout document, livre, publication ou autre enregistrement de la communication humaine ; tout groupe de documents ou toute partie d'un document traitée comme une entité)" [Gredley and Hopkinson 1990].

Les éléments ou données bibliographiques qui constituent ces notices n'ont pas tous le même poids informationnel. Ce poids varie, en effet, en fonction de l'analyse documentaire effectuée, de l'utilisateur et de l'opération qu'il compte effectuer sur cet élément. Hagler et Simmons¹ définissent les données bibliographiques comme des "éléments d'information qui aident à définir une partie d'un enregistrement de communication comme objet physique" et identifient trois types de données remarquables par leurs fonctions :

- ♦ données qui identifient uniquement un document particulier pour le distinguer des autres (par exemple, le titre, la date de publication ou le nombre de pages d'un livre).
- ♦ données qui soulignent une association de deux documents ou plus (par exemple, l'auteur commun ou le fait qu'un livre soit une réédition d'un autre).
- ♦ données qui décrivent quelques caractéristiques du contenu sémantique d'un document (par exemple, l'énoncé de son sujet ou une liste de mots clés).

Ce chapitre a pour but de définir les données "pertinentes" des notices bibliographiques qu'il serait intéressant de représenter graphiquement. Pour cela, nous allons nous intéresser plus particulièrement aux données qui décrivent quelques caractéristiques du contenu sémantique d'un document, c'est à dire le troisième type de données selon Hagler et Simmons.

Le choix de ces données doit prendre en compte diverses utilisations et opérations qui sont généralement effectuées lors d'une interrogation et la nature même de l'entité bibliographique (document, livre, image, document audiovisuel, conférence,...).

¹ HAGLER R. and SIMMONS P. - The bibliographic record and information technology - Chicago. American Library Association, 1982.

2.1.1. Structure et contenu des notices

2.1.1.1. Structure physique d'une notice

A l'origine du traitement par ordinateur, la plupart des formats² utilisés étaient en "zones fixes", c'est à dire utilisant des zones de longueurs fixes : chaque donnée avait un nombre précis de caractères. Une amélioration fut apportée par les formats en "zones variables", c'est à dire utilisant des zones de longueurs variables précédées par un séparateur unique ou plusieurs séparateurs simples. Ces deux types de formats ne pouvaient évidemment pas convenir à la complexité d'une notice bibliographique. Il a donc été nécessaire de mettre au point un format en zones variables, identifiées de manière beaucoup plus précise, à l'aide d'étiquettes. Les notices bibliographiques ont trois caractéristiques :

- ♦ les zones sont de longueur variable,
- ♦ la présence de certaines zones est facultative,
- ♦ certaines zones peuvent être répétées.

Physiquement, les notices enregistrées sur support informatique comprennent :

- ♦ Un **guide** composé d'un certain nombre d'informations générales sur la notice, informations rangées dans des positions fixes, permettant aussi bien son traitement informatique que bibliographique.
- ♦ Le **répertoire**, c'est à dire l'index des zones utilisées dans la notice, contenant autant d'éléments que de zones; chaque élément permettant d'identifier la zone et de la repérer en donnant sa position et sa longueur.
- ♦ Le **texte de la notice**, découpé en zones; on trouve d'abord les zones de contrôle (numéro d'identification de la notice, ISBN, etc.), puis les zones des informations bibliographiques.

Ce sont ces zones qui contiennent les données bibliographiques, chacune étant identifiée par un indicateur.

2.1.1.2. Contenu d'une notice

Le contenu d'une notice varie surtout en fonction de la nature de l'entité bibliographique qu'elle représente. En effet, entre un article et une image animée, de nombreux éléments bibliographiques diffèrent. Cependant, certains éléments sont, de part leur nature, présents dans toutes les notices; comme le titre, le créateur (auteur, peintre, imprimeur,....), les mots clés,...

L'exemple suivant nous présente une notice bibliographique du catalogue de la bibliothèque de l'ENSSIB :

| | |
|--------------|-------------|
| OCLC | ocm22113205 |
| > CHAMP_FIXE | |
| Statut | a |
| LANG | Français |
| DATE_TP | s |
| DATES | 1989 |
| > NUMEROS | |
| + ISBN | |
| ISBN_ERRONE | 2867420240 |

² La description est basée sur le format MARC (MACHINE Readable Catalog).

```

SOURCE      DLC
+ Dewey
  Indice  686.2
> COTE_LOCALE
COTE      686.2
  Suffixe  NOD
+ VEDETTES
> NOM_PERSONN
NOM       Nodier, Charles 1780-1844
+ MENTION_TITR
TITRE_PROPRE Critiques de l'imprimerie par le docteur Néophobus
MENTION_RESPO Charles Nodier ; textes choisis et présentés par
Didier Barrière
> ADRESSE_BIBL
+ EDITEUR
LIEU_PUBLICA Paris
EDITEUR_DIFF Editions des Cendres
DATE_PUBLICAT c1989
> COLLATION
NBRE_VOL_PAGE 148 p.
DETAILS_MATER ill.
DIMENSIONS   22 cm
+ COLLECTION
+ COLLECTION
COLLECTION   Collection Technique (Editions des Cendres)
FORME_EDITEE Collection Technique
+ NOTES
> NOTE_SUR_BI
NOTE_DE_BIBL Includes bibliographical references (p. [111]-137) and
index
+ VEDETTES_MAT
+ NOM_COMMUN
NOM         Bibliophilie
+ NOM_COMMUN
NOM         Imprimerie
SUBD_GENERAL Histoire
+ NOM_COMMUN
NOM         Industries graphiques
+ NOM_COMMUN
NOM         Livres
SUBD_GENERAL Mise en page artistique
+ VEDETTES_SEC
+ NOM_PERSONN
INDIC1      1
INDIC2      0
NOM         Barrière, Didier 1956-

```

Comme nous l'avons dit précédemment, le début de la notice contient des zones de contrôle (numéro d'identification, ISBN, etc.) qui servent à l'identification des notices. Les

zones décrivant les caractéristiques physiques de la notice (nombre de page, taille, etc.) ne donnent aucune information sur le contenu sémantique du document.

Les champs à haut pouvoir informationnel sont le titre, les mots clés et leurs subdivisions, ainsi que les personnes ayant un rapport avec ce document.

Nous allons étudier, dans les paragraphes suivants, certains éléments qui, de part leur nature, décrivent le document et sont susceptibles de pouvoir être représentés sous forme graphique.

2.1.2. Le titre

Parmi les éléments descriptifs, le titre apparaît comme la version la plus concise du contenu d'un document, et surtout telle que l'auteur l'a rédigée et telle qu'il a voulu qu'elle soit. Il devient alors un élément documentaire de premier ordre, puisque reflet à la fois du document et de son auteur. La nature du titre est défini comme la "mention, au début d'un livre, du sujet ou de la formule par laquelle l'auteur entend le désigner"³. Cet aspect a conféré au titre une place à part, car aucun autre élément descriptif (mot clé, référence, etc.) ne semble à priori réunir de façon aussi naturelle une telle densité d'informations.

Cependant, le titre d'un document peut être un produit de l'imagination beaucoup plus que de la réflexion, ce qui rend son intérêt très discutable en tant qu'élément d'entrée dans un système documentaire automatisé. Cela, a amené l'UNESCO à rappeler aux auteurs que : "les titres des articles doivent être suffisamment informatifs et précis quant au contenu pour présenter un intérêt pratique dans les listes de titres ainsi que pour le codage aux fins de stockage et de la récupération de l'information."⁴

Nombreux sont les spécialistes et les responsables de services d'information qui ont orienté leurs recherches sur le titre afin d'en connaître la valeur comme base de travail dans un tel système. En effet, au début des traitements informatiques l'espace mémoire alloué aux notices ne permettait que l'enregistrement du titre). Les deux questions autour desquelles paraît s'articuler toute réflexion sur le sujet, sont les suivantes :

- ♦ les informations fournies par le titre sont-elles suffisantes pour permettre l'identification du contenu réel des documents sélectionnés?
- ♦ quel est l'intérêt, la valeur de cet élément à l'entrée du système, ou, en simplifiant, le titre possède-t-il les "mots", les formes qui permettront de retrouver les notions recherchées?

On peut résumer ces objectifs en utilisant les formules : "apport en information" (*Information content*) et "valeur du titre en recherche documentaire" (*Information retrieval capability of titles*).

Lors d'études effectuées sur le titre, différentes méthodes ont été utilisées pour établir la valeur informationnelle des titres, chaque méthode correspondant à une manière différente d'aborder le problème. Dans certaines méthodes, le titre est comparé à des éléments ne figurant pas dans une notice bibliographique (résumé, texte intégral, tables de matières, ...) mais dans des bases de données ou dans des textes numérisés. C'est dans cette optique, que nous avons relaté les méthodes intégrant ces éléments.

PREMIERE METHODE : comparaison entre les mots contenus dans divers produits documentaires.

Cette méthode a pour but de rechercher le meilleur élément d'entrée d'information, donc d'établir le meilleur produit d'analyse documentaire à introduire dans la mémoire du système de traitement.

La première proposition testant la valeur des titres, compare les mots clés choisis lors de l'indexation d'un document aux mots du titre de ce même document. Elle a été faite par

³ Définition extraite du "Dictionnaire Quillet de la Langue française". 1965.

⁴ "Guide pour la rédaction des résumés d'auteurs destinés à la publication" - *Documentaliste*, 1969, 6 (1), p.26.

Maizell⁵ en 1960, puis reprise par Ruhl⁶ en 1964 et enfin par Bottle⁷ en 1970. Cette méthode repose sur le procédure suivante : "les informations apportées par le titre d'un document sont évaluées à la suite d'un comparaison de caractère qualitatif entre les mots présents dans le titre et les mots clés que l'on a jugés utiles d'extraire ou de choisir pour caractériser le contenu de ce même document" [Fondin 1982]. La recherche de correspondance se fait mot par mot.

On établit ainsi une corrélation, un rapport statistique, qui permet d'estimer la valeur relative des titres, selon que :

- ♦ les titres ont tous les mots clés,
- ♦ les titres n'ont pas tous les mots clés,
- ♦ les titres ont des mot-notions pour lesquels ne figure aucun mot clé.

Cette comparaison a très vite abouti à la notion de synonymie ou plus tard à celle de notion voisine. Les personnes ayant effectué cette comparaison, pensaient que cela donnait un résultat intéressant pouvant être amélioré si les titres étaient mieux écrits.

La deuxième proposition liée à cette méthode, élargit la recherche car elle intègre d'autres éléments : les mots clés choisis par l'auteur, le résumé ou même le texte intégral. Comme précédemment, on compare mot à mot en ajoutant les mots synonymes. Cette proposition montre que le titre est un support d'entrée utile mais moins efficace que les autres produits de l'analyse documentaire (mots clés, résumé, etc.). En effet, C. Schultz, W. Schultz et R. Orr⁸ lui préfèrent les mots clés choisis par l'auteur. Tell⁹ estime que les auteurs feraient bien d'écrire des titres qui reflètent le contenu de leurs publications.

Cependant, ces résultats paraissent toutefois discutables car il existe divers problèmes méthodologiques à ces propositions, concernant :

- ♦ la validité de l'analyse documentaire faite par l'homme en tant que critère de comparaison,
- ♦ le choix des mots clés,
- ♦ le nombre de mots clés sélectionnés lors de l'indexation.

DEUXIEME METHODE : comparaison entre différents résultats de recherche.

Cette méthode est apparue assez tôt, dès 1961, et elle a profité du développement des techniques de traitement automatique des documents et de l'accroissement du stockage des données en mémoire. En effet, Hansen¹⁰ travaille dès 1968 sur plus de 250 000 notices. L'approche dans cette méthode est alors indirecte. On n'étudie plus le titre en tant que tel. L'objectif est alors de rechercher le meilleur élément d'entrée dans un système documentaire automatisé en évaluant à la sortie, les performances réciproques des diverses possibilités.

⁵ MAIZELL R.E. - "Value of titles for indexing purposes" - Revue de la documentation, 1960, 27 (3), pp.126-127.

⁶ RUHL M.J. - "Chemical documents and their titles : human concept indexing versus KWIC machine indexing" - American Documentation, 1964, 15 (2), pp.136-141.

⁷ BOTTLE R.T. - "The information content of titles in engineering literature" - IEE Transactions on Engineering Writing and Speech, 1970, EWS-13 (2), pp.41-45.

⁸ SCHULTZ C.K., SCHULTZ W.L. and ORR R.H. - "Comparative indexing : terms supplied by biomedical authors and by document titles" - American Documentation, 16 (4), Oct 1965, p. 299-312.

⁹ TELL B.V. - "Document representation and indexer consistency : a study of indexing from titles, abstracts and full texts using UDC and keywords" - ASIS Proceedings, 6, 1969, p. 285-292.

¹⁰ HANSEN I.B. - "Evaluation of the database CA Condensates compared with Chemical Titles" - Journal of Chemical Documentation, 1972, 9 (4), pp. 201-205.

On comparera donc les références obtenues en réponse à des questions posées au système documentaire en travaillant soit :

- ♦ sur le titre,
- ♦ sur des mots clés libres ou contrôlés, c'est à dire extraits ou non d'un langage documentaire,
- ♦ sur le résumé,
- ♦ sur le texte intégral,
- ♦ sur la notice bibliographique complète.

Ces cinq sources d'informations différentes pouvant, par la suite, être combinées entre elles.

Les études effectuées grâce à cette méthode, nous montrent que l'emploi du titre serait moins efficace que celui des autres éléments, qu'ils soient primaires comme le texte ou secondaires comme le résumé.

Néanmoins, il semble que les résultats obtenus soient basés sur des critères de comparaison bien mouvants. En effet, le nombre de documents récupérés détermine-t-il la qualité ou la pertinence de chaque élément? Sinon, comment est mesurée cette pertinence?

TROISIEME METHODE : recherche du concept-vedette dans le titre.

Il ne s'agit pas ici de comparer deux éléments comme lors de la première méthode. On se borne à constater si le mot servant de vedette-matière, donc de rubrique de classement, se retrouve ou non dans les titres placés sous cette vedette. La base de comparaison est plus simple, car elle repose sur l'établissement de corrélation à partir d'un seul mot-concept. Ce choix accroît sans doute les chances de récupération des titres, mais rend plus que relative l'estimation qui en découlera.

L'appréciation des chercheurs utilisant cette méthode, est largement favorable. Kraft¹¹ et Ghosh¹² considèrent que l'élément "titre" est intéressant pour une recherche rapide d'informations. Cependant, Montgomery et Swanson¹³, malgré les bons résultats obtenus, admettent que le titre pourrait ne pas être suffisant.

QUATRIEME METHODE : ajout de mots clés au titre.

Cette méthode consiste à faire suivre le libellé du titre par des mots clés représentant des concepts ignorés par le titre mais exprimés dans le document et dont la présence est jugée utile par l'analyste. L'étude revient ensuite à relever le nombre de titres pour lesquels l'indication de notions complémentaires a été nécessaire, par rapport à ceux dont le libellé a paru suffisant.

¹¹ KRAFT D.H. - "A comparison of KWIC indexing of titles with a subject heading classification system" - American Documentation, 15 (1), 1964, p. 48-52.

