

## Dominique Ménillet

Institut national de l'information scientifique et technique

# THÉSAURUS ET INDEXATION

**C**E SONT des documentalistes représentant des producteurs de bases de données scientifiques et techniques et des utilisateurs qui constituent le groupe qui se consacre aux travaux concernant l'analyse, l'indexation et des langages documentaires.

### Analyse et indexation

L'indexation est l'opération qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans ce document, c'est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse.

La transcription en langage documentaire se fait grâce à des outils d'indexation qui appartiennent à deux grands types :

– *classificatoire* : les éléments d'information sont classés selon un

ordre logique regroupant les notions plus particulières sous des notions plus générales. Il y a des emboîtements de classes l'une dans l'autre ;

– *combinatoire* : les notions sont rangées par ordre alphabétique des mots qui les expriment. Ces outils combinatoires rassemblent un ensemble de termes, descripteurs ou mots clés, pouvant se combiner entre eux lors de l'indexation pour exprimer des notions complexes.

Ces derniers ont l'avantage de permettre une description fidèle du

contenu du document sans la contrainte d'une structure de classification préalable qui rend difficile l'introduction de tout aspect nouveau.

La recherche bibliographique automatisée et l'utilisation de la logique booléenne pour formuler les questions ont favorisé la recherche par mots clés. Les performances de cette recherche (pertinence et rappel) sont directement en liaison avec la qualité de l'indexation, c'est-à-dire de la saisie des données. Ceci implique des

#### Liste des normes ISO et AFNOR

<b>ISO 2788</b> : 1986 Documentation - Principes directeurs pour l'établissement et le développement de thésaurus monolingues	pour l'analyse des documents, la détermination de leur contenu et la sélection des termes d'indexation	pour l'établissement des thésaurus multilingues
<b>NF Z 44-004</b> : 1984 Recommandations aux auteurs des articles scientifiques et techniques pour la rédaction des résumés	<b>NF Z 47-102</b> : 1978 Principes généraux pour l'indexation des documents et EP jusqu'au 20 décembre 1992	<b>Z 47-103</b> : 1980 Thésaurus monolingues et multilingues : symbolisation des relations
<b>NF Z 47-100</b> : 1981 Règles d'établissement des thésaurus monolingues	<b>ISO 5964</b> : 1985 Documentation - Principes directeurs pour l'établissement et le développement de thésaurus multilingues	<i>Pour mémoire</i> <b>Z 44-070</b> Documentation - Indexation analytique par matière
<b>ISO 5963</b> : 1985 Documentation - Méthodes	<b>Z 47-101</b> : 1990 Principes directeurs	<b>Z 47-200</b> Documentation - Liste d'autorité de matières - Structure et règles d'emploi

\* Suite à la réorganisation du comité technique TC 46 de l'ISO, la commission CN5 « Thésaurus et règles d'indexation » a été regroupée en 1987 avec deux autres commissions, la CN6 « Eléments de données bibliographiques » et la CN7 « Présentation des publications ». C'est ainsi qu'a été créée, en 1987, la CN9 « Présentation, identification et description des documents », qui couvre un domaine très vaste.

règles précises d'indexation d'une part, et l'utilisation d'un langage documentaire normalisé d'autre part.

## Les thésaurus

C'est dans cette nécessité que se sont développés les outils linguistiques de type combinatoire dont le principal est le thésaurus.

Un thésaurus est un ensemble de termes appartenant à un domaine précis de la connaissance et représentant chacun un concept ou un ensemble de concepts. Les termes sont reliés entre eux par des relations qui permettent de déterminer leur environnement sémantique. Ces relations sont de trois ordres : relations d'équivalence, relations hiérarchiques et relations associatives. La valeur d'un thésaurus en tant qu'outil documentaire réside moins dans le choix de ses termes – toujours empreint d'une certaine convention – que dans le choix des relations conceptuelles qui en définissent les modalités d'utilisation.

