

E.N.S.S.I.B
ECOLE NATIONALE SUPERIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHEQUES

UNIVERSITE
CLAUDE BERNARD
LYON I

DESS en INFORMATIQUE DOCUMENTAIRE

Note de Synthèse

Les logiciels de création et de gestion de thesaurus

Par Majd KAADE

Monsieur Jean-Pierre LARDY

1992

TD
35

E.N.S.S.I.B
ECOLE NATIONALE SUPERIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHEQUES

UNIVERSITE
CLAUDE BERNARD
LYON I

DESS en INFORMATIQUE DOCUMENTAIRE

Note de Synthèse



Les logiciels de création et de gestion de thesaurus

Par Majd KADE

Monsieur Jean-Pierre LARDY

1992

ED

35

1992

LES LOGICIELS DE CREATION ET GESTION DE THESAURUS

Majd KAADE

RESUME

En 1947 l'histoire du thesaurus a débuté, en 1959 le thesaurus a pris sa forme actuelle. Les dernières années ont vu l'apparition de logiciels de création et gestion de thesaurus qui ont facilité la tâche fastidieuse de réaliser un thesaurus. Ces logiciels varient dans leurs capacités à gérer certaines fonctions qui leurs sont demandées.

DESCRIPTEURS

Logiciel, lexique.

ABSTRACT

The thesaurus history started in 1947. It take his current shape in 1959. The latest years have witressed the comming of software to create and manage that have made easier the dull lask of realizing a thesaurus. This software fluctuate in their ability to manage some functions that are wanted to then.

KEYWORD

Software, thesaurus.

REMERCIEMENTS

Je remercie vivement Monsieur Jean-Pierre LARDY de m'avoir proposé ce sujet et enrichie de ses bons conseils.

TABLE DES MATIERES

INTRODUCTION.....	1
STRATEGIE DE RECHERCHE.....	2
1. LA RECHERCHE MANUELLE.....	2
2. LA RECHERCHE AUTOMATISEE.....	2
a. PRESENTATION DE LA BASE DE DONNEES PASCAL.....	3
b. PRESENTATION DE LA BASE DE DONNEES INSPEC.....	3
c. PRESENTATION DE LA BASE DE DONNEED LISA.....	4
d. PRESENTATION DE LA BASE DE DONNEES ISA.....	4
e. EQUATIONS DE RECHERCHE.....	4
I. L'HISTOIRE DU THESAURUS.....	7
1. LE DICTIONNAIRE DE MOOERS.....	7
2. LES ASSOCIATIONS DE BERNIER ET CRANE.....	8
3. LE THESAURUS DE LUHN.....	9
4. LES THESAURUS DE BERNIER.....	10
a. LE THESAURUS LIMITE.....	10
b. LE THESAURUS TECHNIQUE.....	10
c. L'INDEX FIGURE.....	10
5. CONTRIBUTIONS PRATIQUES.....	11
6. LA CONTRIBUTION ANGLAISE.....	11
7. THESAURUS OPERATIONNEL.....	13

II. LES LOGICIELS DE GESTION DE THESAURUS.....	14
1. CHAMPS.....	15
2. RELATIONS.....	16
a. TYPE DE RELATIONS.....	16
b. RECIPROCITE DES RELATIONS.....	17
c. CONTROLES.....	18
d. PREVENTION DE L'ENCOMBREMENT.....	18
3. MODIFICATION DES DONNEES.....	18
4. RECHERCHE DES DONNEES D'UN THESAURUS.....	20
5. SORTIE ET IMPRESSION.....	20
a. AFFICHAGE A L'ECRAN.....	21
b. SORTIE PAR IMPRIMENTE.....	22
6. AUTRES DOMAINES.....	23
a. CHANGEMENTS GLOBAUX.....	23
b. APPROBATIONS.....	23
c. ORPHELIN.....	23
d. CLASSIFICATION DES TERMES.....	24
e. TRI.....	24
f. INTERFACE DE L'UTILISATEUR.....	24
g. MESSAGE D'ERREUR.....	25
h. NOTES.....	25
i. DATES.....	25
j. REACTION.....	26
k. ACTUALISATION DU VOCABULAIRE.....	26
l. INTRODUCTION ET TRAITEMENT EN DIRECT PAR RAPPORT AU TRAITEMENT DE LOTS.....	26

III. LES LOGICIELS DISPONIBLES.....	27
1. LOGICIELS DOCUMENTAIRES INTEGRENT LA GESTION DE THESAURUS.....	28
2. LOGICIELS AUTONOMS.....	34
CONCLUSION.....	41
BIBLIOGRAPHIE.....	42
ANNEXES.....	43

INTRODUCTION

Les documentalistes ayant besoin d'une aide à l'indexation et à l'interrogation, ils ont essayé durant des années de transformer les dictionnaires habituellement utilisés en un outil efficace et pratique à la documentation. Différentes idées ont été adaptées pour élaborer cet outil.

Les logiciels de gestion de thesaurus ont commencé à apparaître dans les dernières années, à la suite de changements et de développements technologiques. Quelques uns de ces logiciels sont disponibles sur le marché, d'autres sont en cours d'élaboration.

Le premier chapitre de cette synthèse présente l'historique du développement du thesaurus ; le deuxième chapitre présente les critères d'évaluation d'un logiciel de gestion de thesaurus ; le troisième est consacré aux logiciels disponibles sur le marché.

Enfin, les lecteurs voudront bien m'excuser pour les fautes de français se trouvant dans le texte de ce document.

STRATEGIE DE RECHERCHE

Le demandeur m'a prêté des articles de périodiques ; ils ont été un support fort intéressant sur lequel l'étude s'est appuyée.

1. LE RECHERCHE MANUELLE :

J'ai consulté la documentation de la bibliothèque de l'ENSSIB souhaitant trouver des informations concernant le sujet, ainsi que les périodiques de la Bibliothèque Universitaire de la DOUA.

Les résultats de ma recherche manuelle ont été des informations très générales.

2. LA RECHERCHE AUTOMATISEE :

la première étape a consisté à ~~était de~~ consulter le Répertoire des banques de données professionnelles et le Guide des banques de données professionnelles ASCII et vidéotex. Grâce à eux j'ai choisi les bases de données concernant le sujet en question.

Quatre bases de données ont été interrogées : PASCAL, INSPEC, LISA, INFORMATIONS SCIENCES ABSTRACTS.

a. PRESENTATION DE LA BASE DE DONNEES PASCAL :

Elle est produite par l'INIST - CNRS, les principaux domaines sont :

Sciences physiques, sciences de l'ingénierie, chimie pure et appliquée, physique, chimie, cristallographie, sciences de la terre, sciences de la vie et médecine, sciences de l'information, technologie, sciences exactes et sciences appliquées.

Analyse de tous les articles de périodiques majeurs français et étrangers ainsi que de rapports scientifiques, thèses, comptes-rendus de congrès.

Elle peut être interrogée en français, anglais et espagnol à partir de 1987.

b. PRESENTATION DE LA BASE DE DONNEES INSPEC :

Elle est produite par l'Institution of Electrical Engineers (IEE), Royaume-Uni.