¹² GHOSH J.H. - "Content representation in document titles : a case study with prostaglandin literature" - Aslib Proceedings, 26 (2), 1974, p. 83-86.

¹³ MONTGOMERY C. and SWANSON D.R. - "Machine-like indexing by people" - American Documentation, 13 (4), 1971, p.318-321.

A la lecture des travaux de Bottle¹⁴, les titres seraient bons, voire très bons car on ne relève que 22% d'ajouts de mots clés. Cependant, il faut considérer ces résultats avec prudence. En effet, toute analyse documentaire faite par l'homme pose le problème de la subjectivité de l'indexeur. Certains considéreront comme bon, un titre qui sera jugé insuffisant par d'autres.

En conclusion, si le titre semble être un concentré du contenu du document, le manque de rigueur ou l'imagination des auteurs font que cet élément n'est pas toujours "pertinent". De plus, si on considère une base de données regroupant des articles d'un même domaine, la représentation des titres est concevable; en revanche, la pluralité des domaines abordés dans les bibliothèques publiques rend cette représentation beaucoup plus difficile, à fortiori si elle est faite dans une carte navigationnelle.

En revanche, le titre doit être nécessairement pris en compte dans le cadre d'une recherche par "titre" ou dans le cadre d'une recherche multi-critères. Cela implique, la nécessité d'envisager la représentation d'un titre sous une forme graphique.

De plus, si le titre ne peut apparaître sous sa forme complète, un certain nombre d'organismes conseillent, dans les recommandations pour l'analyse documentaire, de s'inspirer tout simplement du titre et de ses "mots" pour sélectionner les mots clés. Le titre apparaît donc sous une forme décomposée par l'intermédiaire des mots clés.

¹⁴ BOTTLE R.T. - "Titles indexes as alerting services in the chemical and life sciences" - JASIS, 21 (1), 1970, p.16-21.

2.1.3. Le mot clé

Le mot clé est le produit d'une opération - l'indexation - qui consiste à représenter le contenu d'un document par un ou plusieurs mots. Cette représentation doit être effectuée sans aucune ambiguïté, même s'il est vrai que dans un contexte documentaire, à la différence de la vie courante, celles-ci soient en fait peu nombreuses.

En simplifiant, il existe deux grandes catégories :

- ♦ Le *mot clé libre* est celui qui, dans un langage naturel quelconque, exprime une notion contenue dans un document. Il peut être choisi soit arbitrairement par l'indexeur, soit extrait du texte.
- ♦ Le *mot clé contrôlé*, appartient généralement à un outil lexical de contrôle - orthographe, sémantique - du vocabulaire d'indexation.

Le mot clé ne sert pas à conserver l'information, mais il sert à retrouver un document contenant le concept qu'il illustre. C'est une clé de recherche indispensable. Un document contient nécessairement plusieurs concepts, lesquels peuvent être identifiés par la grille d'analyse. Normalement, on devrait choisir autant de mots clés qu'il y a de concepts parfaitement identifiés

La valeur informative du mot clé n'est donc pas à prouver et la représentation des mots clés paraît implicite. De ce fait, aucune autre étude ne sera effectuée sur la représentativité du mot clé.

2.2. Résumés et tables des matières

Nous allons aborder rapidement deux éléments, le résumé et la table des matières, qui font parties de bases de données ou de textes numérisés. Néanmoins, avec l'évolution des techniques informatiques, ces éléments seront sans doute ajoutés aux notices bibliographiques dans les années futures. Que peuvent donc nous apporter le résumé et la table des matières?

2.2.1. Le résumé

Le résumé est défini par la norme française NF 44-044 (1984) comme "un texte concis reflétant fidèlement, sans interprétation ni critique, le contenu d'un document". Son but primordial est "d'aider le lecteur à cerner la pertinence du document vis-à-vis de l'information recherchée". La finalité du résumé peut être décrite en recensant trois buts spécifiques :

- ♦ **Auxiliaire de décision** : le résumé constitue un auxiliaire de décision pour le lecteur et chercheur : "Le résumé doit lui permettre d'identifier rapidement et de manière précise le contenu fondamental d'un document pour décider, s'il y a lieu, de sa lecture complète" (AFNOR Z 44-004).
Dans certains cas le résumé peut se substituer au document lui-même. En effet, il existe des résumés de longueur variable qui poursuivent l'objectif de se substituer aux documents. Le plus souvent rédigés par des services d'information, ces résumés peuvent se présenter sous la forme de compte rendus sommaires.
- ♦ **Auxiliaire de traduction** : Les résumés se rédigent dans la langue de travail de l'organisme qui les éditent ou dans la langue du public visé par les auteurs de documents primaires. Bon nombre de résumés remplissent ainsi une fonction de traduction succincte permettant au lecteur qui ne maîtrise pas la langue du document primaire d'en connaître au moins l'information essentielle.
- ♦ **Auxiliaire de recherche en ligne** : En ce qui concerne la récupération de documents et de références bibliographiques, le résumé ne peut pas être dissocié de recherches dite en "texte libre" (i.e. non indexé). Chaque terme d'un résumé peut servir alors à sélectionner des informations. L'affichage de résumés à l'écran permet en outre, de vérifier la pertinence des références bibliographiques. Cet affichage peut être réalisé dans une fenêtre extérieure à la représentation graphique de la structure conceptuelle proprement dite.

Le résumé prend une place de plus en plus importante dans l'analyse documentaire. La mise à disposition de divers supports informatiques le rend indispensable. En effet, dans le cas de documents audiovisuels, une description sommaire du contenu est souhaitable voire impérative. L'apport informationnel du résumé est donc important.

2.2.2. La table des matières

La table des matières d'un document (scientifique ou technique) est toujours disponible en même temps que ce document et elle est aussi, à priori, toujours rédigée par l'auteur. En effet, si on émet parfois des doutes sur la valeur du résumé effectué par l'auteur, la table des matières est le reflet exact de la structure du document comme l'a voulu l'auteur.

La table des matières est intimement liée au résumé car on peut considérer que les titres des chapitres constituent un "bref résumé" de ces chapitres. La table des matières est donc la sommation de ces brefs résumés, formant ainsi une entité complète décrivant le document.

L'extraction des mots clés d'une table des matières peut contribuer à améliorer le confort et l'efficacité de la recherche documentaire, essentiellement en autorisant des recherches beaucoup plus fines. Les mots clés extraits de la table des matières d'un document sont beaucoup moins nombreux que ceux extraits du texte intégral, ce qui permet de réduire considérablement le bruit, sans pour autant omettre aucune information puisque les notions retenues dans la table des matières sont, à priori, les notions essentielles.

L'utilité des tables des matières peut être envisagée sous différentes formes, quelle soit directe (représentation graphique arborescente,...) ou indirecte (extraction de mots clés pour adjonction à une requête) mais la valeur informationnel de ces tables des matières est en tout cas très élevée.

En conclusion, la pluralité des informations extraites de l'analyse documentaire permet une description détaillée des documents et par la même une recherche pertinente, surtout dans le cas d'une recherche multi-critères. Cependant, Wilson¹⁵ préconise que "maintenant que les catalogues en ligne rendent, en principe l'accessibilité aux informations contenues dans les enregistrements extrêmement facile, de porter les efforts sur la qualité de la représentation du contenu". Il faut augmenter la profondeur de l'analyse et penser à introduire là les résumés ou les tables des matières comme supports pour la recherche.

¹⁵ WILSON P. - "The catalog as access mechanism: background and concepts." - Library Resources and Technical Services, 1983, 27 (1), p.4-17.

2.3. Le concept de cooccurrence

2.3.1. Définition

Au sens littéraire du terme une cooccurrence est définie comme "la présence simultanée de deux ou plusieurs éléments ou classe d'éléments dans un même énoncé". Dans un système de recherche documentaire on parle d'occurrences de mots soit dans la totalité des champs (titre, résumé, descripteurs) des enregistrements, soit seulement dans le champ "descripteurs".

Les fréquences des mots dans le texte d'un article dépendent beaucoup du sujet qu'a choisi l'auteur. En effet, si l'auteur écrit sur un sujet spécifique, il utilise alors des mots avec une fréquence particulière, ce qui conduit à des cooccurrences inhabituelles de certaines paires de mots.

L'utilisation des cooccurrences permet d'associer ou de créer de nouveaux concepts, qui peuvent être ajoutés à ceux préexistants. Cet apport d'information repose sur le principe que la présence répétée de deux termes dans un document définit un concept.

2.3.2. Ajout de termes cooccurrents dans une requête

Un des problèmes majeurs de la recherche en langage naturelle, est la difficulté qu'a l'utilisateur à donner tous les termes qui décrivent le sujet qui l'intéresse. En effet, la requête d'un utilisateur consiste, la plupart du temps en quelques mots en rapport avec son sujet. Il existe alors des procédés, qui en partant des mots de cette requête, ajoutent des synonymes ou une autre orthographe des mots entrés par l'utilisateur. Cette expansion de requête est réalisée à l'aide de thésaurus et d'un contrôle du vocabulaire. Cependant, ces procédés bien qu'extrêmement longs, ont un intérêt technique considérable pour l'identification automatique de paires ou de groupes de mots statistiquement associés.

On peut faire l'hypothèse que, si deux mots apparaissent simultanément dans un même document c'est qu'il existe entre eux un certain rapport sémantique et que cette présence simultanée traduit un concept. Les termes cooccurrents, ainsi obtenus par l'analyseur, peuvent être utilisés pour identifier des relations sémantiques existantes entre ces termes. Ajoutés à la requête, ils sont généralement de bons discriminants et permettent une meilleure recherche des documents pertinents.

Néanmoins, le problème affaissant à cette expansion de requête réside dans le choix des termes à ajouter. Différentes approches ont été testées, prenant en compte le poids, la fréquence des termes ou la distance par rapport à leurs voisins. Helen Peat et Peter Willett ont établis certaines limites à l'utilisation des termes cooccurrents dans l'expansion de la requête :

- ♦ un terme donné ressemble le plus aux termes qui ont une fréquence d'occurrence comparable dans la collection de documents étudiés.
- ♦ les termes de hautes fréquences tendent à ne plus différencier les documents pertinents des documents non pertinents. Ces termes ajoutés à la requête se sont donc pas suffisamment discriminants.

L'utilisation du concept de cooccurrence pour une expansion de requête doit être réalisée dans un domaine de recherche précis pour pouvoir appliquer les restrictions précédentes. Néanmoins, l'expansion des requêtes ne serait être la solution unique à la recherche documentaire, du fait qu'elle engendre une dérive des résultats. En effet, l'augmentation des termes à l'entrée du système renvoie à la sortie une multitude de réponses si on effectue l'union des réponses pour chaque terme, ou aucune réponse si on effectue le croisement des réponses pour chaque terme. Certaines limites doivent donc être appliquées afin d'obtenir des résultats pertinents.

2.3.3. La méthode des mots associés

La méthode des mots associés est une technique de représentation des structures d'informations, contenues dans les bases de données bibliographiques ou dans les bases de brevets. Le principe de la méthode des mots associés et des logiciels Leximappe qui la mettent en oeuvre est simple : "un texte scientifique ou technique quel qu'il soit peut se réduire à l'ensemble des cooccurrences entre les mots qui le composent" [Callon and Courtial and Pénan 1993].

Leximappe est un outil de classification de données documentaire développé par le Centre de Sociologie de l'Innovation des Mines de Paris [Callon and Courtial and Laville 1989] et l'INIST/CNRS [Turner 1991] [Michelet 1988].

La méthode des mots associés repose sur le comptage des cooccurrences de mots indexant les différents documents de la base bibliographique. En comptabilisant toutes ces cooccurrences et en représentant les liens entre mots qu'elles tracent, il est possible d'obtenir un réseau de mots associés.

2.3.3.1. La mesure du lien entre les mots clés

On définit un indice pour mesurer l'intensité relative des cooccurrences de mots clés. Cet indice est obtenu grâce à la méthode d'agrégation dite du saut minimal. Cette méthode construit une hiérarchie de classes d'agglomération d'objets ou groupes d'objets. Les objets sont représentés sous la forme d'un graphe complet valué. Les valeurs des arcs du graphe définissent un indice de dissimilarité.

L'indice de dissimilarité de deux mots i et j est calculé suivant la formule suivante :

$$\frac{C_{ij}^2}{C_i C_j}$$

avec :

- ♦ C_{ij} est le coefficient de cooccurrence, c'est à dire le nombre de document(s) indexé(s) à la fois par le mot i et le mot j .
- ♦ C_i et C_j sont les nombres de documents indexés respectivement par les mots i et j .

Leximappe permet à l'utilisateur de définir un seuil de cooccurrence, tel que tout C_{ij} inférieur à ce seuil annule l'indice de dissimilarité.

"Cet indice de dissimilarité est en fait un cosinus au carré, les mots étant plongés dans un espace dont chaque dimension représente un document. La $n^{\text{ième}}$ coordonnée d'un mot a une valeur égale à un, si le mot indexe le $n^{\text{ième}}$ document, sinon la valeur est nulle. Dans ce cas, C_{ij} est le produit scalaire des vecteurs i et j , C_i (respectivement C_j) étant la norme élevée au carré du vecteur i (respectivement j)." [Georgel 1991].

Après avoir établi un indice mesurant une dissimilarité entre deux objets, il est nécessaire de définir une distance entre groupes d'objets. L'algorithme d'agrégation dit de saut minimal de Leximappe utilise une distance ultramétrique. La distance entre groupes d'objets est alors définie par le plus petit indice de dissimilarité entre deux objets appartenant à chacun des

groupes. "L'algorithme consiste, après avoir créé autant de groupes d'objets que d'objets, à constituer un nouveau groupe en agrégrant les deux groupes les plus proches au sens de la distance ultramétrique. Le processus est répété en tenant compte de ce nouveau groupe." [Georgel 1991].

2.3.3.2. Construction de la carte stratégique

La méthode Leximappe va donc rechercher les sous-graphes connexes (c'est à dire qu'entre deux mots, il existe au moins un chemin) les plus fortement liés, de 10 mots maximum. Pour cela, après avoir calculé l'indice de dissimilarité, les paires de mots clés sont rangées par ordre de dissimilarité croissant. Le programme parcourt alors cette liste depuis le début pour construire des doublets, puis des triplets, etc. de mots associés de façon à fournir un graphe connexe (réseau). L'agrégat ainsi formé est alors retiré de la liste et le programme relance sa procédure.

La carte stratégique est une représentation graphique où les agrégats formés sont positionnés dans un plan définis par deux indices : la centralité et la densité.

- ♦ La **centralité** rend compte pour un agrégat, de l'intensité de ses liens avec d'autres agrégats. Plus ces liens sont nombreux et plus l'agrégat est considéré comme un point de référence pour l'ensemble des recherches entreprises dans le réseau.
- ♦ La **densité** vise à caractériser l'intensité des liens qui unissent les mots qui composent un agrégat donné. Plus ces liens sont forts et plus l'agrégat constitue un ensemble cohérent et intégré.

Ces deux notions permettent de représenter de façon synthétique et simplifiée la morphologie du réseau en rangeant les agrégats horizontalement (suivant l'axe des x) par ordre de centralité croissante, et verticalement (suivant l'axe des y) par ordre de densité croissante. Cette opération permet de classer, à l'aide de valeur médiane, tous les agrégats en quatre catégorie qui correspondent aux quatre quadrants du diagramme (figure 1).