Dans ce contexte, le groupe de travail « thésaurus et indexation » s'est fixé comme but d'établir des documents qui fassent référence pour lutter contre l'anarchie dans ce domaine et qui puissent être utilisés à des fins didactiques. Quatre normes ont ainsi été élaborées, qui sont plutôt des recommandations et / ou des énoncés de principes directeurs. Elles ont un enchaînement logique :

- règles d'établissement des thésaurus monolingues = norme NF Z 47-100 ;
- principes directeurs pour l'établissement des thésaurus multilingues = norme NF Z 47-101 ;

- principes généraux pour l'indexation des documents = norme NF Z 47-102 ;
- recommandations aux auteurs des articles scientifiques et techniques pour la rédaction des résumés = norme NF Z 44-004.

Pour les thésaurus monolingues, l'important est la définition des relations entre les termes. De ces relations découle la structure du thésaurus avec la possibilité de l'utiliser automatiquement, par exemple, pour la génération automatique des termes génériques à partir des termes spécifiques (autopostage) ou pour l'interrogation avec les synonymes d'un descripteur.

par des concepts et les mots clés correspondants. Ils devraient permettre une coopération plus facile entre différents services d'information et l'établissement de règles adaptées à chacun des systèmes particuliers.

L'ensemble thésaurus-règles d'indexation devrait servir de base à la constitution de systèmes experts et de bases de connaissances pour l'indexation ou la recherche des documents ; la connaissance de ces normes est un préalable indispensable.

Enfin, si l'indexation doit s'effectuer automatiquement, un résumé informatif bien rédigé en devient le sup-

---

### *Pour les thésaurus monolingues, l'important est la définition des relations entre les termes*

---

Les thésaurus multilingues ont été élaborés en raison de l'importance des collaborations entre producteurs de bases de données de différents pays et aussi de la volonté de fournir aux utilisateurs ne parlant pas la langue d'indexation la possibilité de l'utiliser à l'interrogation.

Le problème essentiel est celui de l'équivalence entre termes de langues différentes qui va de l'équivalence exacte à l'équivalence inexacte ou partielle à la non-équivalence (aucun terme équivalent n'existant dans la langue).

Les principes généraux pour l'indexation des documents ont pour objet d'établir une méthodologie cohérente pour la caractérisation d'un document

port essentiel. C'est dans cette optique que s'inscrivent les recommandations aux auteurs des articles scientifiques et techniques pour la rédaction des résumés. Le résumé permet au lecteur d'identifier rapidement l'intérêt d'un document et permet aussi, par son intégration comme tel dans une base de données, d'être interrogeable directement, tout comme les mots clés.

Ces quatre normes françaises ont été élaborées avec le même souci : mieux construire les systèmes d'enregistrement de l'information pour mieux répondre aux besoins des utilisateurs.

Mai 1993

### Commission Description bibliographique : AFNOR Z44A

#### LISTE DES MEMBRES

**Président** : Françoise Leresche, Bibliothèque nationale, CCBT

**Secrétaire** : Jean-Claude Hesling-AFNOR

Geneviève Boisard, Bibliothèque Sainte-Geneviève (Paris)

Mlle Bosc, Bibliothèque nationale, CCBT

Mme Boulogne, Françoise Boustouler, Bibliothèque interuniversitaire Cujas (Paris)

Mlle Branger, Bibliothèque nationale, Phonothèque

Marie-France Calas, Bibliothèque nationale, Phonothèque

Martine Comberousse, Ministère de l'Enseignement supérieur et de la Recherche  
M. Lebaron, DIST

Marie-Renée Cazabon, Bibliothèque de France

Pierre-Yves Duchemin, Bibliothèque nationale, Cartes et plans

Mme Gana, Bibliothèque nationale, Dépôt légal

Mme Gauchet, Bibliothèque nationale, CNEPS

Mme Ghersin, Bibliothèque Forney (Paris)