Elle couvre les domaines :

Électricité et ingénierie de contrôle : Physique atomique et moléculaire ; programme, application, systèmes et équipements informatiques, propriétés électriques, magnétiques et optiques ; propriétés structurales, thermiques et mécaniques, technologie de contrôle ; appareillages électriques et magnétiques ; circuits électromagnétiques et optique ; gaz, dynamiques des fluides et plasmas ; physique des particules élémentaire ; mesures et instruments ; sciences et technologies de l'information et de la communication ; mathématiques, physique nucléaire ; systèmes de puissance et applications.

Elle est constituée à 80% d'articles de périodiques et de 20% d'actes de congrès, rapports techniques, ouvrages, brevets, thèses universitaires.

Langue de la base anglais.

c. PRESENTATION DE LA BASE DE DONNEES LISA :

Domaines Sciences de l'information
Sciences de l'information : automatisation, bibliothéconomie,
distribution du livre, édition, reproduction, stockage et
recherche de l'information, utilisateurs.

Données : 550 périodiques, ouvrages, communications
scientifiques, rapports.

Début : 1969.

d. PRESENTATION DE LA BASE DE DONNEES ISA :

Domaines Sciences de l'information
Bibliothéconomie ; catalogage ; classification ; édition,
enseignement de l'information ; indexation ; recherche de
l'information ; reproduction ; résumé, sciences de
l'information ; stockage de l'information ; technologie des
microformes ; télécommunications, traduction.

Données : 450 périodiques, ouvrages, rapports, compte-rendu de
congrès, brevets.

Début : 1966.

e. EQUATIONS DE RECHERCHE :

L'interrogation a été faite sur les bases de données
groupées au moyen de la commande ONESEARCH de DIALOG :

b2,144,61,202

Une commande qui permet d'effectuer la recherche sur les quatre bases de données.

STHESAURU? (W) SOFTWARE? ?

J'ai obtenu 26 réponses.

En faisant

RD

Remote Duplicat, une commande qui permet de supprimer les doublons, le nombre de réponses s'est réduit à 13.

Parmi les 13 références obtenues, 7 étaient les plus intéressantes ; elles concernaient essentiellement le sujet. 2 d'entre elles sont les mêmes que j'ai empruntés du demandeur. Mon ignorance de l'allemand m'a empêché de lire une référence, ainsi qu'une autre parmi les 6 références restantes, et non pertinentes.

En plus, la base de données PASCAL a été interrogée de la façon suivante :

* Sur CD-ROM à partir de 1986 à la Bibliothèque Universitaire de la DOUA.

* En ligne de 1980 à 1986 pendant les TD à l'ENSSIB.

Selon l'équation de recherche déjà mentionnée :

STHESAURU? (W) SOFTWARE? ?

S1/1980:1986.

Les références obtenues par l'interrogation de la base de données PASCAL sur CD-ROM étaient au nombre de 25 références.

* 3 références pertinentes, deux d'entre elles étaient des doublons par rapport aux références déjà obtenues.

* 22 références non pertinentes, deux étaient des doublons, deux autres en langue allemande. Malgré la non pertinence, les références semblant intéressantes ont été demandées par le prêt inter bibliothèque souhaitant trouver un signalement d'un logiciel ou autres informations. La plupart n'a jamais été reçue.

Bien évidemment, les trois références obtenues par l'interrogation de la base de données PASCAL en ligne étaient toutes des doublons.

Les bibliographies citées à la fin de chaque article lu ont été une aide pour acquérir d'autres références.

Il faut signaler que les références non pertinentes concernaient : soit, simplement les logiciels documentaires, soit les logiciels qui intègrent la gestion du thesaurus en indexation et interrogation. Tandis que le sujet demandé était les logiciels autonomes de création et gestion du thesaurus.

les références pertinentes étaient en langue anglaise ; il n'était pas facile de traiter le sujet, en s'appuyant sur des informations écrites uniquement en anglais. Il a été indispensable de confier quelques articles à un bureau de traduction, afin de présenter le sujet d'une façon convenable.

I. L'HISTOIRE DU THESAURUS

On peut faire remonter l'histoire moderne du thesaurus à 1947. Même à cette première étape, une confusion de la terminologie et du but était évidente. La situation fut partiellement mise en évidence seulement à l'émergence du premier thesaurus complet et opérationnel en 1959. La période produisit de nombreuses contributions théoriques et pratiques, qui ont donné forme au concept du thesaurus d'utilisation opérationnelle¹.

1. LE DICTIONNAIRE DE MOOERS :

En février 1947, C.N.MOOERS, a fait référence à l'utilisation d'un thesaurus en association avec un système automatisé d'extraction. Dans ce but, le thesaurus était apparemment destiné à indiquer les nuances de sens associées à des termes d'index dans l'usage linguistique normal.

Un autre projet était un dictionnaire, une encyclopédie et un système d'extraction aidé par l'ordinateur, dont le but n'était pas autre que celui qu'un thesaurus destiné à être utilisé par les indexeurs et les utilisateurs.

Vers la fin de 1947, MOOERS conçut un langage développé de récupération dans lequel chaque descripteur est relié à un nombre de termes. Les descripteurs et tous les termes en relation sont listés dans un dictionnaire.

La pratique d'avoir un terme générique associé à un ensemble de termes plus ou moins relatifs reflète la forme suffisante d'un thesaurus des contemporains à attacher le nom de DICTIONNAIRE DE MOOERS.

1. Référence N° 21.

2. LES ASSOCIATIONS DE BERNIER ET CRANE :

Le crédit de ré-introduction de l'idée d'un thesaurus dans une documentation professionnelle disponible va à C.L.BERNIER et E.J.CRANE, de CHEMICAL ABSTRACTS SERVICE, en août 1947 ; ces derniers n'ont pas connaissance de l'étude de MOOERS. Ils n'ont pas non plus discuté de leurs idées avec H.P.LUHN, un autre concurrent à revendiquer l'introduction du thesaurus en recherche documentaire.

Il semble évident que les buts avancés par BERNIER et CRANE en 1947 étaient originaux dans leur application à la récupération de documents. Leurs idées étaient suffisamment nouvelles pour attirer l'attention des spécialistes de l'information et établir le terme de thesaurus dans le vocabulaire des documentalistes.

BERNIER et CRANE ont adopté une définition beaucoup plus stricte par rapport à celle habituellement employée. Ils ont proposé que des thesauri différents devraient être requis pour les différents types de questions. Contrairement aux listes d'entêtes du sujet extérieurement imposées et prescrites et communes, les thesauri proposés ont à être construits tel un processus continu à travers la compilation d'enregistrements d'ensembles de termes de recherche employés avec succès pour répondre aux questions mises au système. Chaque recherche d'un catalogue génère un ensemble ou un tableau des termes d'index caractérisés par des relations subordonnées, coordonnées ou d'équivalence. Ces ensembles de termes et de relations sont enregistrés sur des cartes perforées et fichés pour l'utilisation future. Si la même question ou une question similaire est répétée, alors la consultation du fichier révélerait l'ensemble antérieurement réussi des termes d'index. Comme une alternative des cartes perforées, les tableaux exécutés pourront être imprimés dans un livre qui ressemblerait à un thesaurus².