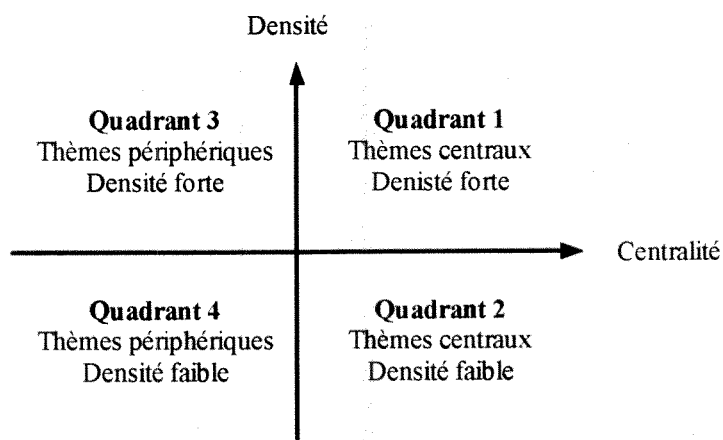


Figure 1

- ♦ Les agrégats de type 1 sont à la fois centraux dans le réseau et général (ils sont fortement connectés à d'autres agrégats) et parcourus par d'intenses liens internes qui manifestent leur fort degré de développement et d'intégration. Ces agrégats constituent en quelque sorte le centre du domaine.
- ♦ Les agrégats de type 2 sont centraux, mais la densité de leurs liens internes est relativement faible. Ils méritent l'attention car ils sont susceptibles de devenir centraux et développés (et par conséquent de se déplacer vers le centre).

- ♦ Les agrégats de type 3 sont peu centraux (qualifier de périphériques) et possèdent une forte densité. Il peut s'agir d'agrégats qui, dans une phase antérieure, étaient centraux mais qui tout en demeurant l'objet d'investissement importants se sont progressivement marginalisés.
- ♦ Les agrégats de type 4 sont à la fois périphériques et peu développés. Ils représentent les marges du réseau.

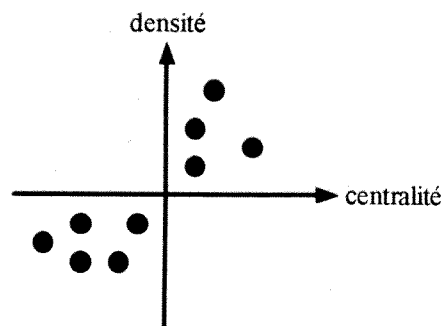
Cette classification des agrégats, visualisée sous la forme d'un diagramme stratégique, fournit une description plus fine de l'état d'un réseau donné, de la position et du degré de développement des thèmes qui le constituent. Le fait que deux agrégats soient proches l'un de l'autre sur le diagramme stratégiques ne signifie pas qu'ils soient liés l'un à l'autre : la seule conclusion qui puisse être tirée de cette observation est que les indices de centralité et de densité ont des valeurs voisines.

2.3.3.3. La structuration du domaine

Le réseau des mots associés permet de qualifier la morphologie d'un domaine de recherche ou d'un secteur techniques. Le diagramme stratégique est un outil très efficace pour fournir une représentation de la structure d'un domaine de recherche. Trois types d'organisation sont envisageables selon Callon, Courtial et Pénan [Callon and Courtial and Pénan 1993] :

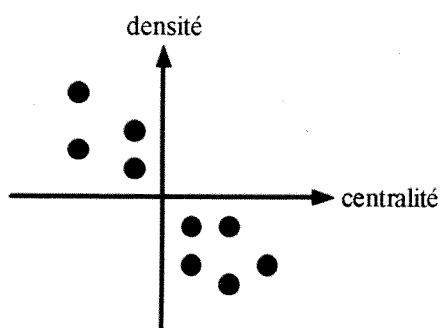
Catégorie 1

Une répartition des thèmes autour de la première bissectrice (quadrant 1/quadrant 4) indique que le domaine s'organise autour d'un noyau de thèmes bien structurés et bien développés, auxquels sont liés une série de thèmes périphériques et peu développés.



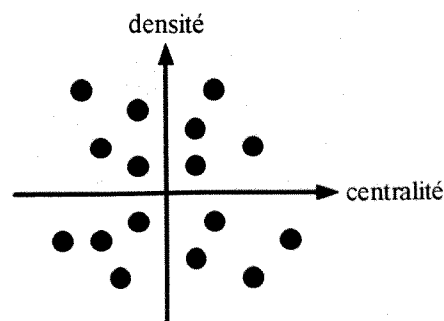
Catégorie 2

Une répartition des thèmes autour de la seconde bissectrice (quadrant 2/quadrant 3) indique un domaine en voie de structuration ou de déstructuration : peu de thèmes sont en position centrale et le réseau est très polycentrique.



Catégorie 3

Une répartition égale des thèmes dans les différents quadrants caractérise un domaine dont la structure est très complexe et riche puisqu'on trouve toutes les familles de thèmes : certains sont centraux et d'autres périphériques, et les divers degrés de développement sont présents. Une telle configuration, qui rassemble toutes les familles possibles d'agrégats (agrégats stabilisés, en émergence ou en cours de développement), suggère une dynamique importante du domaine.



L'intérêt, en ce qui nous concerne, de la méthode des mots associés réside dans l'établissement d'une mesure du lien entre les mots clés. En effet, l'indice de dissimilarité permet de quantifier les liens de cooccurrences existant entre les mots clés. Cette classification, appelée encore classification par simple lien, présente l'avantage sur une classification courante de ne pas forcer les mots à se regrouper en classes les plus homogènes possibles et les plus hétérogènes d'une classe à l'autre.

En reconsidérant la limite de 10 mots maximum pour les agrégats, la présentation d'un agrégat sous la forme d'une carte graphique pourrait donner une représentation intéressante et différente de celle obtenue par un thesaurus. Par exemple, en considérant un mot clé entré par un utilisateur, la représentation des concepts liés à ce mot clé par des indices de dissimilarité élevés permettrait d'obtenir une représentation basée sur la force des cooccurrences par rapport au mot central. Cette représentation devra être étudiée pour définir son intérêt réel et si son implémentation est facilement réalisable.

3. Types de relations à représenter

Les différentes relations à représenter dépendent automatiquement des concepts que l'on veut prendre en compte. En effet, les relations relient les concepts entre eux, elles traduisent le rapport qui les lie.

Les relations peuvent être le résultat d'une attribution humaine ou dérivées de mesures statistiques. Elles peuvent être de type sémantique ou de type syntaxique, et peuvent avoir une direction.

Nous allons donc étudier successivement les relations sémantiques pouvant exister dans un thesaurus, puis les relations de cooccurrences, définies de manière statistique.

3.1. Les relations sémantiques

Les relations sémantiques sont employées, entre autre, pour construire les thesaurus. En effet le thesaurus est un dictionnaire de termes unis par des relations. Ces relations peuvent être regroupées dans les catégories suivantes :

- ♦ relations d'équivalence
- ♦ relations hiérarchiques
- ♦ relations associatives
- ♦ relations de définition

3.1.1. Les relations d'équivalence

Les relations d'équivalence ou relations de substitution renvoient des synonymes ou quasi synonymes vers un descripteur. Ces relations permettent soit :

- ♦ de traduire les termes équivalents en descripteurs
- ♦ de connaître pour un descripteur l'ensemble sémantique qu'il peut recouvrir

Dans le cas d'un thesaurus, ce sont les relations EMPLOYER (USE) et EMPLOYER POUR (USE FOR).

3.1.2. Les relations hiérarchiques

Les relations hiérarchiques servent à exprimer les rapports de supériorité et de subordination entre les concepts exprimés.

Les relations hiérarchiques constituent la structure essentielle du thesaurus autour de laquelle s'organise l'indexation des documents et des questions. Elles se subdivisent en relation générique et relation partitive.

La relation générique

La relation générique, appelée également relation genre/espèce, permet d'organiser autour d'un descripteur la classe dont il fait partie. Cette organisation se fait en terme génériques et en termes spécifiques :

SIEGE Terme générique MEUBLE (Broader Term)

SIEGE Terme spécifique FAUTEUIL (Narrower Term)

Les relations génériques peuvent être à caractère soit monohiérarchique (un terme ne peut avoir qu'un seul père) soit polyhiérarchique (un terme peut avoir plusieurs pères).

La relation partitive

La relation partitive ou relation tout/partie est utilisée pour découper un concept en ses constituants, c'est à dire l'entité et ses parties.

3.1.3. Les relations associatives

Les relations associatives ou relations de voisinage servent à indiquer les analogies pouvant exister entre deux descripteurs ou termes associés.

Les types de relations associatives sont fort nombreuses. Parmi celles-ci on peut noter les relations Objet/Propriété, Objet/Objet, Objet/Application, Propriété/Processus, etc.

La relation associative est souvent utilisée pour éviter l'emploi de la polyhiérarchie.

3.1.4. La relation de définition

La relation de définition est introduite par la Note d'usage ou Note d'application dans un thesaurus. Cette relation est utilisée pour :

- ♦ limiter l'emploi d'un descripteur,
- ♦ expliciter une abréviation,
- ♦ exclure le possible d'un terme.

Par exemple : FICHER (à n'utiliser que pour les fichiers sur supports magnétiques).

Ces différentes relations créent une hiérarchie entre les concepts, permettant de structurer l'information qui traduisent.

3.2. Le thesaurus dynamique (H. Kimoto et T. Iwadera)

Kimoto et Iwadera¹⁶, dans le cadre de la recherche documentaire, ont développé un nouveau système appelé AIRS (Associated Information Retrieval System) utilisant une approche connexioniste. Le thesaurus dynamique est composé de noeuds, qui représentent chaque terme du thesaurus statique, et de liens, qui représentent les connexions entre les noeuds.

La caractéristique distinctive du système AIRS réside dans le fait qu'il détermine les intérêts des utilisateurs. En effet, chaque utilisateur définit un échantillon de documents qu'il considère pertinent. Une information sur chaque terme, automatiquement extraite de l'échantillon de documents, est alors utilisée pour changer le poids des noeuds, générer de nouveaux liens, grâce en particulier aux cooccurrences, et construire le thesaurus dynamique qui génère des mots clés associés. Ces mots clés associés sont alors utilisés pour rechercher des documents qui regroupent précisément les intérêts des utilisateurs.

L'information sur le terme comprend des informations sur les mots clés et des informations sur les relations entre mots clés :

- ♦ L'information sur un mot clé correspond à des mots clés et leur rang, et est rangée en fonction de l'importance de ce mot clé dans un échantillon particulier.
- ♦ L'information sur la relation entre mots clés correspond au type de la relation et à sa force.

Le thesaurus dynamique est construit selon la structure du réseau. Chaque noeud du réseau, qui possède un poids, représente un terme du thesaurus. Chaque lien représente une relation entre les termes. Le poids d'un noeud est calculé en utilisant le rang du mot clé contenu dans l'information sur le terme. Il existe cinq types de relation entre les noeuds :

1. relation générique
2. relation spécifique
3. relation d'usage
4. relation de synonymie
5. relation de cooccurrence

Les relations 1, 2, 3 et 4 sont obtenues par le thesaurus statique. La relation 5, la relation de cooccurrence, est obtenue par les relations des mots clés contenus dans l'information sur le terme. La relation de cooccurrence est définie comme la relation des paires de mots clés qui cooccurrent dans l'échantillon de documents.

L'utilisation des relations de cooccurrence dans le thesaurus dynamique permet de personnaliser le thesaurus en modifiant les poids des noeuds et les liens. Le système AIRS utilise le thesaurus dynamique pour générer des mots clés associés au mot clé rentré par l'utilisateur lors de sa requête.

¹⁶ KIMOTO H. and IWADERA T. - "Construction of a dynamic thesaurus and its use for associated information retrieval" - 13th International Conference on research and development in information retrieval, 1990, pp.227-240.

Les effets attendus du système AIRS sont les suivants :

- ♦ Les mots clés associés, qui reflètent l'intérêt des utilisateurs, sont générés par le thesaurus dynamique.
- ♦ Un taux de rappel élevé et un taux de précision élevé sont obtenus en utilisant ces mots associés pour une recherche documentaire.
- ♦ Il est possible d'utiliser les transitions d'état du thesaurus dynamique pour mettre en évidence le changement d'intérêt d'un utilisateur au cours de la recherche. Ainsi, la recherche documentaire reflète les intérêts principaux de l'utilisateur.

Le système AIRS comprend quatre phases :

1. L'information sur le terme est extraite automatiquement de l'échantillon de documents. Les mots clés, le rang des mots clés et les cooccurrences entre mots clés sont extraits en utilisant le système INDEXER¹⁷.
2. Le thesaurus statique ou traditionnel est modifié par le terme d'information pour former le thesaurus dynamique. Des liens sont générés et le poids des noeuds est calculé pendant que le thesaurus dynamique est fabriqué.
3. Les mots clés associés sont générés à partir du mot entré par l'utilisateur, lors de sa requête, grâce au thesaurus dynamique. Le thesaurus dynamique commence avec le mot entré, puis avec les mots associés sélectionnés basés sur le noeud du poids et ses liens.
4. Les documents sont recherchés en utilisant les mots associés. Les documents obtenus sont classés en utilisant les informations du thesaurus dynamique.

3.2.1. Algorithme de génération des liens

Si deux mots clés apparaissent dans un document, un lien est alors généré entre les deux noeuds correspondant à ces mots clés, si aucun lien n'existe préalablement entre ces noeuds. Ainsi, si les noeuds A, B et C cooccurrent dans un document, on obtient alors la configuration suivante :

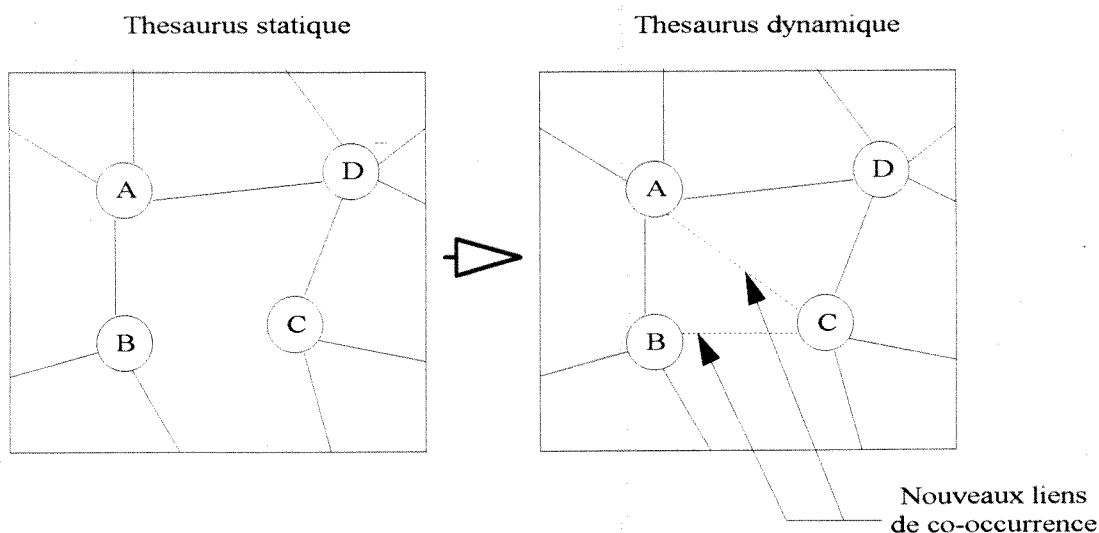


Figure 2

¹⁷ KIMOTO H. and NAGATA M. and KAWAI A. - "Automatic Indexing System for Japanese Text" - Review of the Electrical Communications Laboratories, 1989, 37 (1), pp.51-56.

