

Motbis

de l'analyse à la recherche documentaire

Sylvie Moussaoui et François Feyler

DRT – DIVISION INGÉNIERIE ÉDUCATIVE ET DOCUMENTAIRE, DIVISION LANGAGES DOCUMENTAIRES
CNDP

*Histoire et perspectives du thésaurus
spécifiquement conçu pour l'Éducation
nationale, désormais familier
aux professeurs-documentalistes.*

La norme AFNOR définit un thésaurus comme une « liste d'autorité organisée de descripteurs et de non-descripteurs obéissant à des règles terminologiques propres et reliés entre eux par des relations sémantiques (hiérarchiques, associatives, ou d'équivalence). Cette liste sert à traduire en un langage artificiel dépourvu d'ambiguïté des notions exprimées en langage naturel¹ ».

Un thésaurus est donc un langage documentaire contrôlé assurant une homogénéité du vocabulaire lors de l'analyse du contenu intellectuel d'un ensemble de documents. Ce langage contrôlé permet au final une recherche efficace sur ce corpus de références documentaires.

Motbis est le thésaurus de l'Éducation nationale. Son premier champ d'application est l'établissement scolaire et les documents qui y sont décrits. Son corpus rassemble les termes utiles à la description de ces documents. D'une part, il répond au vocabulaire de l'Éducation nationale, de sa structuration, de son fonctionnement par disciplines et niveaux ou par notion d'activité pédagogique ; d'autre part, il possède une vocation encyclopédique puisque les documents utilisés dans les établissements scolaires par les enseignants et pour les élèves ne sont pas uniquement des référentiels pédagogiques mais également des ouvrages documentaires, monographies ou articles de périodiques, traitant de l'ensemble des sujets abordés au cours d'une sco-

larité ou nécessaires à la constitution d'une culture personnelle.

Historique

Motbis existe depuis maintenant quinze ans : sa version 1, éditée par le CNDP en octobre 1989, a été conçue sur le socle de Thélyce, publié par les CRDP de Lyon et de Grenoble, et de Mémobase, créé par le CRDP de Poitou-Charentes pour répondre aux besoins des indexeurs qui dépouillaient les articles des premières MémoFiches.

Il est actuellement utilisé dans plus de 10 000 centres de documentation français : CDI des collèges et lycées, centres du réseau Scérén, réseau des CAFA et la plupart des IUFM (dans ce dernier cas, il est souvent utilisé avec RAMEAU).

Depuis sa première édition, il a été employé pour l'indexation de plusieurs centaines de milliers de références documentaires et, en particulier pour celle des MémoFiches qui, en septembre 2004, représentent 143 291 notices de dépouillement d'articles de périodiques, établies depuis 1988 (1986 pour la version imprimée) et diffusées dans plus de 5 000 CDI.

Notons à ce sujet que les plus anciennes de ces MémoFiches, plusieurs milliers, ont été indexées avec le thésaurus Mémobase, avant même la parution de Motbis 1 : pour ces références les plus anciennes, Motbis 2005 constitue donc aujourd'hui le cinquième langage documentaire (après Mémobase, Motbis 1, Motbis 2 et Motbis 3), avec autant de réindexations automatiques de l'un à l'autre.

Les statistiques d'indexation (Motbis 3.1 et Id-Mémo²) des 130 000 notices (correspondant à un nombre de 250 000 à 500 000 occurrences de descripteurs) présentes dans la base MémoFiches fin septembre 2003 ont été recensées dans la perspective d'une approche analytique pour l'évolution de Motbis.

Sur les 9 650 descripteurs correspondant à l'ensemble Motbis 3/Id-Mémo, environ 1 560 n'avaient jamais été utilisés alors que près de 700 l'avaient été 100 fois ou plus, dont 28 plus de 1 000 fois. Par rapport à un thésaurus et à une base de données

**« Depuis la
généralisation des accès à
Internet [...], le champ
d'intervention de Motbis
déborde le CDI en les murs
pour "s'attaquer" aux
documents en ligne. »**

1. AFNOR, 1996, NF Z 47-100, p. 479.

2. Id-Mémo : complément de termes (noms de personnes, d'organismes et de sites) à Motbis.

3. Dico : dictionnaire thématique d'interprétation du plein texte.



bibliographique de cette importance, l'idéal serait d'avoir des descripteurs utilisés chacun entre 10 et 100 fois. C'est la notion de « concept utile ».