BERNIER et CRANE offraient une version manuelle d'un système qui attira l'attention vers les associations ; ils préconisaient une forme hybride qui tenta d'introduire le

2. L'idée des tableaux est probablement due en quelque sorte à un concept radical d'indexation associative introduite par V.BUSH.

concept associatif dans un plan conventionnel de relations conceptuelles prédéterminées. Leur but nécessitait l'enregistrement systématique des stratégies de recherche employées par des individus.

Les essais associatifs auraient impliqué une quantité énorme de travail d'entretien dans des systèmes toujours manuels. L'idée n'a pas paru pratique pour la majorité des expérimentateurs, ils n'ont pas accordé aucune importance.

3. LE THESAURUS DE LUHN :

LUHN de IBM, proclama personnellement que l'application d'un thesaurus fut introduite pour la première fois par lui-même en 1952.

La date préférée de LUHN et qui diffère de celle avancée par WALL et FOSKETT présente un problème. L'évidence suggère que, en faisant la dite réclamation, LUHN attirait l'attention vers un article publié en janvier 1953, portant le titre : "nouvelle méthode d'enregistrer et de rechercher une information".

Par incidence, ceci était sa première étude publiée. Le contenu de cet article a été soumis à l'AMERICAN CHEMICAL SOCIETY le 11.9.1951.

LUHN aurait avancé 1952 comme date marquant sa première étude à une information publique plus large, car il avait envoyé le brouillon de son exposé à l'AMERICAN DOCUMENTATION à cette date.

La nouvelle méthode de LUHN utilise le principe de caractériser un sujet par un ensemble d'éléments d'identification ou critères. Ces éléments pourront être de toutes dimensions, autant de possibilités d'enregistrement que désiré et sans importance de l'ordre dans lequel ils sont donnés.

Vers 1957, LUHN avait raffiné et développé ses idées de récupération. Il appliqua l'étiquette thesaurus à une aide à la recherche de documents qui groupait des termes assignées du codage. Ces catégories ont constitué le thesaurus. Un dictionnaire isolé alphabétique des termes et des codes a été fourni.

On a reconnu à LUHN le mérite de l'introduction d'un dictionnaire de type thesaurus .

4. LES THESAURUS DE BERNIER :

a. LE THESAURUS LIMITE :

En janvier 1957, BERNIER proposa un "thesaurus limité" pour attirer l'attention de l'utilisateur vers les termes principaux si le vocabulaire s'avère être trop large et impossible à lire à chaque fois pour la sélection. Le thesaurus limité bien que pauvre en détails de compilation, était une évolution modification évidente. Il représentait un raffinement qui jusqu'à lors manquait d'une forme agréée et qui encore avait à démontrer sa capacité d'application dans des circonstances opérationnelles.

En 1957, de même qu'en 1947, BERNIER était le pionnier de son temps.

b. LE THESAURUS TECHNIQUE :

BERNIER continua à développer l'idée du thesaurus récupératif. En juillet 1957, il introduisit le terme de "thesaurus technique", et développa un aspect du thesaurus qui était négligé par les contributeurs précédents au débat. Il a dit que, le système récupératif avait besoin d'un certain système, pour montrer les relations entre les sémantèmes ; de telle façon que l'utilisateur de l'index, puisse aller de ces sémantèmes qu'il connaît à tout les autres, qu'il aura besoin de connaître pour une recherche complète.

c. L'INDEX FIGURE :

Dans ces premiers traitements du concepté du thesaurus, BERNIER avait prêté peu d'attention à la nature des langages récupératifs. Il semble qu'il avait assumé que la forme du vocabulaire du thesaurus serait similaire à celle des listes des sujets. Vers 1957, des différences importantes ont émergé. Le listage et le regroupement des sémantèmes en forme de

thesaurus était une étape préparatoire dans la compilation d'un index opérationnel. Ce dernier index fut nommé par BERNIER index figuré. Il s'agit d'une forme d'index dans lequel chaque terme est relié à autres thèmes de plusieurs façons.

5. CONTRIBUTIONS PRATIQUES :

Vers 1957, les travaux de LUHN et de BERNIER avaient entraîné la discussion fréquente de l'idée du thesaurus. Toutefois, à l'exception de l'acceptation générale du nom, un doute considérable est demeuré à propos de l'utilisation du thesaurus. Il y avait un manque énorme d'expérience pratique.

F.R.WHALEY, de la LINDE COMPANY RESEARCH LABORATORY, NEW YORK, vers 1965, introduit un TERM FILE... constitué de fiches ordonnées alphabétiquement. Les notations sont faites sur les fiches en vue de montrer les synonymes, les termes génériques, l'étendue d'un terme est également été expliquée.

WHALEY n'était pas seul à ré-affirmer les valeurs de la recherche conventionnelle ; SCHULTZ également produisit un dictionnaire dans lequel la recherche couvre non seulement un terme spécifique, mais aussi les termes génériques ou les synonymes.

A la même époque, J.P.WADINGTON introduit un système de recherche, dans lequel les termes sont également arrangés par classe en vue d'assurer une indexation consistante, et d'indiquer les relations génériques et spécifiques si désiré.

6. LA CONTRIBUTION ANGLAISE :

S.WHELAM, un libraire, avait abouti à la conclusion qu'une classification de libraire en arbre, était UN modèle de classification non approprié aux fins documentaires. Les relations conceptuelles étaient rarement aussi simples qu'on ne les obtienne par les relations bibliographiques. La complexité des relations pourra être mieux exprimée à travers un modèle de treillis à cause de la grande liberté des relations conceptuelles impliquées par une telle forme.

WHELAM favorisa l'utilisation de vocabulaires limités. Tandis que le personnel de CLRU (CAMBRIDGE LANGUAGE RESEARCH) pense que leur utilisation a créé des difficultés d'assignation. Pour la recherche, il est préférable que le langage du thesaurus emploie les mots de la communication documentaire, car ils donnent l'indication la plus fiable du contenu des documents. Malheureusement, adhérer à un tel principe signifie inévitablement un vocabulaire volumineux avec problèmes majeurs de synonymie.

L'intérêt de CLRU n'était pas simplement théorique. A l'instar de WHELAN, ils ont testé un thesaurus sur leur propre collection.

Il a été constaté que le thesaurus ne pouvait pas accommoder facilement des termes nouveaux et de leurs relations, dès que le vocabulaire croît. La construction de treillis de relations se répandant continuellement, les coordinations fausses étaient marquées.

Mais il supportèrent peu de chose, leur travail devant beaucoup à MOOERS, LUHN et WHELAN.

Ce qui explique probablement l'influence de leurs contributions est que les membres de l'équipe CLRU étaient des experts linguistiques. Les thesauri étant des outils de langage, les contributions de ce groupe étaient investies d'une autorité à laquelle les documentalistes en tant que tels ne pouvaient aspirer.

7. THESAURUS OPERATIONNEL :

Au temps où CRANE et BERNIER et l'équipe CLRU apportaient leurs contributions au débat du thesaurus, les développements étaient déjà en chemin qui produiraient le premier thesaurus. Le thesaurus était sur le bout d'être retiré des mains des théoriciens pour être utilisé pour résoudre des problèmes pratiques au E.I. DU PONT NEMOURS AND CO., Inc., WASHINGTON, USA. Les premiers essais, en 1957, avec le système Uniterm, étaient en même temps, un succès, et des problèmes rencontrés par d'autres praticiens, tels WHALEY et SCHULTZ :

A l'achèvement de l'index du DEPARTEMENT DE GENIE il est devenu apparent que certains problèmes de vocabulaire n'ont pas été résolus : problèmes de synonyme, de génériques et d'associations.