3.2.2. Algorithme de calcul du poids d'un noeud

Le poids d'un noeud reflète l'importance des mots clés extraits de l'échantillon de documents. L'importance des mots clés est calculée par le système INDEXER, qui extrait et range les mots clés pour chaque échantillon. Le rang affecté au noeud est fonction de la fréquence et de la situation dans ce document. On définit donc :

D : ensemble complet des documents

T_i : document individuel

K_i : ensemble des mots clés extraits de T_i

KW_{ij} : mots clés individuels

n : nombre de documents

m : nombre de mots clés dans un document T_i

soit

D = {T_i} avec i=1,...,n et

K_i = {KW_{ij}} avec j=1,...,m

L'importance de KW_{ij} par rapport à T_i est définie par KI(ij). La valeur de KI(ij) diminue linéairement quand J (le rang) augmente, et la somme des valeurs de KI(ij) (j=1 à m) est égale à 1 (pour chaque i) afin de normaliser l'importance. Plus la valeur du mot clé est proche de 1, au plus le rang du mot clé est important.

KI(ij) est calculé par la formule suivante :

$$KI(ij) = \frac{2TW_i}{m*(m+1)} * (m+1-j)$$

où TW_i est la valeur donnée à T_i dans D et j est le rang de KW_{ij} dans T_i.

TW_i est calculé par la formule suivante :

$$TW_i = \frac{DW}{n} \text{ (constant)}$$

où DW est la valeur de D donnée par un utilisateur.

Après KI(ij) est calculé pour chaque T_i, le poids du noeud l noté NW_l est calculé comme la somme des KI(ij) dans D.

3.2.3. Algorithme de génération des mots clés associés

Les mots clés associés ont pour but d'étendre le mot clé rentré par l'utilisateur lors de sa requête. Les mots clés associés sont obtenus en parcourant les noeuds et liens du thesaurus dynamique en commençant par le noeud correspondant au terme entré par l'utilisateur. Par la suite, le noeud correspondant au terme entré par l'utilisateur sera appelé "noeud de départ", et l'ensemble de noeuds et de liens parcourus pendant le processus de génération des mots clés sera appelé "chemin de génération". La distance de parcours est définie comme le nombre de liens traversés pour générer les mots associés.

La procédure de génération des mots associés du système AIRS est définie par les phases suivantes :

1. La distance de parcours et le type de liens à parcourir sont prédéfinis. La valeur seuil du poids des noeuds est prédéfinie pour les noeuds sélectionnés. La distance entre deux noeuds dans le thesaurus dynamique est définie comme le nombre de liens entre ces deux noeuds.
2. En partant du noeud de départ, les liens acceptables sont parcourus dans la limite de la distance prédéfinie.
3. Tous les noeuds du chemin de génération deviennent des mots clés associés potentiels.
4. Parmi les mots clés associés potentiels, seulement ceux qui ont un poids supérieur à la valeur seuil sont effectivement sélectionnés pour être des mots clés associés.

Un exemple de génération de mots clés associés est donné dans la figure suivante. On considère dans cet exemple que la distance de parcours est égale à 3, que tous les liens peuvent être parcourus, et que le noeud de départ est A. Un des chemins de génération est composé, par exemple, des noeuds A, B, C et D qui deviennent alors tous les quatre des mots clés associés potentiels. Finalement, seuls les noeuds A et D, qui ont un poids supérieur à la valeur seuil, seront choisis comme mots clés associés.

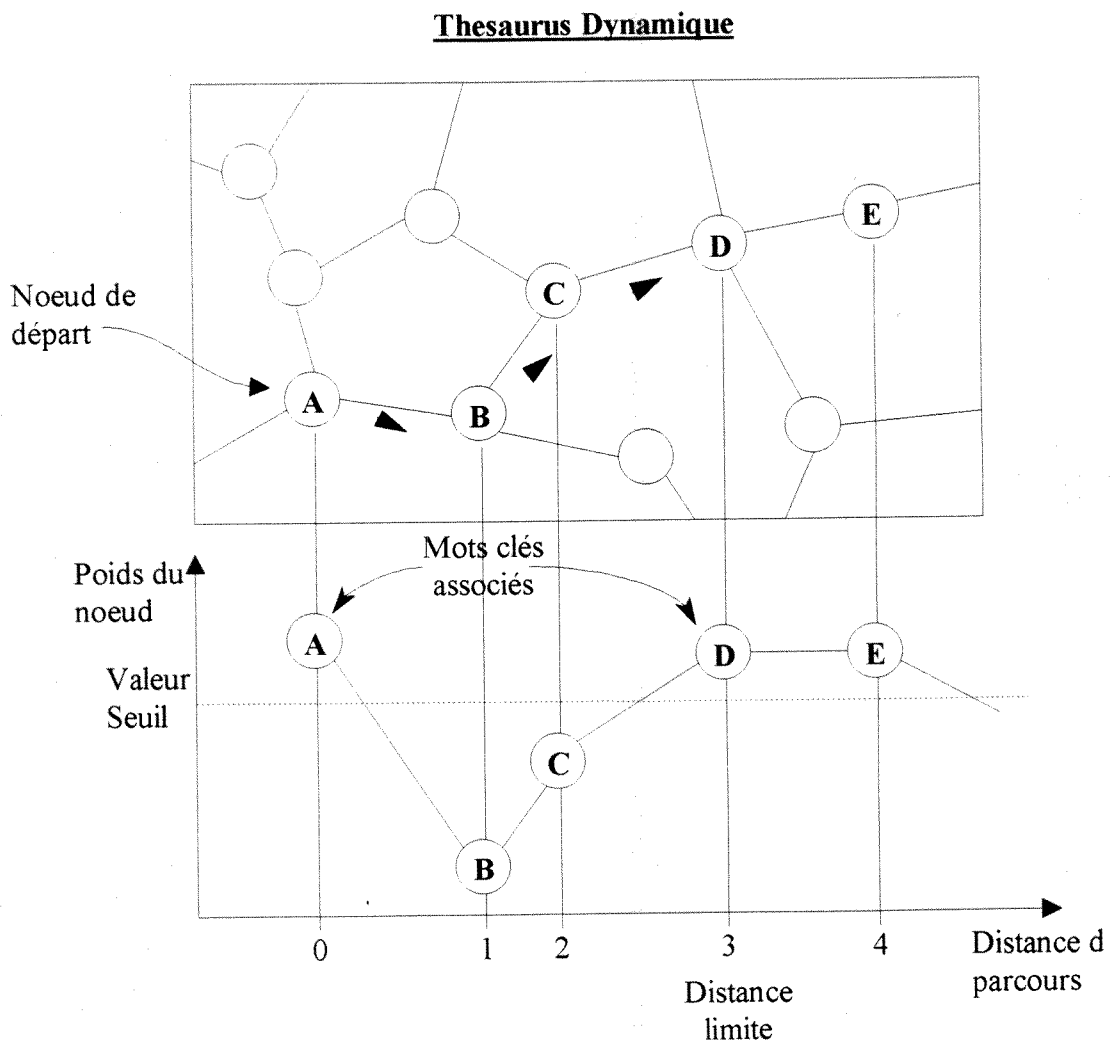


Figure 3

3.2.4. Résultats de l'expérimentation du thesaurus dynamique

Différents tests ont été réalisés pour étudier la pertinence des liens obtenus grâce aux relations de cooccurrence.

Pour la génération des mots clés, le système AIRS a traité des articles de journaux et le thesaurus développé pour la recherche de ces articles a été utilisé comme thesaurus statique. Ce thesaurus statique contenait environ 8000 termes.

Pour la recherche documentaire utilisant les mots clés associés, les expérimentations ont été accomplies sur une base de données de 163 articles de journaux. Trois séries de tests ont été effectuées, faisant varier la taille de l'échantillon de documents utilisé pour extraire l'information sur le terme. Les mots clés furent entrés par un utilisateur, puis les mots associés furent générés par le système AIRS. Une stratégie de recherche booléenne de type "OU" fut adoptée. Puis un taux de rappel et un taux de précision pour la recherche documentaire ont été calculés. Trois types de structures de thesaurus furent utilisés pour trouver la plus appropriée des trois :

- ♦ Thesaurus 1 : noeuds non pondérés et relations de cooccurrences non utilisés (thesaurus statique)
- ♦ Thesaurus 2 : noeuds non pondérés et relations de cooccurrences utilisés
- ♦ Thesaurus 3 : noeuds pondérés et relations de cooccurrences utilisés (thesaurus dynamique).

Les tests où les relations de cooccurrences ont été prises en compte, montrent que le taux de rappel et le taux de précision augmentent considérablement. Cependant, la distance de parcours ne doit pas être trop élevée sinon le taux de précision diminue considérablement. Le thesaurus dynamique (thesaurus 3) s'est montré le plus pertinent pour la recherche documentaire.

Le système implanté génère des liens de cooccurrence quand deux mots clés apparaissent dans le même document pertinent. La génération de liens pourrait aussi refléter le rôle grammatical des mots clés dans une phrase, tel qu'une relation objet-sujet, ou mettre en évidence d'autres types de liens tel que les liens cause-résultat.

4. Ergonomie de la représentation des structures conceptuelles

La représentation des structures conceptuelles pose différentes interrogations quant aux formats dans lesquels elles doivent être affichées :

Comment les noeuds et les arcs doivent-ils être représenté?

- ♦ Le libellé des noeuds n'étant pas de longueur ou de style identique la représentation textuelle doit faire l'objet d'une étude préalable.
- ♦ La forme graphique des noeuds, représentant les concepts, doit permettre la différenciation de ces derniers afin de les classer ou de les hiérarchiser.
- ♦ La forme graphique des arcs doit permettre la représentation des différentes relations pouvant exister entre les concepts.
- ♦ Les noeuds doivent être représentés de manière à permettre la sélection de l'un d'entre eux.

Comment les noeuds et les arcs doivent-ils être placés sur le graphique?

- ♦ La distance entre les boites représentant les noeuds et la disposition des arcs sur le graphique doivent permettre la représentation d'une structure compréhensible par l'utilisateur.
- ♦ Le placement des noeuds ou des arcs peut définir des axes significatifs de la représentation graphique.
- ♦ La géométrie du graphique, dont on a extrait différents modèles, permet de savoir si un ou plusieurs concepts sont pris en compte dans un document particulier.

Les paragraphes suivants s'attacheront à répondre aux questions précédemment posées. De plus, la représentation graphique des thesaurus sera abordée car elle permet de mettre en application ces différentes questions.

4.1. Représentation des noeuds

Le concept représenté par le noeud doit être facilement reconnaissable, c'est à dire que la boîte doit être :

- ♦ assez large pour contenir au moins le ou les mots d'un terme simple,
- ♦ différentiable des autres pour permettre une classification.

4.1.1. Représentation du texte à l'intérieur d'un noeud

En donnant à une boîte une taille limite, il se peut que le texte n'apparaisse pas en entier dans la boîte, en particulier si le texte est une longue expression ou une phrase comme c'est le cas, par exemple, pour les notes d'usage. Que doit-on alors afficher à l'intérieur?

La solution la plus simple est de n'afficher que le début du texte (figure 4).

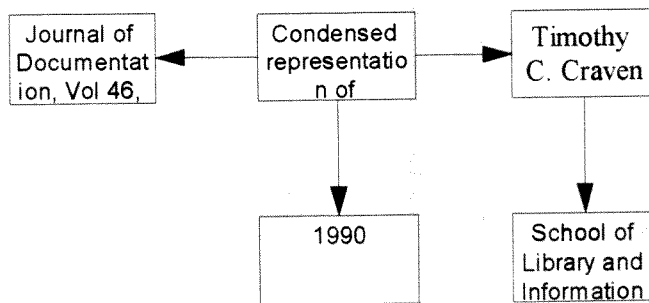


Figure 4

On observe cependant, que le début du texte ne permet pas toujours de différencier les différents concepts ou ne suffit pas à leur compréhension.

Une autre solution est d'abrégé certains termes par des règles simples et automatiques pour permettre ainsi l'affichage de plus d'informations dans l'espace réduit de la boîte. Craven [Craven 1990a] propose trois méthodes pour la représentation de phrases ou de texte (en langue anglaise) dans un graphique:

- ♦ L'écriture rapide des mots dont la longueur dépasse cinq lettres donne une compression d'environ 80% du texte continu original avec peu d'ambiguïté. Cette technique (figure 5) couplée avec deux autres techniques de prise de notes (omission de certains mots, tel que les articles, et la réduction des mots en majuscule à leur initiale) a été implémentée dans le système TEXNET élaboré par Craven.

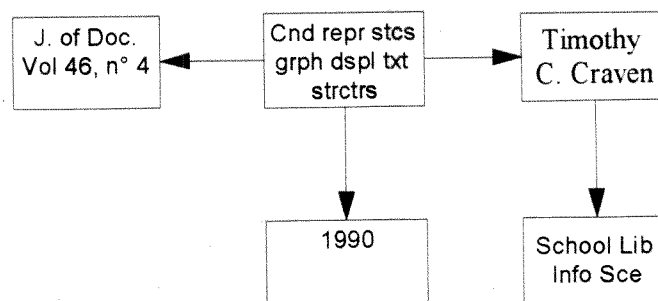


Figure 5

- ♦ Une autre méthode de prise de notes plus complexe (omission de phrases, introduction de phrases adverbiales et de phrases verbe tel que 'has been'; et plus ambitieusement, mise sous forme nominale, passive et de participe) peut être envisagée, mais son temps de traitement serait inacceptable lors de l'application courante.
- ♦ La troisième méthode est l'application de techniques de troncature de texte, qui reste pour l'instant assez approximative.

4.1.2. Ergonomie des boites

La spécificité du noeud peut être déterminée par la forme de la boite. Ainsi, dans le package THSRS [Craven T. 1990b], la bordure des boites est différente si l'on a un terme préféré ou non, et les termes sont d'autant plus importants que les boites sont grandes. Cette différenciation est justifiée car souvent le contenu du noeud le plus important est nécessaire pour la compréhension d'autres noeuds; par exemple, quand le contenu de la classe générique doit faire comprendre les termes représentant les classes spécifiques.

Sur la figure 6, on peut observer que les termes "graphique" et "système" apparaissent dans des boites plus grandes que les autres termes et dans une police de taille et de style différents. Cela permet de désigner la différence de classes (générique/spécifique) qu'il existe entre les différents noeuds.