Repérer l'information pertinente

Motbis, comme les autres thésaurus, a vu son utilité s'accroître avec le développement de la recherche informatisée. Les logiciels permettaient, dans un premier temps, la recherche sur l'unique champ « mots-clés ». L'apport d'un langage contrôlé pour l'indexation était alors évident. Cependant la possibilité, apparue au début des années 90 sur certains logiciels de recherche documentaire, d'effectuer des recherches en plein texte sur un ensemble de champs (mots du titre, du résumé et de l'indexation, contrôlée ou non...) a abouti à trois constats :

- les recherches effectuées uniquement sur une zone d'indexation contrôlée par le langage d'indexation, si elles sont généralement très satisfaisantes en termes de lutte contre le bruit documentaire, sont néanmoins toujours productrices de silence documentaire ;

- les recherches effectuées en plein texte exhumant souvent des références documentaires tout à fait pertinentes qui ne pourraient pas être retrouvées par l'intermédiaire du langage documentaire implanté ;

- les chances d'approcher une certaine exhaustivité dans la recherche sont augmentées si l'utilisateur sait exploiter de façon conjointe l'autopostage (dans le cas d'une recherche thématique avec un thésaurus) et la troncature, tout en restant quelque peu défensif par rapport à l'ambiguïté potentielle de certaines chaînes de caractères. Néanmoins, ce type de recherche « experte » reste un but à atteindre dans une démarche de formation, mais n'est, de toute évidence, pas acquise pour la plupart des élèves, au début de leur cursus secondaire tout au moins.

Le schéma classique d'utilisation des langages documentaires, outils uniques et identiques pour l'indexation et la recherche des références pertinentes, montre lui-même ses limites : il y a en effet une contradiction apparente entre la néces-

sité d'élaborer des langages documentaires utilisés par le plus grand nombre de centres possible pour des raisons d'échange de références documentaires et les besoins particuliers de catégories d'utilisateurs spécifiques. Cette contradiction peut être surmontée si l'on accepte de différencier la notion de « langage d'indexation » de celle « d'interface d'interrogation ». Il est nécessaire que ces interfaces soient élaborées en cohérence avec les langages d'indexation préexistants et que les représentations de concepts contenues dans ces derniers soient sans ambiguïté.

Évolution des langages documentaires

C'est dans cette démarche que s'inscrit le CNDP. En effet, outre l'évolution conjointe du thésaurus Motbis et de l'outil d'interprétation du plein texte Dico³, qui pourraient respectivement devenir TFIDE (Thésaurus francophone pour l'indexation documentaire pour et sur l'éducation) et DFRIE (Dictionnaire français pour la recherche d'information pour et sur l'éducation), il semble souhaitable de prévoir à plus long terme l'intégration de ces deux outils dans le cadre d'un service en ligne SFIDE (Système français pour l'information documentaire pour et sur l'éducation), outil de recherche de l'information et d'aide à l'indexation, donnant accès à un ensemble d'informations et de liens avec des données et des services web relatifs aux notions présentes dans les langages documentaires gérés par le département des Ressources et des Technologies du CNDP (thésaurus, nomenclatures, contenu des métadonnées).

Depuis la généralisation des accès à Internet dans les établissements scolaires, le champ d'intervention de Motbis déborde le CDI en les murs pour « s'attaquer » aux documents en ligne. Or, le Web, en raison de son incroyable profusion, pourrait exploiter avec profit un langage documentaire plus encore que les supports traditionnels de l'information. Les moteurs de recherche, qui se développent actuellement dans les domaines de la linguistique, exploreraient utilement cette piste des langages contrôlés. ●