Les solutions aux problèmes rencontrés vinrent de MORTIMER TAUBE'S DOCUMENTATION. C'est très naturellement qu'ils conseillèrent l'emploi de leur propre système Uniterm. Les essais de TAUBE pour en surmonter les déficiences, menèrent à l'introduction d'une matrice d'association des mots, qui fut générée par la création à chaque terme, d'un listage secondaire d'autres termes relatifs³. la matrice fut immédiatement marquée telle une liste de rappel qui était en effet un thesaurus.

Vers 1959, la matrice associative fut remplacée par un thesaurus technique, pour résoudre les problèmes de génériques spécifiques et sémantiques.

En style et structure, elle ressemblait à l'appareil déjà employé par WHALEY de LINDE COMPANY, vu qu'il était développé dans le but de créer à chaque terme désiré ses synonymes, ses relations génériques, ainsi qu'une liste de termes relatifs.

En fin de compte thesaurus est devenu une aide opérationnelle, bien efficace pour l'indexation et l'interrogation.

3. Un autre exemple de l'index associatif de BUSH.

II. LES LOGICIELS DE GESTION THESAURUS

Au cours des trente dernières années, le concept de logiciel a connu un changement. Traditionnellement, et même aujourd'hui, le thesaurus a été l'outil principal dans le langage d'indexation développé par des institutions individuelles pour leur propres besoins. C'est pourquoi, les fonctions principales du thesaurus sont relativement restées les mêmes depuis des dizaines d'années ; son but est d'optimiser l'indexation et l'interrogation dans un environnement donné, avec un cercle d'utilisateurs à besoins spécifiques.

Il y a cependant des facteurs qui ont affecté ce concept traditionnel, notamment depuis les dix dernières années :

- * la montée de la technique nationale et internationale et celle de la coopération scientifiques, et politique.
- * L'accroissement de communications des données.
- * L'automatisation des bureaux.
- * L'accroissement de l'importance de l'intelligence artificielle et avec elle le concept de la récupération de la connaissance.

Suite à ces changements importants les logiciels de création et de gestion de thesaurus ont commencé à apparaître. Les critères généraux minimaux d'un logiciel de thesaurus satisfaisant sont :

- * Fonctionnement Opération sur une plate-forme d'ordinateur disponible : MSDOS, UNIX...
- * Intégration avec un système de base de données existant.
- * Interface conviviale à l'utilisateur.
- * Génération automatique des réciproques.
- * Préviation des relations encombrantes.
- * Validation des termes et des relations.
- * Soutien des relations standard.
- * Disponibilité de champs supplémentaires en de texte libre.
- * Confirmation en cas de changements.
- * Une ample variété de sortie et d'états.

On peut ajouter à ces critères des fonctions spécifiques comme :

* Construction et entretien du thesaurus :

- Quelquefois, des parts de thesaurus existants doivent être intégrés.
- Les thesaurus doivent être régulièrement actualisés.

* Edition du thesaurus :

L'édition du thesaurus ou de ses parties est nécessaire à toutes les phases du travail.

* Echange des vocabulaires :

Il est toutefois donné une attention spéciale, car il gagne de plus en plus d'importance dans le futur proche.

Nous allons détailler les critères d'évaluation de logiciels de thesaurus⁴.

1. CHAMPS :

Tout logiciel de gestion de thesaurus devra aider l'utilisateur à créer une terminologie convenant à ses propres besoins. De ce point de vue, les longueurs des champs descripteurs et des champs de notes doivent être définies par l'utilisateur.

La limite minimale du champ descripteur ne devra pas être inférieure à soixante caractères cent est préférable.

En effet, quelques logiciels restreignent la longueur du champ du terme à 50 caractères ou moindre, fait qui sera un problème pour deux raisons :

4. Références N° 9 et 15.

* Plus le vocabulaire d'un thesaurus est spécifique plus les termes complexes sont probables, mais aussi les noms propres, les noms de pays par exemple, qui peuvent avoir une longueur considérable.

* L'échange des données pourra être rendue impossible, quand des mots d'un thesaurus source avec plus de caractères que ceux admis dans le fichier cible, devront être transférés.

Plusieurs systèmes limitent la longueur des notes. Il pourra arriver qu'il existe également des limites du nombre d'occurrences d'un champ à occurrences multiples, ou du nombre total des relations fixées pour un terme donné. Lorsque de telles limites existent, il sera important de vérifier si elles sont assez généreuses pour répondre au besoin.

2. LES RELATIONS :

L'intérêt d'un logiciel réside dans la spécification et la gestion des relations entre les termes. Trois zones devront être considérées :

- * Types de relations autorisées ;
- * Génération des réciproques ;
- * Contrôle des relations.

a. TYPES DE RELATIONS :

On a quatre sortes de relations :

- * Terme générique/spécifique *Broader/Narrower* - BT/NT ;
- * Utilisé/Utilisé pour *USE/UF* ;
- * Terme relatif *RT*.

On a deux variantes de la relation BT/NT :

- * BTG/NTG (générique) ;
- * BTP/NTP (partitif).

Aucun logiciel ne gère directement ces relations, mais elles peuvent être incluses dans les relations définies par l'utilisateur chez LEXICON, COLLECTION et STRIDE.

On a également l'option d'utiliser un indicateur de facettes, un terme qui n'est pas descripteur autorisé, mais utilisé en vue d'organiser les termes .

Exemple :

Automobiles

Par dimensions/Prix

Automobiles compactes
Automobiles de dimensions moyennes
Automobiles subcompactes

Par fabricants

Automobiles Chevrolet
Automobiles Nissan

Par genres

Familiale
GIT
Cabriolet

Souvent les logiciels ne gèrent pas les facettes, mais de nouveau, ceci pourra être défini par l'utilisateur dans certains des plus puissants.

Enfin, il sera possible de définir d'autres relations pour une application particulière. Ceci ne doit pas être une tâche difficile, car les relations spécialisées sont habituellement des variantes des trois types majeurs. Par exemple, il y a des applications où il sera souhaitable de spécifier des relations secondaires, une variante de *BT/NT* ou d'indiquer qu'un terme est une abréviation d'une autre variante de *USE/UF*.

b. RECIPROCITE DES RELATIONS :

La réciprocité des relations doit être automatique. L'utilisateur devra introduire un côté des relations, le système doit générer l'autre côté.

c. CONTROLES :

Les contrôles de la cohérence relative à la structure de la relation, ainsi que de la terminologie sont de première importance pour un en thesaurus. Tout logiciel devra inclure des contrôles de cohérence en vue d'éviter les caractéristiques illogiques de la terminologie et de la structure d'un thesaurus.

Les contrôles les plus importants sont ceux qui préviennent :

- * Les entrées multiples du même terme *Duplicates* ;
- * Des relations incomplètes, la relation d'un terme à un autre sans réciprocité ;
- * Le rejet des relations contradictoires entre des termes à travers plusieurs niveaux d'hierarchie.