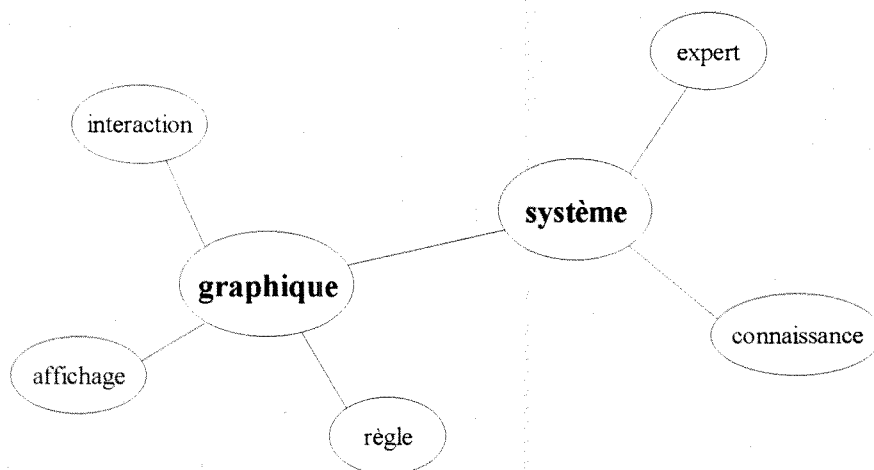


Figure 6

L'utilisateur doit être capable de repérer n'importe lequel des noeuds de l'affichage; par exemple, pour l'effacer, modifier son contenu, le lier à un autre noeud, le délier, ou sélectionner le terme pour une recherche (opérations effectuées lors de la gestion ou de la mise à jour de thesaurus sous forme graphique). Ceci implique que chaque noeud ait un bref identifiant d'affichage, ou simplement qu'il existe une zone claire dans l'écran pour cliquer sur chaque noeud affiché. Les noeuds emboîtés (voir 4.2.1. Subordination) peuvent poser dans ce cas un problème.

4.2. Placement des noeuds

Les noeuds doivent être placés de telle façon qu'il soit facile de voir globalement quels concepts sont associés, et de quelle manière ils sont liés ensemble.

Ces notions peuvent être réalisées, du moins en partie, en plaçant les noeuds liés proches les uns des autres. Les noeuds, et leurs arcs associés, doivent néanmoins être placés de telle sorte qu'ils n'interfèrent pas avec les liens d'autres noeuds, et chaque noeud doit être suffisamment reconnaissable.

Le placement des noeuds diffère suivant l'aspect que la représentation graphique veut mettre en évidence.

4.2.1. La subordination

Afin de montrer la subordination existante sur un noeud, ce noeud est placé à l'intérieur de boîtes assignés à d'autres noeuds (noeuds subordonnés au présent noeud). Cet arrangement est connu sous la notation des ensembles emboîtés et est employé par le système hypertexte IGD [Feiner 1988]. Cette "vue aérienne" est quelquefois plus flexible que de rassembler les noeuds liés de façon non hiérarchique ou polyhiérarchique.

En plus de la subordination, il existe d'autres raisons à l'emboîtement des boîtes des noeuds; par exemple, pour proposer à l'utilisateur une vue détaillée en faisant apparaître les noeuds les moins importants.

4.2.2. Les directions

L'association d'une direction spatiale logique avec les directions des liens rend les relations entre concepts plus claires. Par exemple, des stratégies de recherche sont couramment développées dans lesquelles les termes sont OU/verticalement et ET/horizontalement, ou dans l'arrangement contraire [Chang 1990].

Certains types de liens n'ont pas de direction intrinsèque particulière, tel que les liens basés sur une cooccurrence à l'intérieur d'une même phrase, alors que dans le cas contraire, la direction du lien peut être un trait dominant d'un schéma conceptuel. Ainsi, un ensemble de liens dirigés peut définir une hiérarchie qui, en grande partie, détermine la structure et l'utilisation d'un schéma conceptuel.

A un niveau paradigmatique, on peut citer la direction depuis le terme générique jusqu'aux termes spécifiques, souvent défini en terme de relation genres-espèces; au niveau syntagmatique la direction du concept qualifié vers le qualifiant dans l'indexation en chaîne; ou encore, dans le système TOSAR [Fugman 1974], la relation entre premier et dernier dans le temps.

Cependant, les exigences des directions peuvent affecter sérieusement la position relative des noeuds. Elles peuvent rendre la représentation de certains liens difficile ou les résultats incompréhensibles.

4.2.3. Un ordre préférentiel

Les noeuds peuvent être placés pour refléter un ordre particulier qui est préférable pour certaines raisons. Un exemple pourrait être l'ordre systématique dans l'agencement d'une classification. Un autre est l'ordre alphabétique, tel qu'il est décrit dans le système de Cahn [Cahn 1980]. Dans ce cas là, le but n'est pas autant de clarifier les relations entre les concepts que de faire des noeuds plus facile à trouver dans un graphique dans lequel plusieurs concepts sont représentés.

4.2.4. Autres modèles

Différents modèles ont été étudiés pour assurer une place adéquate à chaque noeud lors de l'affichage.

L'application d'un quadrillage de formes uniformes simples (triangles, rectangles, ou hexagones) a été envisagé par plusieurs chercheurs. Chaque noeud occupe alors une case où tout autre noeud est exclu.



Une case rectangulaire, comme dans la méthode de Watanabe [Watanabe 1989], est probablement la plus simple, par rapport à une hexagonale, ou à un arrangement en nid d'abeille comme le suggère Duncan [Duncan 1990].

Un autre modèle possible, est un ensemble de ronds concentriques (voir figure 10). Cet arrangement semble bien fonctionner quand on est en présence d'un ou deux noeuds de haut degré, c'est à dire d'un noeud avec beaucoup d'arcs associés.

4.3. Styles des arcs

Différents styles d'arcs dessinés indiquent les différents types de liens qui existent entre les noeuds. Il n'existe aucune convention générale, cependant, les liens directionnels sont indiqués par des flèches.

Une flèche pointe généralement vers le bas dans une hiérarchie, comme dans le système de Chaumier et Fourteau pour la représentation des thesaurus à schémas fléchés [Chaumier and Fourteau 1979] où la représentation des diverses relations sémantiques est la suivante :

- ♦ relations hiérarchiques :
 - ⇒ relations génériques : trait doublement fléché : 
 - ⇒ relations partitives : trait simplement fléché : 
- ♦ relations d'associations :
 - ⇒ relations d'associations intra-schémas : trait non fléché à l'intérieur du schéma.
 - ⇒ relations d'associations inter-schémas : trait non fléché vers l'extérieur du cadre.
- ♦ relations d'équivalence : caractère de types différents entre descripteurs et termes équivalents.

L'utilisation d'une grille permet de localiser et de coder les descripteurs par leurs coordonnées dans le plan.

Dans la figure suivante on peut observer les différents types de relations et leur adéquation à l'intérieur d'un schéma¹⁸ :

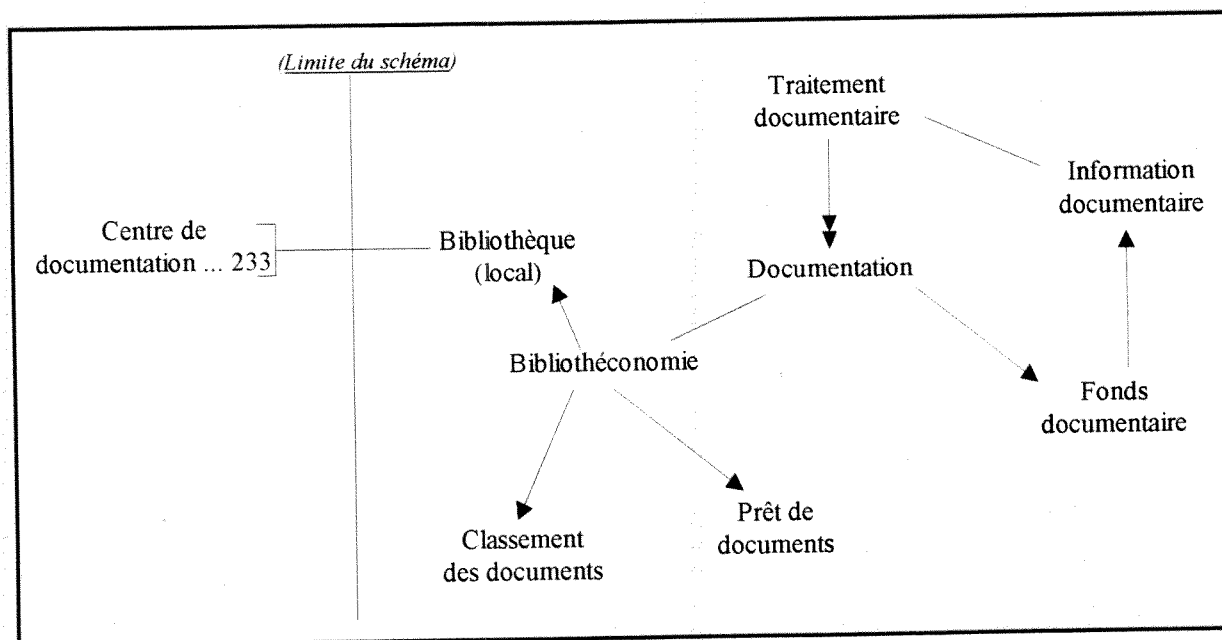


Figure 7

¹⁸ Schéma extrait du thesaurus à schéma fléchés d' Electricité de France (4ème édition) - Champs sémantique : "Traitement documentaire".

Les arcs peuvent représenter le degré de proximité des termes. Par exemple, Doyle [Doyle 1962] utilisait des flèches pour montrer l'ordre des termes cooccurrent dans les expressions. Il utilisait pour cela, des arcs de nature différente (ligne pleine, ligne pointillée, double ligne, ligne fléchée ou non fléchée et d'épaisseur variable) pour indiquer les degrés variables de cooccurrence. De plus, un numéro était affecté à chaque arc, reflétant la valeur du coefficient de corrélation (figure 8).

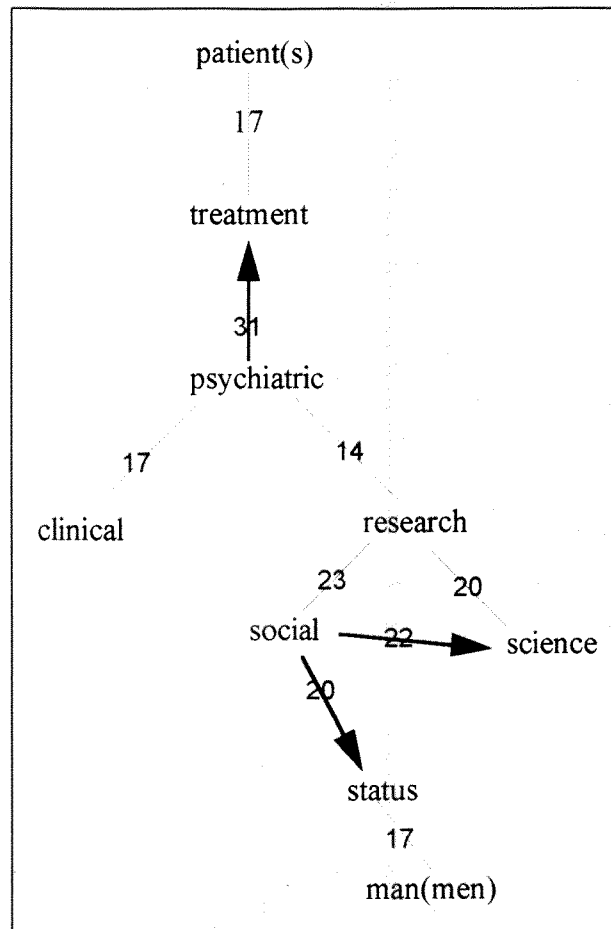


Figure 8

Un problème peu exploré pour l'instant a été soulevé par Craven : "Jusqu'à où la représentation des arcs peut augmenter la perception d'une troisième dimension dans le cas où un affichage doit représenter une vue en perspective?"

4.4. Représentation graphique de thesaurus

La représentation graphique est le type de représentation qui a provoqué le plus de recherche de la part des concepteurs de thesaurus. Des nombreuses représentations graphiques proposées, deux essentiellement sont encore couramment utilisées :

- ♦ les schémas graphiques
- ♦ les tableaux graphiques

La représentation graphique des descripteurs et de leurs relations dans un thesaurus offre de considérables avantages, comme l'a décrit Chaumier [Chaumier 1988] :

- ♦ Les représentations graphiques donnent une vue d'ensemble immédiate de l'environnement, d'un concept, voire du domaine. Les représentations graphiques sont appréhendées beaucoup plus rapidement qu'une séquence de mots, ou même qu'une table d'information.
- ♦ La représentation graphique permet la dispersion des termes associés, ce qui produit différents bénéfices :
 - * Les représentations graphiques en deux dimensions permettent de mettre en valeur la diversité des relations autour du concept central (diversité qui est impossible à retranscrire par une simple énumération linéaire, où un terme peut seulement être le voisin de son prédécesseur et de son successeur dans une relation simple : alphabétique, numérique ou hiérarchique).
 - * Le volume du thesaurus est réduit, particulièrement avec les types de représentations graphiques compactes. En effet, dans une structure de cercles concentriques, les noeuds peuvent se situer tout autour du noeud central, permettant ainsi plus de deux voisins.
- ♦ Les représentations graphiques augmentent la qualité de l'indexation en terme d'exhaustivité et d'uniformité.
- ♦ Les représentations graphiques augmentent la qualité de la recherche documentaire, c'est à dire une augmentation du rappel et de la précision.
- ♦ Les représentations graphiques sont particulièrement abordables pour les non spécialistes qui ne sont pas familiarisés avec le vocabulaire ou qui ne connaissent pas les conventions utilisées par les indexeurs.

Néanmoins avec l'accroissement du nombre de descripteurs il est parfois difficile de tous les représenter avec leurs relations, dans une forme claire et intelligible. Il en est de même pour les thesaurus multilingues.

4.4.1. Les schémas graphiques

4.4.1.1. Les thesaurus à schémas fléchés (arrowgraphs)

C'est le type de représentation graphique le plus ancien et le plus répandu. L'enveloppe des concepts d'un même domaine, souvent appelée champ sémantique, est définie de façon pragmatique; le nombre de descripteurs doit être tel que le schéma soit lisible et clair : 30 à 50 termes constituent une bonne norme.

Le repérage d'un descripteur dans un schéma se fait grâce à une réticulation du schéma en cent cases numérotées de 00 à 99.

Un descripteur est identifié par un code constitué du numéro du schéma suivi du numéro de la case où il se trouve. Dans la figure suivante¹⁹ (figure 9), le descripteur 236.38 correspond au champ 236 ("langage d'indexation") et à la case 38 de ce champ 236, soit au descripteur "classification hiérarchique". Ce système de numérotation est particulièrement efficace pour des thesaurus multilingues, car il sert de notation pivot entre les différentes langues, le descripteur ayant la même localisation dans les différentes versions linguistiques.