Elisabeth Giuliani, Bibliothèque de France

Annette Grousseau, Bibliothèque d'art et d'archéologie

Mme Guinamard, La Documentation française, Bibliothèque

Souad Hubert, BPI - CNAC Georges Pompidou

Mme Huby, Institut Français du pétrole

Mme Landrieau, Ministère de l'Économie et des Finances

Danielle Le Nan, Bibliothèque nationale, Périodiques

Françoise Leresche, Bibliothèque nationale, CCBT

Catherine Mattenet, AFNOR

M. Menage, Cercle de la librairie, SBD

Céline Ménil, Ministère de l'Enseignement supérieur et de la Recherche, DISTB

Marion Mouchot, Bibliothèque nationale, CCBT

Monique Pelletier, Bibliothèque nationale, Cartes et plans

Mme Pross, Fondation nationale des sciences politiques

Mme Robberechts, Fondation nationale des sciences politiques

Suzanne Santiago, Centre international d'enregistrement des publications en série (ISDS)

Pascal Sanz, Centre national de la documentation pédagogique

Nicole Simon, Bibliothèque nationale, Service étranger

Mme Sumian, Bibliothèque municipale (Annecy)

Mme Thompson, Bibliothèque nationale, Estampes

Catherine Trémauville, Bibliothèque de la Sorbonne (Paris)

Corinne Verry, Institut Pasteur, Bibliothèque

Maria Witt, Cité des sciences et de l'industrie

#### 1 DESCRIPTION BIBLIOGRAPHIQUE

**Z 44-050**, Déc. 1989, Documentation - Catalogage des monographies - Rédaction de la description bibliographique

**NF Z 44-063**, Avr. 1979, Documentation - Catalogage des publications en série - Rédaction de la notice bibliographique

**Z 44-065**, Juin 1980, Documentation - Catalogage des images animées - Rédaction de la notice bibliographique

**Z 44-066**, Déc. 1988, Documentation - Catalogage des enregistrements sonores - Rédaction de la notice phonographique

**Z 44-067**, Sept. 1991, Documentation - Catalogage des documents cartographiques - Rédaction de la description bibliographique

**Z 44-073**, Mai 1991, Documentation - Catalogage des monographies - Rédaction de la description bibliographique allégée

**Z 44-074**, Oct. 1986, Documentation - Catalogage des monographies anciennes - Rédaction de la description bibliographique

#### 2 ACCÈS À LA DESCRIPTION BIBLIOGRAPHIQUE

**Z 44-059**, Déc. 1987, Documentation - Catalogage - Choix des accès à la description bibliographique

**NF Z 44-060**, Oct. 1983, Documentation - Catalogage d'auteurs et d'anonymes - Forme et structure des vedettes de collectifs- auteurs

**NF Z 44-061**, Juin 1986, Documentation - Catalogage - Forme et structure des vedettes noms de personnes, des vedettes titres, des rubriques de classement et des titres forgés

**NF Z 44-070**, Août 1986, Documentation - Indexation analytique par matière

#### 3 CLASSEMENT BIBLIOGRAPHIQUE

**NF Z 44-062**, Juil. 1963, Documentation - Classement des noms propres étrangers comportant des particules ou d'autres éléments accessoires - Répertoire de quelques usages nationaux

**Z 44-080**, Fév. 1986, Documentation - Règles de classement bibliographique

#### LISTE DES PROJETS À L'ÉTUDE

##### 1 Révision

**Z 44-065** - Documentation - Catalogage des images animées - Rédaction de la notice bibliographique

**Z 44-068** - Documentation - Catalogage des documents cartographiques - Manuel d'application du fascicule de documentation Z 44-067 et de la norme NF Z 44-081

**Z 44-072-1** - Documentation - Notice signalétique - Partie 1 : Identification d'un document pour le prêt entre bibliothèques

**Z 44-072-2** - Documentation - Notice signalétique - Partie 2 : Identification d'un document pour une acquisition

##### 2 Création

**Z 44-069** - Documentation - Catalogage de la musique imprimée - Rédaction de la notice bibliographique

**Z 44-077** - Documentation - Catalogage de l'image fixe - Rédaction de la description bibliographique

**Z 44-078** - Documentation - Catalogage des parties composantes - Rédaction de la description bibliographique

**Z 44-079** - Documentation - Catalogage - Forme et structure des vedettes titres musicaux

**Z 44-081** - Documentation - Catalogage - Forme et structure des vedettes noms géographiques

**Z 44-082** - Documentation - Catalogage des documents électroniques - Rédaction de la description bibliographique