Ces contrôles de cohérence devront sûrement être appliqués lorsque les termes sont introduits, et les relations déterminées, mais au moment où un terme où une relation sont changés ou annulés, ces changements doivent être correctement et d'une façon cohérente, reflétés dans le thesaurus aussi.

Lorsqu'une relation est mal définie, il y aura deux choix dans la réponse de système :

- * Rejet de la relation, exigeant que le terme soit ajouté si il est vraiment validé ;
- * Une réponse indiquant que le terme n'est pas actuellement dans le thesaurus et demandant s'il doit être ajouté.

La première de ces deux options est une sécurité d'intégrité, car elle prévient qu'un terme erroné ou encombrant ne s'ajoute éventuellement sans vérification. La seconde option n'est pas tellement à sécurité intégrale, mais elle est plus conviviale pour l'utilisateur. Les termes ne pourront être ajoutés accidentellement, mais il sera possible, lorsqu'on traite un terme, d'introduire un autre dans le système, pour une amélioration ultérieure.

d. PREVENTION DE L'ENCOMBREMENT :

Il est facile de produire des encombrements dans les relations lors du développement d'un thesaurus. Les logiciels préviendront de ces encombrements. On a pu éviter quatre types généraux d'encombrement :

* Un non descripteur peut n'avoir d'autres sortes de relations.

Exemple :

Agents oxydants
Oxydants USE
Réactions d'oxydation RT

* Un descripteur ne peut se référer à lui-même.

Exemple :

Agents oxydants
Agents oxydants BT

* Les hiérarchies ne sauraient être circulaires.

Exemple :

Transportation de surface
Transportation de terre NT
Transportation de terre
Transportation de surface NT

* Le même couple de termes peut être tenu à la fois à la relation BT/NT et RT l'un par rapport à l'autre.

Exemple :

Transportation de surface
Transportation de chemin de fer NT
Transportation de chemin de fer RT

Ces exemples apparaîtront rares dans le vie réelle, mais de telles erreurs peuvent avoir lieu dans le développement de thésaurus. Une addition ou un changement qui semblent être admis du côté de la relation lorsqu'ils sont introduits, pourraient encombrer. Les encombrements sont particulièrement faciles à introduire tôt à la fois dans un développement quand plusieurs relations sont ajoutées ou changées, ou plus tard, lorsqu'un terme aura un nombre de relations établi, de façon que l'oeil en oublie une en explorant la liste.

Les capacités de prévention de l'encombrement ne sont pas disponibles en général dans les logiciels qui fonctionnent sur des ordinateurs personnels.

3. MODIFICATION DES DONNEES :

La modification et l'annulation des données devront être faciles à effectuer, il est souhaitable que l'on soit capable de naviguer en utilisant les flèches et les touches PgUp et PgDn ou leurs équivalents.

4. RECHERCHE DES DONNEES D'UN THESAURUS :

Le logiciel devra assurer une recherche convenable des données du thesaurus. Cette nécessité pourra être donnée par l'un ou les deux moyens suivants :

- * Les listes permutées, hiérarchisées ou alphabétiques basées sur un terme donné.

- * La recherche par l'opérateur OU, la recherche de proximité, etc...

Cette possibilité est d'une grande importance pour les interrogations chercheurs en permettant d'examiner les termes du thesaurus et leur relations, et d'introduire des termes dans les recherches.

5. SORTIE ET IMPRESSION :

La récupération des données est nécessaire à toutes les étapes de construction et d'entretien du thesaurus et devra être possible par imprimante ou dans un fichier.

La présentation du vocabulaire suit d'habitude des standards et des conventions qui se sont développés durant plus de trente ans. La majorité des thesaurus a au moins une section alphabétique et une autre systématique, ou présentation hiérarchisée, parfois même graphique. La représentation de trois types de relation : hiérarchie, équivalence, relation associative est également devenue une trace commune de la majorité, mais certainement pas de tous les thesaurus.

a. AFFICHAGE A L'ECRAN :

Dans toute phase de construction et d'entretien, l'affichage à l'écran est nécessaire :

- * Vérifier si les termes sont déjà entrés.
- * Etablir si toutes les relations nécessaires ont été établies.
- * Vérifier si les annulations des termes ou des relations ont été correctement faites par le programme.

On aura besoin d'une grande variété de format d'affichage à l'écran et d'états. Nous considérerons quatre types généraux de présentation : alphabétique, hiérarchique, micro-thesaurus et entretien.

PRESENTATION ALPHABETIQUE :

Le listage alphabétique est fréquemment le plus complet ; il montre habituellement tous les champs associés à un terme en plus des relations. Les termes descripteurs et non descripteurs seront à la fois présents.

PRESENTATION HIERARCHIQUE :

Les hiérarchies pourront être basées sur les relations BT/NT ou sur la classification des termes en catégories par rapport aux systèmes en arbres.

Les relations BT/NT peuvent impliquer une grande étendue de duplication, car la hiérarchie entière peut être répétée à chaque terme. A cet effet, un listage hiérarchique entier de chaque terme est rarement produit tel quel état édité. Par contre les hiérarchies peuvent être intégrées avec un listage alphabétique et tronquées au même niveau. Par exemple trois niveaux supérieurs et inférieurs. Le listage d'un terme de sommet est une variante plus compacte du listage hiérarchique ; ses points d'entrée consistent uniquement en ces termes au sommet de leur hiérarchie.

Une structure en arbre est similaire à une structure hiérarchique, mais est basée sur la classification explicite des termes, même pour la classification imposée par les relations BT/NT. Elle inclut en général des informations non hiérarchiques. Les listages en catégories sont typiquement très brefs, du fait qu'ils contiennent tout juste les termes à

l'intérieur d'une catégorie ; pour plus d'informations, l'utilisateur se référera à une liste alphabétique.

MICRO-THESAURUS :

Il pourrait être utile d'extraire d'un thesaurus plus ample un sous-ensemble des termes décrivant une zone étroite particulière.

ETATS D'ENTRETIEN :

Une variété d'états statistiques sont nécessaires pour le gestionnaire : nombre de termes, relation etc..., de même que les changements avec le temps, par exemple les annulations et les changements de relations. Pareillement, les listages spécifiques des termes par une variété de critères : date de l'addition ou de la modification ; présence ou absence d'un type de relation ; états d'approbation.

b. SORTIE PAR IMPRIMANTE :

La récupération des listes de mots par imprimante est nécessaire pour plusieurs raisons, dont :

- * Les limites physiques de l'écran offrent une vue restreinte du vocabulaire en une fois.

- * Le contrôle de listes des mots sera plus facile sur matériel imprimé.

- * Toutes les institutions ne font pas usage d'un logiciel de thesaurus intégré, mais veulent utiliser un thesaurus imprimé.

Facultativement, le programme pourrait être capable de générer une représentation graphique du thesaurus. D'ailleurs, les programmes standards de thesaurus n'offrent pas cette option. Des questions plus spécifiques de l'impression sont relatives aux détails de la présentation du vocabulaire. L'utilisateur devra accorder autant de liberté possible pour qu'il soit capable d'imprimer un thesaurus conformément à ses propres besoins.

L'impression spéciale d'un thesaurus peut être librement défini en ce qui concerne la pagination, la différenciation typographique des descripteurs et des non-descripteurs, de la densité des lignes etc...