La figure suivante, présente une représentation graphique d'un thesaurus sous forme de schéma fléché :

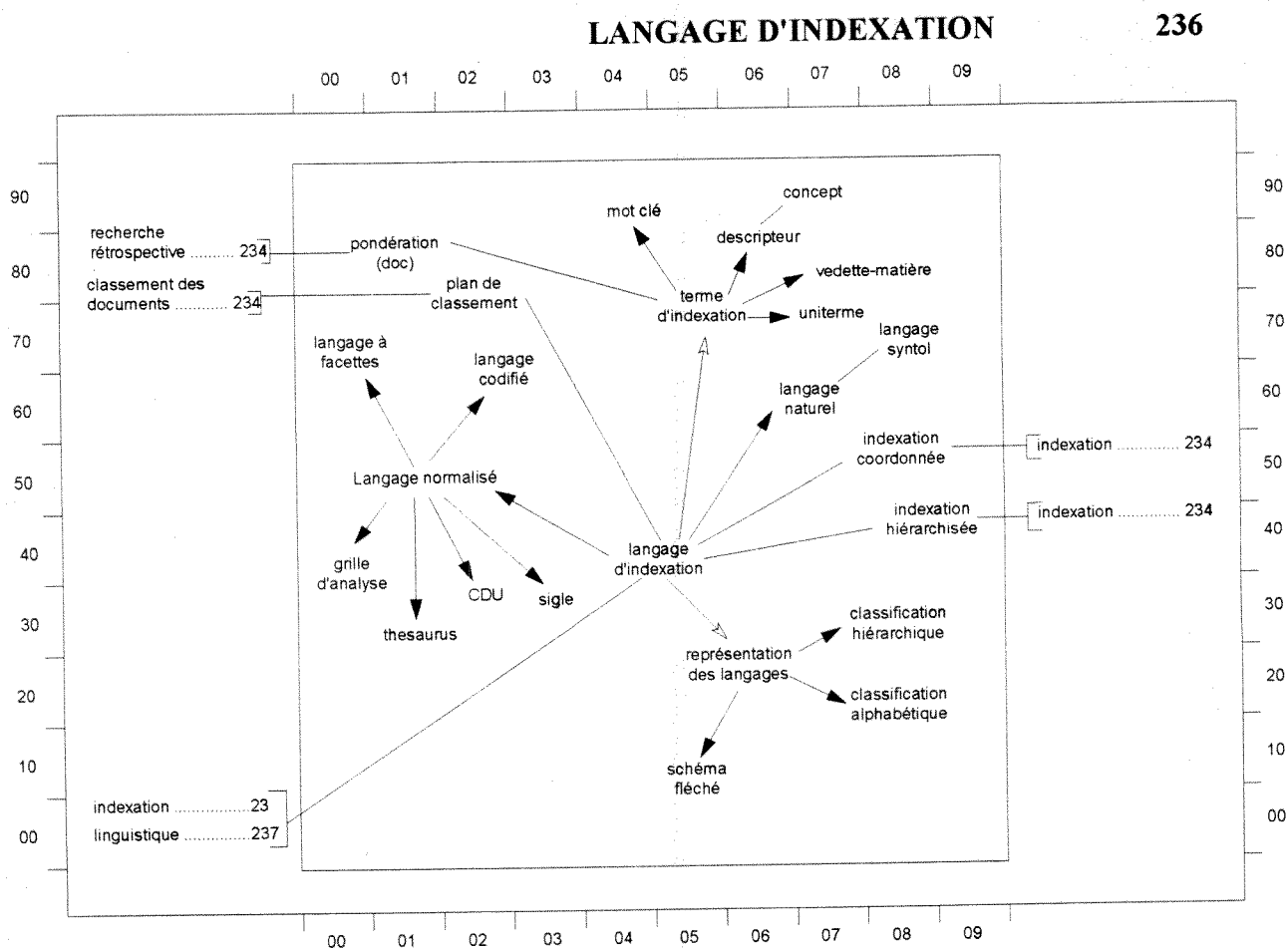


Figure 9

L'agencement des descripteurs à l'intérieur d'un schéma correspond à une structure rayonnante à partir du centre, constitué de plusieurs arborescences. Il s'agit d'une hiérarchie plus ou moins imbriquée, correspondant aux divers critères de subdivision possibles. Schématiquement, on peut dire que le descripteur générique du schéma (tête d'arborescence), son titre, figure au centre, c'est à dire qu'il est placé dans l'une ou l'autre des quatre cases

¹⁹ Schéma extrait du thesaurus à schéma fléchés d' Electricité de France (4ème édition) - Champs sémantique : "Langage d'indexation".

centrales possibles du schéma (cases 44-45-54-55). A partir du terme générique, les termes spécifiques sont placés dans une case, de plus en plus éloignée vers la périphérie, en fonction de leur degré de spécificité.

Le type de relations sémantiques établies entre les descripteurs est précisé par une représentation conventionnelle des liaisons, comme nous l'avons vu précédemment [Chaumier and Fourteau 1979]. Les relations entre termes de différents schémas sont représentées par des liaisons avec les descripteurs placés hors du cadre du champ. Les descripteurs ainsi placés sont précédés de leur localisation dans leur propre champ. De ce fait, un descripteur donné n'apparaît que dans un seul schéma, mais il est rappelé, autant de fois que nécessaire, dans la marge d'autres schémas avec lesquels il est en relation.

Les synonymes - au sens documentaire - figurent en général dans une liste annexée au schéma, mais s'ils ne sont pas trop nombreux, on peut les porter sur les schémas eux-mêmes en recourant à un graphisme différent.

Le schéma fléché permet d'avoir simultanément sous les yeux l'ensemble du vocabulaire disponible dans un domaine pour l'indexation d'un document ou d'une question. La structure du thesaurus est ainsi mise en relief, en particulier par des noeuds du graphe relationnel.

L'un des inconvénients majeurs de ce type de représentation reste la difficulté de mise à jour des schémas. Chaque schéma est généralement accompagné de la liste alphabétique des descripteurs qu'il renferme. Cette liste donne le numéro de la case du descripteur, le descripteur et les termes équivalents, précédés de la relation EP, que ce descripteur recouvre.

4.4.1.2. Les schémas circulaires

Les schémas circulaires (figure 10)²⁰ sont une intéressante représentation graphique. Le terme générique figure au centre du schéma, puis, sur les couches concentriques du schéma sont placés les termes spécifiques. De plus, les secteurs du schéma délimitent des sous-ensembles sémantiques dans le champ sémantique.

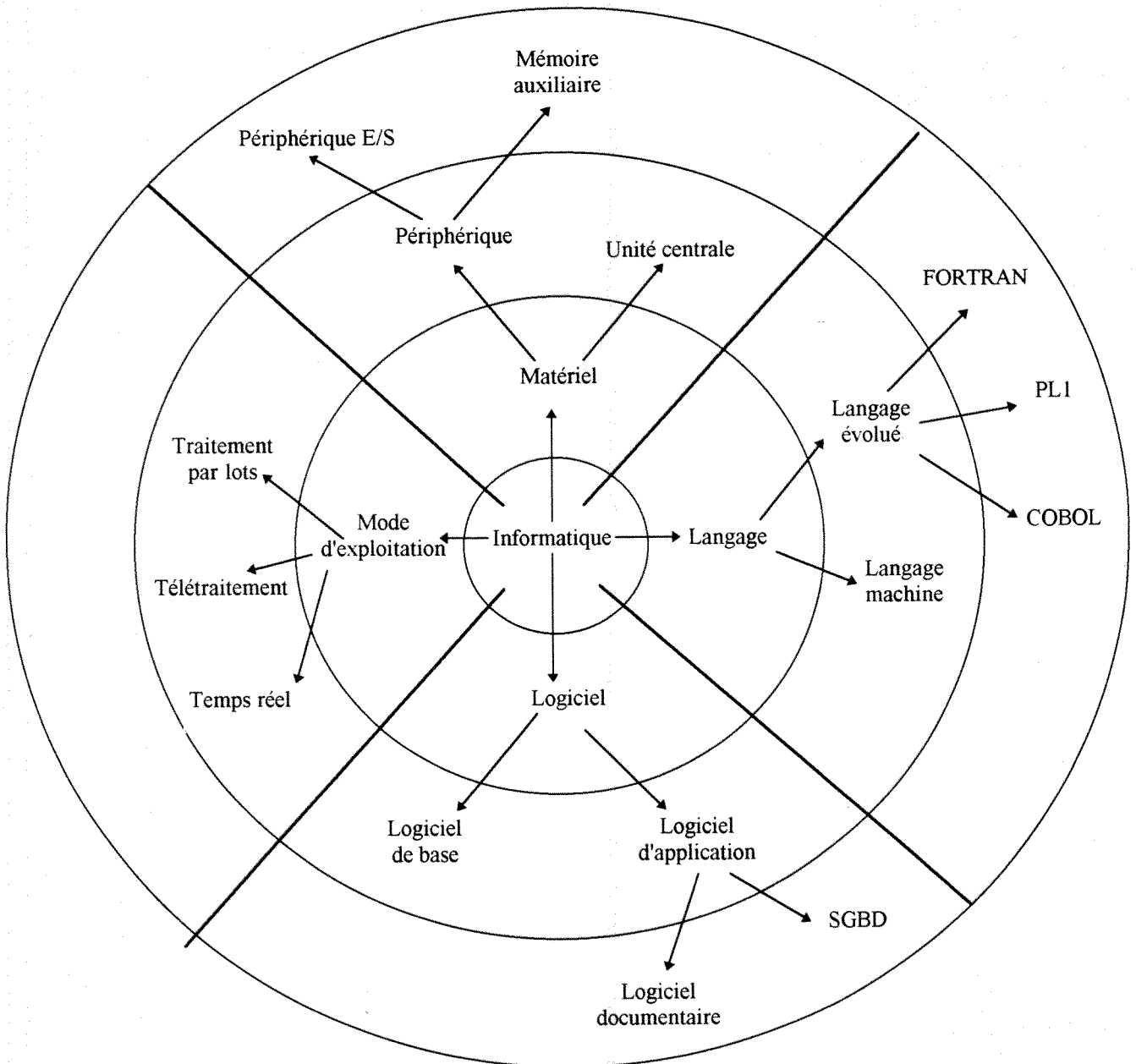


Figure 10

²⁰ Schéma extrait de : Le traitement linguistique de l'information - Chaumier J. 1988.

4.4.1.3. Les représentations graphiques arborescentes

L'une des représentations graphiques les plus simples reste la représentation directe des arborescences (figure 11). Mais si ce système a l'avantage d'être simple et, par conséquent, très clair à l'emploi, son inconvénient majeur réside dans l'encombrement des hiérarchies.

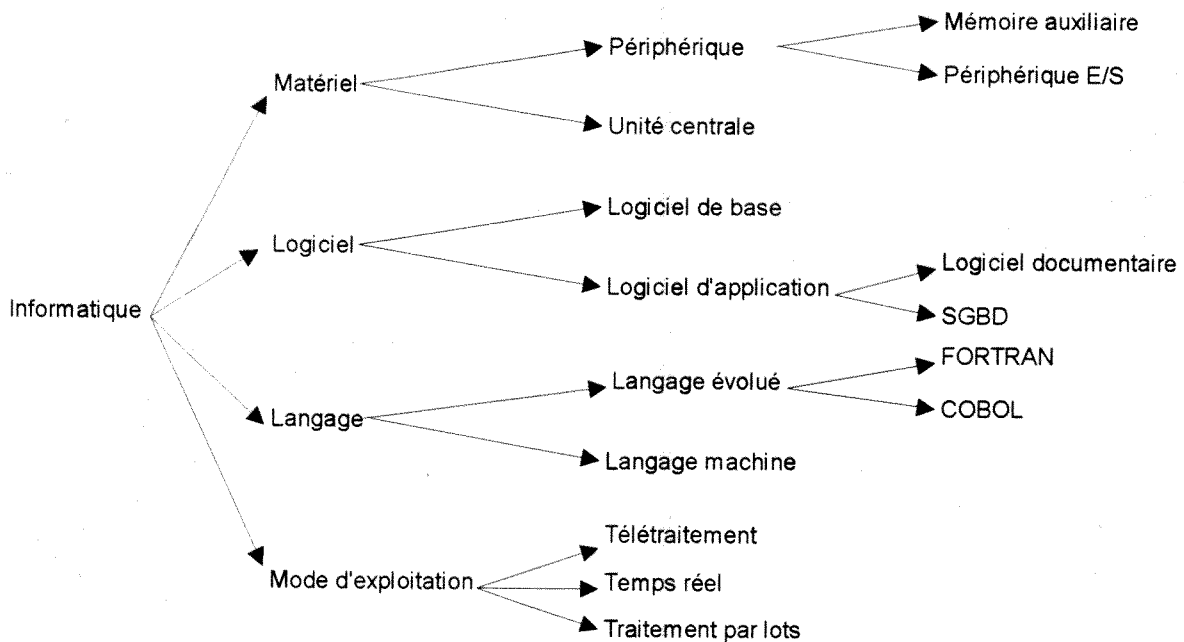


Figure 11

La représentation graphique ne se développant que dans un seul sens (à l'encontre des schémas qui utilisent le plan dans toutes ses orientations), si le nombre de termes génériques est important, il devient difficile d'utiliser ce mode de représentation. De même, s'il y a disproportion du nombre de termes spécifiques d'un descripteur à l'autre, la lecture devient également difficile.

4.4.2. Les tableaux graphiques (terminographs)

Ce mode de représentation a été plus particulièrement employé pour des thésaurus au volume important. En effet, l'une des caractéristiques réside dans la densité de chaque tableau.

Comme dans les représentations précédentes, le vocabulaire est découpé en champs sémantiques. Chaque champ fait l'objet d'un tableau sur lequel on fait figurer l'ensemble du vocabulaire, c'est à dire descripteurs et termes équivalents. Les termes équivalents sont repérés par le signe "-" qui les précède (figure 12).

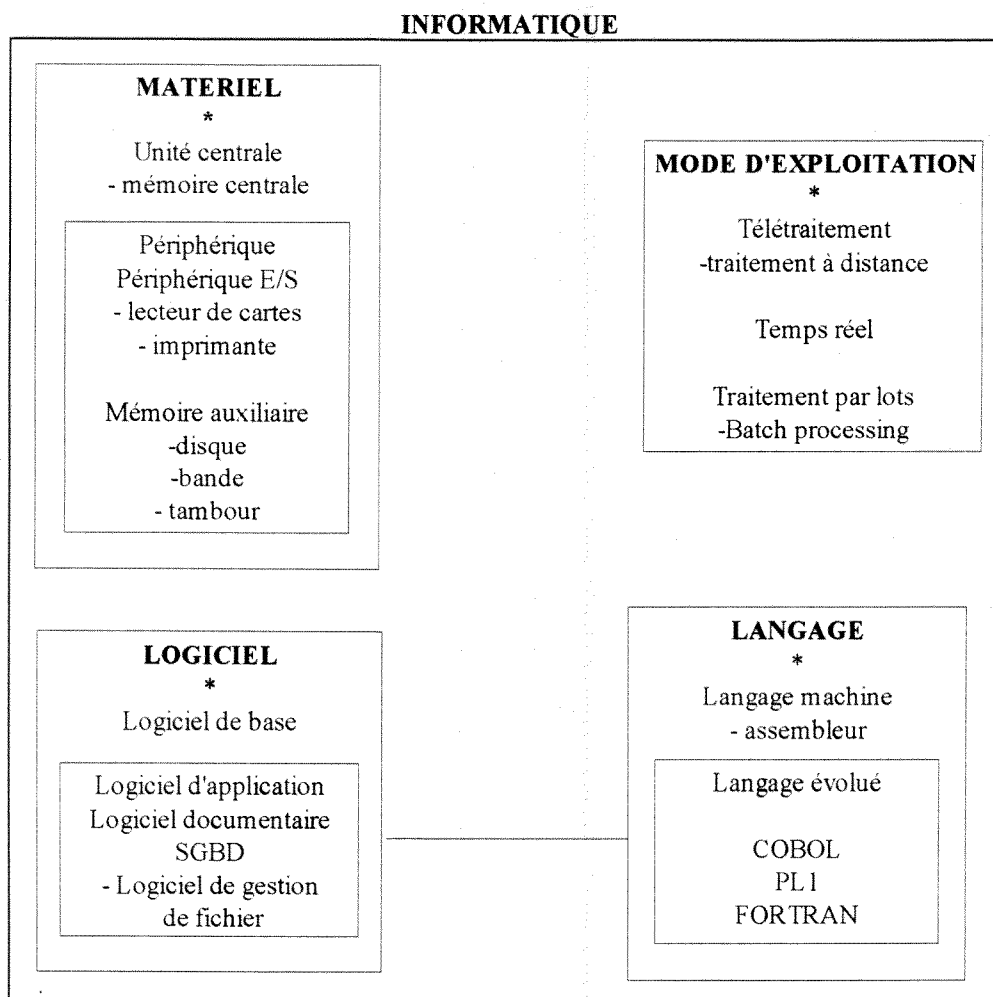


Figure 12

Les sous-ensembles sémantiques sont délimités sur le tableau par des carrés ou rectangles renfermant le vocabulaire du sous-ensemble. L'imbrication des sous-ensembles permet de représenter la structure hiérarchique du champ sémantique.

Cependant la lecture de ce type de thésaurus nécessite un certain apprentissage.

4.4.3. Les macro-schémas

Certains systèmes ont cherché à représenter le thesaurus en un seul schéma permettant, ainsi, d'appréhender d'un seul coup d'oeil la structure générale du langage. Ces macro-schémas étaient généralement destinés à présenter le "macro-vocabulaire", c'est à dire l'ensemble des descripteurs génériques (top terms). Cette représentation est réalisée dans le but de différencier le vocabulaire d'indexation, plus particulièrement destiné à la diffusion sélective, c'est à dire un vocabulaire assez générique, et le vocabulaire plus spécifique, plus particulièrement orienté vers la recherche rétrospective.

5. Caractéristiques graphiques particulières

Certaines des représentations graphiques que nous avons abordé dans les paragraphes précédents, proposent des caractéristiques spéciales qui permettent de modifier leur affichage graphique ou de créer de nouveaux affichages à partir des données qu'elles contiennent. Ces options d'affichage, en représentant différemment les concepts et les relations, offrent à l'utilisateur de :

- ♦ clarifier la représentation graphique,
- ♦ choisir un critère de sélection à partir duquel sera redéfini la représentation graphique,
- ♦ définir le terme central de la représentation,
- ♦ choisir une vue différente de celle représentée.

Nous allons étudier certaines de ces caractéristiques pour comprendre l'intérêt qu'elles offrent dans l'optique de les utiliser dans les cartes navigationnelles.

5.1. Affichages sélectifs

L'utilisation d'affichages sélectifs est une solution partielle aux problèmes créés par de grandes structures ou des structures complexes. Par exemple, le système CALIBAN [Frei and Jauslin 1983] est limité à une simple structure arborescente à trois niveaux pour l'affichage des relations d'un thesaurus.

La sélection peut donc s'envisager pour des raisons de présentation, c'est à dire pour permettre une reconnaissance de la structure en délimitant les informations présentées, et rendre ainsi la représentation graphique plus compréhensible. Elle peut aussi s'envisager pour permettre la mise en valeur d'un type de concept ou de relation, afin d'obtenir une information différente de celles présentées dans une vue non sélective.

Les affichages dont la sélection s'effectue par le nombre de noeuds inclus, sont produits en rayonnant à partir d'un noeud central simple. Le choix des noeuds environnants est déterminé suivant divers critères :

- ♦ par un nombre maximum de noeuds qui détermine l'étendue d'une sélection.
- ♦ en ne retenant que les noeuds ayant un lien direct avec le noeud central.
- ♦ par le type de lien (les noeuds ayant le même type de lien que le noeud central peuvent être regroupés dans un style compact, comme dans le système décrit par Cahn [Cahn 1980]).
- ♦ par le type de coordonnées (dans une situation de scaling multidimensionnel, une solution consiste à choisir les noeuds ayant des coordonnées semblables autres que les deux représentées à l'affichage).

Une sélection peut être effectuée sur le type des liens ou des concepts que l'on veut représenter :

- ♦ Pour une structure basée sur des mesures d'association, telle que les cooccurrences, une approche est d'afficher seulement les liens les plus forts pour chaque noeud, comme le suggère Doyle [Doyle 1962]. En effet, dans le système envisagé par Doyle, chaque arc se voyait affecter la valeur de son coefficient de corrélation (voir figure 6). On pouvait, par la suite, choisir un seuil minimal d'affichage des noeuds à partir duquel la sélection était effectuée.

- ♦ Dans les structures contenant des liens hiérarchiques, les liens non hiérarchiques peuvent être exclus. Lors d'une représentation graphique d'un thesaurus, par exemple, les relations d'associations peuvent être ignorées afin de représenter uniquement les relations génériques et partitives.

La sélection est aussi très utile pour la présentation des résultats de recherche. On peut alors sélectionner suivants différents critères (date de parution, langue, etc.) pour permettre une classification ordonnée des documents.

Afin de réaliser ces différentes sélections l'utilisateur doit pouvoir choisir le(s) concept(s) ou la(es) relation(s) qu'il désire voir s'afficher. Par exemple, le système KEATS [Eisenstatd, Domingue, Rajan and Motta 1990] permet à l'utilisateur de sélectionner le type de liens à afficher.

L'affichage sélectif est un outil très précieux mais qui nécessite, de la part de son utilisateur, la connaissance suffisante des effets espérés.

5.2. Scaling : Le concept de "fisheye view"

L'utilisateur doit avoir la possibilité d'ajuster la taille des boîtes pour découvrir plus ou moins la structure complète de la représentation graphique ou le texte d'un noeud individuel. Un ajustement global doit permettre à l'utilisateur de changer l'échelle de l'affichage. En agrandissant, une vue d'ensemble peut être accomplie. En zoomant, l'utilisateur peut avoir accès à une zone particulière, visualisant moins de noeuds, mais avec plus de détails.

L'ajustement local peut être associé au concept de "fisheye view" [Furnas 1986]; en général les "fisheye views" sont créées en faisant varier sélectivement la taille, la position, et le style graphique des noeuds ou des arcs.

Une "fisheye view" affiche les objets proches de l'élément central (élément considéré comme le plus important par l'utilisateur), aussi appelé focus, avec plus de détails. L'affichage des différents éléments dépend à la fois de la distance de l'élément par rapport au focus et de l'importance (poids) *à priori* attribuée à l'élément. Ces deux caractéristiques définissent la taille (ou tout autre facteur) de tous les éléments à rattacher au focus.

Pour afficher une "fisheye view", l'utilisateur sélectionne un terme, définissant ainsi le focus. Puis pour chaque noeud, on définit son degré d'intérêt, composé de :

- ♦ son degré, c'est à dire du nombre de liens qui lui sont affectés, et
- ♦ sa distance de vue, c'est à dire la longueur de son chemin par rapport au focus.

Seuls les noeuds ayant une valeur supérieure à une certaine valeur seuil sont alors retenus pour l'affichage. La valeur seuil fixe la taille de la fenêtre d'affichage, c'est à dire le nombre approprié de noeuds qui doivent être affichés. Après avoir déterminé les noeuds qui doivent être affichés, un algorithme de traçage des noeuds pour les petits réseaux est utilisé pour positionner les noeuds.

Dans de la cas de la figure 6, on définit le terme "système" comme focus. On obtient alors la figure suivante :

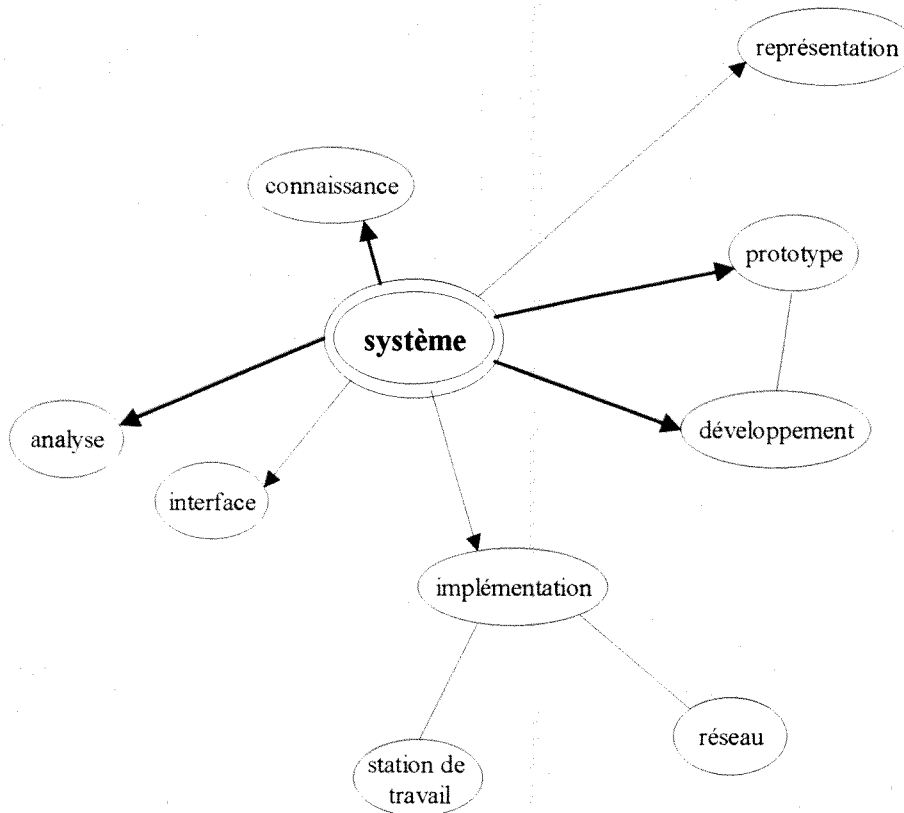


Figure 12

Le terme "système", choisi comme focus, apparaît dans une boîte différente des autres termes. Les nouveaux termes affichés (par rapport au schéma de la figure 6), sont le résultat de la prise en compte du terme "système" comme focus, redéfinissant ainsi leur degré d'intérêt. La longueur des arcs reflète le poids de chaque noeud par rapport au focus.

La forme visuelle des noeuds et des bords transmet des informations additionnelles sur les relations entre les noeuds de la "fisheye view". Le degré d'un noeud est représenté par sa taille. Les noeuds ayant de grandes valeurs sont plus grands, et ainsi plus proéminents dans l'affichage. Le focus doit être le plus gros noeud de la vue. Les noeuds à haut degré sont les noeuds que l'utilisateur peut explorer pour trouver les parties les plus denses du réseau.

La forme des bords des fenêtres transmet aussi de l'information sur la structure du réseau. Les bords sont plus larges pour le focus et de plus en plus fins à mesure que les noeuds connectés s'éloignent du focus (non représenté sur la figure).

Une "fisheye view" peut être construite à l'intérieur d'un affichage de type particulier; par exemple, un affichage hiérarchique peut attribuer plus d'espace à des noeuds hauts hiérarchiquement, et moins d'espace est donner au noeuds bas, qui sont souvent moins importants.

Le concept de "fisheye view" peut amener un autre type d'information. En effet, si ce concept est placé à l'intérieur d'un modèle navigationnel, l'enregistrement des différents focus sélectionnés par l'utilisateur tout au long de sa recherche permet de définir un chemin de recherche. Ce chemin de recherche, une fois enregistré et traité de manière efficace (suppression des branches de distance inférieure à une valeur seuil, etc.), peut être proposé à d'autres utilisateurs. En effet, si un autre utilisateur débute sa recherche sur un noeud faisant partie d'un chemin déjà existant, ce chemin peut lui être présenté, ce qui lui permet d'arriver aux résultats plus rapidement si ce chemin correspond à la recherche qu'il désirait effectuer. Le chemin présenté peut aussi élargir son champs de recherche en faisant intervenir des notions auxquelles l'utilisateur n'avait pas pensées.

Par la suite des mesures statistiques pourraient être développées en considérant un même noeud, départ de plusieurs chemins, et les différents noeuds d'arrivée de ces chemins, afin de déterminer la ou les recherches les plus souvent accomplies.

Le concept de "fisheye view" nous semble donc être une caractéristique importante des cartes navigationnelles et son étude approfondie permettra sans doute de définir de nouveaux outils pour la recherche documentaire.

5.3. Le procédé Visual Recall de Xerox

Visual Recall incorpore les nouvelles technologies du centre de recherche de Xerox à Palo Alto (PARC) dans l'analyse linguistique et la visualisation de l'information. Le résultat est un système compréhensif qui est à la fois facile à utiliser et sophistiqué.

Les groupes de travail utilisant Visual Recall peuvent créer, partager et utiliser leurs documents d'organisation plus efficacement. Une base de données textuelles donne des résultats de recherche plus complet.

La visualisation unique des différentes possibilités de l'application dans une interface intuitive permet de voir les modèles et les relations d'un résultat de recherche, donnant ainsi un sens à un grand nombre de documents en un coup d'oeil. Cette interface permet donc d'utiliser toutes les capacités :

- ♦ La fenêtre de **recherche** permet de définir et d'exécuter une recherche de documents. Les recherches booléennes sont construites et affichées graphiquement, la recherche pouvant s'effectuer sur plusieurs serveurs.
- ♦ La fenêtre de **travail** est un emplacement permanent pour stocker les documents. On peut utiliser les documents individuellement ou les regrouper pour une utilisation ultérieure.
- ♦ La fenêtre **d'affichage** montre les documents obtenus à la suite de la recherche. On peut alors naviguer dans les documents pour en visualiser le contenu.

Une vue arborescente regroupe les documents par propriétés et présente le résultat dans une hiérarchie qui facilite la représentation des relations entre les documents. L'utilisateur peut sélectionner les propriétés et les arrangements qui permettront de classer les documents. Il peut aussi sélectionner un ou plusieurs documents pour avoir une vue rapprochée.

Une vue quadrillée (*grid view*) dispose les documents le long d'une propriété linéaire, comme la date ou la version. Le quadrillage rend la détection des ensembles de documents et des anomalies plus facile, permettant ainsi la sélection des documents dont l'utilisateur a besoin. L'utilisateur peut sélectionner les attributs et les propriétés d'affichage linéaire.

Une visualisation en trois dimensions est disponible, permettant à l'utilisateur de naviguer dans les documents directement à partir de la fenêtre d'affichage. Cette visualisation est possible grâce à une approche intéressante de représentation d'un arbre complexe sous forme de cônes. En effet, les noeuds de même niveau sont regroupés sous un cône, c'est à dire que les concepts sont représentés sur la base du cône, donnant ainsi une représentation tridimensionnelle, ce qui permet de réduire la place des branches de l'arbre. Puis par un système de projection sur un plan on observe la partie de l'arbre projetée sans faire apparaître le reste de l'arbre. L'utilisateur peut alors faire tourner les cônes pour projeter une autre partie de l'arbre.