6. AUTRES DOMAINES :

a. CHANGEMENTS GLOBAUX :

Il y a des fois où il sera nécessaire de changer la même caractéristique sur plusieurs termes, la possibilité de changement global facilite la tâche.

b. APPROBATIONS :

Les procédures d'approbation de nouveaux termes et relations à la fois ont besoin d'être données. Le responsable des approbations ne saurait être celui qui fait les saisies par clavier. Même si une seule personne exécutait toutes ces opérations, il sera sage de déterminer l'impact réel d'une addition ou d'un changement avant de l'intégrer définitivement aux données.

c. ORPHELIN :

Un terme est orphelin quand il est seul sans aucune relation. Bien qu'il soit souhaitable que chaque terme soit une part d'une hiérarchie, par exemple qu'il ait au moins une BT ou une NT, ceci n'est pas toujours possible. Cependant, un RT devra être présent. Les logiciels varient du point de vue de leur traitement des orphelins. Quelques-uns ne permettent pas leur inclusion en thesaurus ; d'autres les accepteront sans commentaire.

d. CLASSIFICATION DES TERMES :

Les termes peuvent être rangés selon une classification simple ou complexe. Elle pourra prendre la forme d'une structure d'arbre très élaborée, ou quelque part entre les deux.

Si un classement est souhaité, le logiciel devra produire à la fois un champ pour le stockage du code *nom de la catégorie, nombre de classifications etc...* et la possibilité de générer des états en ordre classifié. Un listage de catégorie donne un groupement supplémentaire de termes qui pourra aider à trouver des termes utiles à ajouter à la recherche ordinairement en vue de l'élargir. Une classification détaillée pourra être utilisée pour éclater des recherches *si le système de recherche le permet* en vue de rechercher sur tous les niveaux de classifications au-dessous de celle choisie.

e. TRI :

La nécessité de trier des chiffres et des caractères spéciaux varie avec l'application. Un descripteur commence-t-il avec le chiffre "911 service d'urgence"? Si oui, comment est il arrangé sous N ou, au commencement ou à la fin de l'alphabet? Par exemple. Où trier "carbon 14" en relation avec les "composés du carbon"? Qu'en sera-t-il à propos des "composés du carbon 14"?.

Si il y a des choix de critères de tri, ce sera une bonne idée.

f. INTERFACE DE L'UTILISATEUR :

La convivialité d'un système a un grand impact sur son acceptation par l'utilisateur.

Les compromis impliquent habituellement une certaine combinaison de défilement, de zoom ou d'addition automatique de lignes.

le défilement pourra être soit vertical : les lignes se transposent au haut de l'écran et de nouvelles lignes apparaissent au fond, ou horizontal : les caractères se meuvent en dirigeant vers la gauche vu que de nouveaux caractères apparaissent sur la droite. Une caractéristique du zoom est de prolonger l'étendue d'un viseur sur base d'écran existant, tandis que l'addition automatique de lignes accroît la dimension du champ sur l'écran en vue de refléter la quantité des données qu'il contiendra. Une certaine forme de l'élargissement automatique est préférable au défilement, vu

que celui-là nécessiterait un effort moindre de la part de l'utilisateur.

g. MESSAGE D'ERREUR :

Ils doivent décrire le type d'erreur de façon la plus claire possible, et doivent être en anglais simple. Ils devront également figurer sur l'écran un temps assez long pour pouvoir être convenablement lus, de préférence jusqu'à ce qu'une frappe de touche appelle l'action suivante.

h. NOTES :

Quelques exemples de types de notes souhaitables :

* Notes d'activité : Pour tous les utilisateurs, montrent la politique d'usage d'un terme dans le système.

* Note d'indexation : Les politiques enregistrées dans l'utilisation d'un terme donné.

* Notes d'autorité : Enregistrement des sources recherchées pour un terme.

* Annulation/note de modification : Enregistrement pour tous les utilisateurs de la période durant laquelle un terme suranné est utilisé pour l'indexation, ou la date à laquelle un nouveau terme a été utilisé pour l'indexation ou la date à laquelle un nouveau terme fut introduit.

i. DATES :

Il faut, au minimum, enregistrer la date à laquelle un terme a été ajouté au thesaurus. Il est souhaitable d'ajouter d'autres dates comme celles des propositions et des approbations, ainsi que celles des plus récentes modifications, ou même celles de toutes les modifications depuis que le terme ait été ajouté. Une date d'annulation sera aussi intéressante.

j. REACTION :

Toutes les transactions d'un logiciel recevront une demande de confirmation le fait que le changement a été fait et qu'il est correct. Le système, au minimum, devra guider la confirmation que l'annulation est prévue. Le guidage pourra même aller plus loin, par exemple : Rappeler à l'utilisateur que l'annulation d'un terme annulera toutes les relations également. Si l'annulation d'un terme produit un orphelin, ceci devra être également signalé.

k. ACTUALISATION DU VOCABULAIRE :

L'actualisation du vocabulaire peut être énormément facilitée lorsque la fréquence d'utilisation des termes de thesaurus en indexation et recherche peut être vérifiée. Les statistiques à propos des conditions et des manières par lesquelles un terme a souvent été utilisé en indexation et recherche peuvent aider à identifier les échecs d'un vocabulaire en ce qui concerne sa spécificité et son étendue.

l. INTRODUCTION ET TRAITEMENT EN DIRECT PAR RAPPORT AU TRAITEMENT DE LOTS :

Aujourd'hui, les moyens populaires d'introduction sont naturellement directs. Mais l'introduction en lots des données de thesaurus est devenue de plus en plus importante, vu que la construction d'un thesaurus peut nécessiter l'adaptation de parts de thesaurus déjà existants et en plus, comme déjà mentionné, il y a un intérêt croissant à l'intégration et à la fusion de thesaurus automatiquement. Actuellement, il n'y a pas de format d'échange généralement accepté pour LES thesaurus, de façon que l'introduction d'un lots de données de thesaurus structuré, générées par un autre programme dépendent largement de la capacité du système source à créer le format nécessaire à un système exécuté.

III. LOGICIELS DISPONIBLES :

Durant les dix dernières années, le nombre de logiciels de gestion de thésaurus pour le développement, l'actualisation et l'application à la fois s'est accru rapidement en général et notamment pour les micro-ordinateurs. Les logiciels aujourd'hui disponibles peuvent être classés soit :

* Par le type de l'ordinateur pour lequel ils ont été développés :

- logiciel pour micro ordinateur, exemple : INDEX, PROTERM, CICADE, LIDOS, TMS, TCS.

- Logiciel pour les ordinateurs d'unité centrale, exemple : DOMESTIC, BASIS.

Ou

* Par les fonctions qu'ils assument : Logiciel autonome pour la construction et l'entretien des thésaurus, exemple : PROTERM, TMS, INDEX.

Plusieurs logiciels sont disponibles sur le marché, On étudiera en détail ceux qui présentent les fonctions les plus évoluées.

1. Logiciels documentaires intègrent la gestion de thesaurus :

ALEXENDRIE :

Distributeur : G.B.C.

EDIBASE :

Distributeur : VOLT Informatique et Prospective.

BRS

Distributeur : SINORG.

FLECS

Distributeur : LOGIMOT.

SUPER DOC :

Distributeur : AIDEL.

XDOC :

Distributeur : COM' X.

CAMPUS DOC :

Distributeur : IBM.

BIBLIO-TECH III, DOCUMENTA sur MACINTOSH :

Distributeur : PLATINUM S.A.

Tous sont des logiciels permettant la gestion de thésaurus.

POLYPHOT :

Distributeur : CONCEPTEURS DISTRIBUTEURS.

Les relations gérées par ce logiciel sont : Générique, spécifique, associés, employé pour, note d'application.

BABYLONE :

Distributeur : ANTHESIS.

Gestion d'un thésaurus sans limitation de nombre de termes et des relations. La modification, la suppression des termes, ainsi que l'édition du thésaurus sont possibles.

DORIS :

Distributeur : PROGIPLUS.

Gestion des relations génériques, spécifiques, synonymes et associés, ces relations sont contrôlées.
Limitation de la longueur du terme à 100 caractères.
L'édition du thésaurus.

LIBER :

Distributeur : LIBER.

40 caractères par descripteur, 5 types de relations.

la polyhiérarchie est permise, nombre de niveaux hiérarchiques non limité.

Possibilité d'affichage, pas de possibilité d'impression.

SACADO :

Distributeur : SYSECA.

24 caractères par descripteur, 2 types de relations, 16 niveaux hiérarchiques.

La polyhiérarchie est permise, pas de possibilité d'impression.

MICRO CAIRS :

Distributeur : aucun en France;

70 caractères par descripteur, 5 types de relations.

Pas de limitation de nombre de niveaux hiérarchiques. La polyhiérarchie, l'affichage et l'impression sont possible.

MINISIS :

Distributeur : DATAWARE.

80 caractères par descripteur, 5 types de relations.

Le nombre de niveaux hiérarchiques illimité. La polyhiérarchie, le contrôle de doublons, la gestion de thesaurus multiling ayant jusqu'à 9 langues, la possibilité d'affichage et d'impression.

MISTRAL :

Distributeur : CLL HONEYWELL BULL.

249 caractères par descripteur, 13 types de relations définies par l'utilisateur, 10 niveaux hiérarchiques.

La polyhiérarchie est permie.

BASIS :

Distributeur : BATTELLE.

51 caractères par descripteur, 13 types de relations.

Génération des réciproques, note d'application, contrôle de doublons, polyhiérarchie, gestion de thesaurus multiling ayant jusqu'à 3 langues, affichage et impression.

ALEXIS :

Distributeur : ST2RIALUX.

256 caractères par descripteur, nombre de relations par descripteur, nombre de niveaux hiérarchiques et nombre de langues sont illimités.

Génération des réciproques, polyhiérarchie, note d'application, contrôle de doublons, affichage et impression.

JLB DOC :

Distributeur : JLB Informatique.

Logiciel permettant de gérer 4 relations générique, spécifique, synonymie, voisinage.

Il limite le nombre de caractère par descripteur à 62 caractères, le nombre de relations par terme à 3 relations, le nombre de niveau hiérarchiques à 10 niveau.

Capable de visualiser les données alphabétiquement et d'imprimer le thésaurus⁵.

TAMIL :

Distributeur : TAMIL.

Gestion de :

- * 1 thésaurus par base.
- * 26 sections par thesaurus.
- * 9 relations : Générique/spécifique, associé, employé, employé pour, code/code pour, note historique, note explicative.
- * 100 000 termes par thésaurus.
- * 50 caractères par descripteur.
- * 500 mots vides, 10 caractères par mot vide⁶.

PAPRIKA :

Distributeur : DECALOG.

Gestion d'un thésaurus ayant toutes les relations essentielles et génération automatique des relations réciproques.

5. Référence N° 22.

6. Référence N° 30.

Le thésaurus est intégré à la recherche. Il peut être enrichi par l'utilisateur en cours d'indexation. Pour chaque mot-clé introduit, le système vérifie son existence et affiche le nombre mot.

La consultation de l'environnement sémantique d'un terme est toujours possible en cours d'indexation et de recherche⁷.

ERGOSUM :

Distributeur : ERGOSUM.

Gestion d'un thésaurus évolué, prise en compte des synonymies, des hiérarchies génériques spécifiques partitives avec un nombre illimité de descendants.
Gestion de polyhiérarchie.

Plusieurs thésaurus peuvent coexister dans une même application. Chaque champ textuel peut posséder un thésaurus qui lui est propre.

Les thésaurus sont interrogables et peuvent être édités. Pour chaque interrogation, choix de niveau de descendance dans la hiérarchie qui devra prise en compte.

Visualisation des hiérarchies sous forme graphique⁸.

BRS/SEARCH :

Distributeur : BRS SOFTWARE PRODUCTS.

Fonctionne sur OP/AT, PS/2, UNIX. Son prix est de \$2500
Version MS/DOS monoutilisateur.

Il permet la validation des termes, il soutient un nombre de relations supplémentaire prédéfinies.

Affichage alphabétique, impression des listes hiérarchisées, alphabétiques et selon KWOC⁹.

7. Référence N° 30.

Voir annexes N° 2 et 3.

8. Référence N° 30.

9. Référence N° 16.

INFORMATION NAVIGATOR :

Distributeur : INFORMATION MANAGEMENT & ENGINEERING.

Fonctionne sur OP/AT, PS/2, UNIX. Son prix est de \$4750 monutilisateur, \$5000-7000 multiutilisateur.

La validation des termes n'est pas possible c'est l'utilisateur qui vérifie l'introduction des termes.

Affichage alphabétique des termes¹⁰.

ASTUTE :

Auteur : CCE/DG XIII LUXEMBOURG 1973-1974.

Matériel : IBM Séries 360-370, Siemens Série 7700.

Langage de programmation / Cobol et PL1.

Distribution : Stérialux.

ASTUTE est gestionnaire de thésaurus, axé sur la construction de thésaurus. Il n'est pas associé à un système de recherche documentaire et fonctionne entièrement en différé.

Le système permet de gérer des thésauri multilingues ayant jusqu'à cinq langues.

Le nombre de niveaux hiérarchiques pour chaque arborescence est limité à sept.

La polyhiérarchie des descripteurs est permise.

Les relations entre deux termes de nombre de 12 maximum, 35 caractères par descripteur, chacune est reliée par un opérateur indiquant sa nature.

L'introduction d'un note d'application est possible.

Au cours des procédures de validation et de mise à jour, des tests stricts sont appliqués pour vérifier les données. Des listes d'anomalies définissant différents types d'erreurs sont produites aux différents stades du traitement.

10. Référence N° 16.

Pas de possibilité de visualiser les données mais il est possible d'éditer chaque thésaurus géré¹¹.

2.LOGICIELS AUTONOMES :

COLLECTION :

Vendeur : Vernon Systems Ltd.

Logiciel autonome, peut être intégré avec un système COLLECTION. Le thésaurus de COLLECTION est une part du système VERNON TOOLKIT, le module est disponible séparément, mais ne fut pas par ce moyen, activement vendu.

Il permet :

La validation des termes introduits par l'utilisateur,

La modification sur les données.

Les champs de date et plusieurs types de notes pour chaque terme.

La visualisation du vocabulaire est alphabétiquement,

Impression du thésaurus hiérarchique ou alphabétique¹².