6. Conclusion

L'objectif de ce travail est de poser un ensemble de questions qui doivent servir à la conception de cartes navigationnelles, puis de présenter des idées, des travaux et des outils qui répondent à certaines de ces questions.

En ce qui concerne le contenu servant à la représentation des cartes, nous avons abordé plusieurs éléments provenant de différentes sources d'information. Nous avons pu déduire de ces études différentes choses.

Le titre semble être un concentré du contenu du document, mais certains éléments, comme le manque de rigueur ou l'imagination des auteurs, remettent en cause la pertinence du titre comme élément de description. La prise en compte du titre doit donc se faire de façon indirecte, par exemple en extrayant les mots pertinents de ce dernier.

En revanche, le mot clé, produit de l'indexation, est sans aucun doute l'élément décrivant le mieux une partie du document. La sommation de tous les mots clés doit permettre la description de chaque partie du document et ainsi former une entité descriptive.

D'autres éléments, moins classiques de l'analyse documentaire, apportent une valeur informationnelle sur le contenu du document. En effet, le résumé et la table des matières, de part leurs formes développées, apportent un complément d'information aux éléments précédents. Le résumé en permettant d'identifier rapidement le contenu d'un document; et la table des matières en donnant une structure hiérarchisée du document. Comme pour le titre, la prise en compte de ces deux éléments peut se faire de façon indirecte grâce aux mots clés qu'ils contiennent.

Enfin, les cooccurrences permettent un apport informationnel supplémentaire en mettant en valeur les liens existant entre certains mots.

Les relations quant à elles traduisent, de part les liens qu'elles forment entre les concepts, une structuration de l'information présentée. Elles sont nécessaires à la bonne compréhension des graphiques.

Tous ces éléments peuvent tendre à apparaître dans les cartes navigationnelles, sous une forme structurée qui combinerait leurs effets sur la recherche documentaire.

Hormis les règles ergonomiques nécessaires à la création de représentations graphiques claires et structurées, deux éléments particuliers ont attiré notre attention.

Tout d'abord, le concept de "fisheye view" ouvre de nouvelles perspectives pour les représentations graphiques. En effet, la mise en place d'un tel concept à l'intérieur des cartes navigationnelles, c'est à dire la possibilité de redéfinir un terme du graphique comme le focus (terme central), permet une navigation implicite dans l'ensemble des concepts. De plus, en mémorisant l'ensemble des focus sélectionnés, on obtient le chemin de recherche qu'a parcouru l'utilisateur. On peut alors imaginer représenter les chemins de recherche pour assister d'autres utilisateurs dans leurs recherches. Ces chemins deviennent donc, une information à part entière. Différentes études peuvent être réalisées sur la pertinence de la représentation de ces chemins, en fonction du degré de satisfaction qu'éprouve l'utilisateur. En

effet, la représentation d'une recherche qui n'a pas satisfait l'utilisateur qui l'a effectuée, doit-elle être proposée à d'autres utilisateurs?

Deuxièmement, la visualisation en trois dimensions disponible dans le procédé Visual Recall, permet à l'utilisateur de naviguer dans les documents directement à partir de la fenêtre d'affichage. Cette notion de navigation, inhérente dans les cartes navigationnelles, est donc à prendre en considération.

A partir de tous ces éléments, quelle représentation des cartes navigationnelles pouvons nous envisager?

Tout d'abord, la notion de multi-fenêtrage nous paraît être une option de représentation intéressante. En effet, en présentant toutes les informations liées à une recherche sur un même écran, l'utilisateur dispose d'une vue d'ensemble qui facilite la cohésion et la compréhension des informations qui lui sont proposées. Le découpage des informations en différentes fenêtres peut être imaginé de la façon suivante :

Dans la fenêtre principale, la carte navigationnelle des concepts est représentée, mettant en scène les éléments bibliographiques vus précédemment. L'utilisateur choisit parmi les concepts représentés par la carte, celui ou ceux (recherche booléenne) qui seront le ou les focus de la prochaine vue. Une fois que l'utilisateur a sélectionné tous les concepts qui semblent décrire ses centres d'intérêt, il lance alors la requête.

Un groupe de trois fenêtres sert à la gestion des documents. Dans une première fenêtre (fenêtre de résultat), les résultats de la recherche sont affichés sous la forme d'une liste de documents. Cette liste contient donc les titres des différents documents auxquels a abouti la recherche. Dans la seconde fenêtre (fenêtre de sélection) contient les documents sélectionnés par l'utilisateur; documents extraits de la précédente liste (fenêtre de résultat). Enfin, une troisième fenêtre (fenêtre de présentation) présente en détail un document sélectionné, c'est à dire, présente le titre, l'auteur et un résumé de ce document. L'utilisateur peut donc gérer, grâce à ces trois fenêtres, tous les documents que le système lui présente.

Enfin, une autre fenêtre sert à la navigation au sein de la carte navigationnelle. Cette fenêtre affiche, sous forme graphique, le chemin de recherche que l'utilisateur a déjà parcouru. Cet affichage permet à l'utilisateur de naviguer au sein de sa recherche en lui permettant, par exemple, de revenir sur un concept et d'explorer une nouvelle direction. De plus, cette fenêtre est également utilisée pour représenter les chemins de recherche d'utilisateurs antérieurs comme nous l'avons vu précédemment.

Ce multi-fenêtrage permet donc une recherche plus structurée et ainsi plus cohérente, en disposant de toutes les informations et de toutes les fonctionnalités des cartes navigationnelles sur le même écran. L'analyse et le développement de ces cartes détermineront les limites de cette représentation.

7. Bibliographie

[Aigrain P. and Longueville V. 1992]

Evaluation of navigational Links between images.

In Information Processing & Management, 1992, 28 (4), pp.517-528.

[Bauer G. 1989]

Knowledge structures promoting problem solutions. Possibilities of graphic representation of knowledge in thesauri-like fashion.

In International Classification, 1989, 16 (2), pp.73-81.

[Bor-Sheng T. 1993]

Mapping metrics for subject-object coordinated field recognition and representation.

In Library Science with a Slant to Documentation and Information Studies, 1993, 30 (3), pp.91-102.

[Byrne J. 1975]

Relative effectiveness of titles, abstracts, and subject headings for machine retrieval from the COMPENDEX Services.

In Journal of the American Society for Information Science, July-August 1975, 26, pp.223-229.

[Cahn DF. 1980]

Computer-aided visualization of database structural relationships.

In American Society for Information Science proceedings, 1980, 17, pp.358-360.

[Callon M. and Courtial J.P. and Laville F. 1989]

La méthode des mots associés : un outil pour l'évaluation des programmes publics de recherche (Le cas des polymères)

Rapport du CSI/Ecole des Mines de Paris, 1989.

[Callon M. and Courtial J.P. and Pénan H. 1993]

La scientométrie

Que sais-je n°2727, 1993, p.126.

[Chaumier J. 1979]

Le traitement des thesaurus à schémas fléchés par l'informatique graphique interactive.

In Documentaliste, 1979, 16 (1), pp.9-14.

[Chaumier J. 1988]

Le traitement linguistique de l'information

Entreprise moderne d'édition, 1988, p.186.

[Craven T. 1980]

Microcomputer-generated graphic displays as an aid in string indexing.

In Journal of the American Society for Information Science, 1980, 31 (2), pp.123-124.

[Craven T. 1990a]

Condensed representation of sentences in graphic displays of text structures.

In Journal of Documentation, 1990, 46 (4), pp.339-352.

[Craven T. 1990b]

Automatic structure modification in the graphic display of thesauri.

In Proceedings of the 1st International ISKO-Conference, Darmstadt, 14-17 August 1990, pp.146-153.

[Craven T. 1991a]

Algorithms for graphic display of sentence dependency structures.

In Information Processing & Management, 1991, 27 (6), pp.603-613.

[Craven T. 1991b]

Graphic display of larger sentence dependency structures.

In Journal of the American Society for Information Science, 1991, 42 (5), pp.323-331.

[Craven T. 1992a]

A general versus a special algorithm in the graphic display of thesauri.

In Classification research for knowledge representation and organization. Proceedings of the 5th International Study Conference on Classification Research, 1992, pp.179-185.

[Craven T. 1992b]

Concept relation structures and their graphic display.

In Classification research for knowledge representation and organization. Proceedings of the 5th International Study Conference on Classification Research, 1992, pp.49-59.

[Deweze A. 1981]

Réseaux sémantiques. Application à l'indexation et à la recherche de l'information documentaire

Thèse de l'université Claude Bernard Lyon I, 1981, p.451.

[Doyle L. 1962]

Indexing and abstracting by association.

In American Documentation, 1962, 13 (4), pp.378-390.

[Duncan E. and McAleese R. 1987]

Intelligent access to databases using a thesaurus in graphical form.

In Online information 87. Proceedings of the 11th International Online Information Meeting, London, 8-10 December 1987, pp.377-387.

[Duncan E. 1990]

A concept-map thesaurus as a knowledge-based hypertext interface to a bibliographic database.

In Jones, KP: Prospects for information retrieval: Informatics 10: proceedings of a conference jointly sponsored by Aslib, the Aslib Informatics Group, and the Information Retrieval Specialist Group of the British Computer Society, King's College, Cambridge, 21-23 March 1989. London: Aslib. 1990. pp.43-52

[Edmonds E. and O'Brien S. and Bayley T. and McDaid E. 1993]

Construction end-user knowledge manipulation systems.

In International Journal of Man-Machine Studies, 1993, 38 (1), pp.51-70.

[Eisenstatd M. and Domingue D. and Rajan J. and Motta T. 1990]

Visual knowledge engineering.

In IEEE transactions on software engineering, 1990, 16 (10), pp.1164-1177.

[Feiner S. 1998]

Seeing the forest for the trees: hierarchical display of hypertext structure.

In Conference on Office Automatisation Systems, Palo Alto, California, March 23-25 1988, pp.205-212.

[Fondin H. 1982]

Le titre comme élément de description du contenu d'un document. Recherche sur les méthodes d'évaluation.

In Documentaliste, 1982, 19 (1), pp.3-15.

[Fowler R. and Fowler W. and Wilson B 1991]

Integrating query, thesaurus, and documents through a common visual representation.

In ACM, 1991, pp.142-151.

[Freivalds D. and Carson S. 1991]

The role of microcomputer in the evolution of the OPAC.

In Library Software Review, 1991, 10 (2), pp.110-114.

[Fugman R. 1974]

Representation of concept relations using the TOSAR system of the IDC: treatise iii on information retrieval theory.

In Journal of the American Society for Information Science, 1974, 25 (5), pp.287-307.

[Furnas G.W. 1986]

Generalized fisheye views.

In Proceedings of Human Factors in Computing Systems (CHI'86), 1986, pp.16-23.

[Gastaldy S. and Davidson C. 1986]

Improved design of graphic display in thesauri through technology and ergonomics.

In Journal of Documentation, 1986, 42 (4), pp.225-251.

[Georgel A. 1991]

ACLA-Leximappe: comparaison de deux méthodes statistiques appliquées aux données documentaires.

In Revue française de bibliométrie, 1991, 9, pp.81-97.

[Gilyarevskii R. and Subbotin M. 1993]

Russian experience in Hypertext: Automatic compiling of coherent texts.

In Journal of the American Society for Information Science, 1993, 44 (4), pp.185-193.

[Gredley E. and Hopkinson A. 1990]

Exchanging bibliographic data. MARC and other international formats

Published by Library Association Publishing Ltd, 1990, p.329.

[Guillot M. 1971]

Utilité de la représentation d'un langage documentaire sous forme graphique du type "schéma fléché".

In Documentaliste, 1971, 8 (1), pp.19-24.

[Kimoto H. and Iwadera T. 1990]

Construction of a dynamic thesaurus and its use for associated information retrieval.

In 13th International Conference on research and development in information retrieval, 1990, pp.227-240.

[Komorowski HJ. 1988]

Browsing and authoring tools for a unified medical language system.

In RIA088. Centres de Hautes Etudes Internationales d'Informatique Documentaire, 1988, pp.624-641.

[Lin X. and Liebscher P. and Marchionini G. 1991]

Graphical representations of electronic search patterns.

In Journal of the American Society for Information Science, 1991, 42 (7), pp.469-478.

[Lowe R. 1993]

Diagrammatic information: techniques for exploring its mental representation and processing.

In Information Design Journal, 1993, 7 (1), pp.3-17.

[Michelet B. 1988]

L'analyse des Associations

Thèse de Doctorat (nouveau régime). Université de Paris VII, U.F.R. de Chimie.

[Nazim Ali S. 1993]

Subject relationship articles determined by co-occurrence of keyword in citing and cited titles.

In Journal of Information Science, 1993, 19, pp.225-232.

[Osipov G. 1992]

Formulation of subject domain models. Part I. heterogeneous semantic nets.

In Soviet Journal of Computer and System Sciences, 1992, 30 (2), pp.1-12.

[Peat H.J. and Willett P. 1991]

The limitations of term co-occurrence data for query expansion in document retrieval systems.

In Journal of the American Society for Information Science, 1991, 42 (5), pp.378-383.

[Pejtersen A. 1992]

New model for multimedia interfaces to online public access catalogues.
In Electronic Library, 1992, 10 (6), pp.359-366.

[Rolling L. 1965]

The role of graphic display of concept relationships in indexing and retrieval vocabularies.
In Atherton P. : Classification Research: proceedings of the Third International Conference on Classification Research, Bombay, India, 6-11 January, 1975? Sarad Ranganathan Endowment for Library Science, 1979, pp.188-203.

[Searc P. 1993]

HyperGlyphs : using design and language to define hypermedia navigation.
In Technical Communication, 1993, 40 (3), pp.414-421.

[Shabahat H. and O'Brien A. 1992]

Recent trends in subject access to OPACs : an evaluation.
In International Classification, 1992, 19 (3), pp.140-145.

[Slack F. 1983]

Subject searching on OPACs: problems and help provision.
In Vine 1983, pp.4-9.

[Tanyukhin I. 1994]

Sufficient criteria for inheritance correctness of attributes in semantic nets.
In Programming and Computer Software (US), 1994, 19 (5), pp.238-240.

[Tsai B. 1994]

Formulas for constructing an animation-supported navigation system.
In Journal of Library and Information Science, 1994, 20 (1), pp.1-26.

[Turner W.A. 1991]

Scientometrics in France.
In Numéro Spécial de Scinetometrics, 1991, ed. Elsevier.

[Ungvary R. 1983]

Application of the thesaurus method to the communication of knowledge.
In International Classification, 1983, 10 (2), pp.63-68.

[Van Rijsbergen C. 1977]

A theoretical basis for the use of co-occurrence data in information retrieval.
In Journal of Documentation, 1977, 33 (2), pp.106-119.

[Watanabe H. 1989]

Heuristic graph displayer for G-BASE.
In International journal of man-machine studies, 1989, 30 (3), pp.287-302.

[Weiland W. and Shneiderman B. 1993]

A graphical query interface based on aggregation/generalization hierarchies.
In Information Systems, 1993, 18 (4), pp.215-232.

BIBLIOTHEQUE DE L'ENSSIB



8022163