STRIDE :

Vendeur : BNF Metals Technologie Center.

Logiciel autonome, peut être intégré à STATUS, son prix est £1500.

11. Référence N° 22.

12. Référence N° 16.

Traitement des données en direct, il fournit une validation des termes introduits par l'utilisateur quand il est intégré avec le système SPEED du même vendeur.

Les modifications sont très faciles à faire.

Les relations sont définies par l'utilisateur mais la validation n'est pas possible.

Un champ de note pour chaque terme est fourni, ainsi qu'un champ de date.

Impression du thésaurus hiérarchique ou alphabétique¹³.

LEXICON :

Logiciel autonome, peut être intégré à d'autres systèmes. Il a été précédemment sur unités centrales et mini ordinateur, actuellement disponible en version PS/2.

Traitement des données en direct, modification facile à faire.

Les relations sont définies par l'utilisateur.

La validation des relations est possible.

Possibilité illimitée de notes et de date définies par l'utilisateur.

Une large variété de types d'impression de thésaurus est offerte¹⁴.

TCS :

Vendeur : LIU-PALMER.

Logiciel autonome, il peut être intégré à d'autres systèmes. Le prix de BASIC VERSION est \$150, le prix de PROFESSIONEL VERSION est \$450. La dernière version est encore au test.

Traitement des données en direct.

La gestion de relations standards est possible ainsi que les facettes.

La validation de relation est soutenue par TCS.

L'introduction de notes et de dates est possible.

La modification de données est limitée.

La visualisation du vocabulaire est hiérarchisée, des affichages plus flexibles seront une caractéristique de la version professionnelle de TCS.

13. Référence N° 16.

14. Référence N° 16.

L'impression d'un thésaurus hiérarchique, alphabétique ou selon KWOC (Key Word On Connection).

J'ai manipulé quelques fonctions de ce logiciel en utilisant la version de base qui m'a été prêtée par le demandeur.

De nouvelles versions de LEXICO et de TCS avec plusieurs caractéristique nouvelles sont encore au stade d'expérimentation ; il est raisonnable d'attendre qu'elles arrivent sur le marché¹⁵.

TMS :

TMS Version 1.0, est un logiciel anglais distribué par MAINTENANCE SYSTEM, développé par PYRAMID IN READING.

Il est autonome ou part de CALM, un thésaurus développé par TMS pourra être transféré dans le système CALM qui servira de liste d'autorité pour valider les termes à l'introduction des données.

Le logiciel est accompagné d'une INTRODUCTION D'UTILISATION en langue anglaise. Son prix est 1.480DM.

Le matériel demandé :

- * PC-XT ou AT.
- * Disque dur de 512 KB de mémoire.
- * Système d'exploitation MS-DOS 2.1 ou plus.

Le menu principal se compose de deux parties : CONSTRUCTION et REVIEW. La partie CONSTRUCTION comporte les options de l'introduction des termes, les relations et les scopes notes, l'option de changer et supprimer les termes, l'option d'ajouter des facettes et l'option de Re-indexation. La deuxième partie comporte les fonctions de sortie.

L'introduction des termes se fait en tapant le terme ou son code.

Le nombre de relations par terme n'est pas limité, 5 type de relations sont gérés par TMS.

15. Référence N° 16.
Voir annexes N° 4 à 11.

Le logiciel est capable de reverser les relations d'un terme et les assigner automatiquement au deuxième terme après la sauvegarde.

Il a des grandes capacités de classification des termes.

Il est capable de réorganiser tous les enregistrements, et de produire des statistiques du vocabulaire, ainsi que certains type d'impression de thésaurus qui figurent dans la liste à la fin de ce chapitre.

TMS a des points faibles comme :

Les données sont limitées à 65.535 termes.

Les modifications sont limitées.

La longueur du terme est limitée à 40 caractères, les scopes notes et les facettes à 20 caractères.

Le contrôle de relations n'est pas assez stricte.

L'impression de thésaurus à facettes n'est pas possible.

La gestion du thésaurus multilingue est limitée à deux langues¹⁶.

PROTERM :

Il a été logiciel spécialement développé pour la constitution, la mise à jour et l'édition de thesauri ou de vocabulaires sur compatible PC.

Il existe trois versions de PROTERM :

PROTERM-T pour le développement et la maintenance des thesauri, des vedettes-matières ou des systèmes de classification.

PROTERM-K comprend PROTERM-T et permet la communication avec des bases de données extérieures via un vocabulaire individuel de recherche.

PROTERM-G sert au développement de glossaires ou terminologies et supporte l'organisation du travail scientifique.

16. Références N° 16 et 20.
Voir annexes N° 12 et 14.

PROTERM-T :

PROTERM-T Version 2.5, logiciel développé par PROGRIS à BERLIN .

Il existe 5 modules supplémentaires.
Il est accompagné d'un MANUEL D'UTILISATION en langue allemande de bon niveau. Son prix est 1200DM.

Le matériel demandé :

- * PC-XT ou AT.
- * Disque dur de 512 KB de mémoire.
- * Système d'exploitation MS-DOS 2.1 ou plus.

Le menu principal de PROTERM-T est composé de trois parties l'INTRODUCTION ET SUPPRESSION, SORTIE, GESTION.

Le logiciel PROTERM limite la longueur de terme à 60 caractères, les scopes notes à 150 caractères et les champs de la date de création et de suppression d'un terme à 50 caractères.

Il n'offre que 5 relations différentes, PROGRIS fournit un module supplémentaire pour des autres relations. Le contrôle des relations est assuré.

Une variété de possibilités de recherches, corrections, suppressions des termes est offerte par PROTERM grâce à la possibilité de liste et formulaire.

L'utilisation est facile grâce aux menus d'aide.

Le changement de l'ordre d'un enregistrement, la réorganisation de tous les enregistrements, l'import et l'export de données sont possibles.

La gestion de plus d'un thésaurus, la production des statistiques sur le vocabulaire.

Des possibilités sophistiquées au niveau de la sortie¹⁷.

17. Référence N° 20.
Voir annexes N° 12 et 14 et 15.

INDEX :

INDEX Version 4.1, est un produit de ERNEST LUKAS.

Il existe 8 modules supplémentaires.
Il est accompagné d'un MANUEL D'UTILISATION en langue allemande de très bon niveau. Son prix est 6.200DM.

Le matériel demandé :

- * PC-XT ou AT.
- * Disque dur 512 KB de mémoire.
- * ENHANCED GRAPHICS ADAPTER pour des sortie spéciales.
- * Système d'exploitation MS-DOS 3.3, ou autres ressemble à PC-DOS, UNIX/XENIX, OS2 et autres.

Le menu principal du logiciel INDEX se compose de 7 options. Le menu principal est coloré en bleu, la première linge en gris, tous les couleurs peuvent être changées par l'utilisateur.

INDEX permet l'utilisation des touches de fonction F1 à F10. En plus, un menu apparait en haut de l'écran expliquant les touches de fonction. Pour chacune un sous menu indique son équivalent. A certaines étapes, un menu apparaît en bas de l'écran donnant des explications supplémentaires.

Il gère 16 petits fichiers associés à chaque terme pour y spécifier des informations, par exemple : Code de langue, code de sujet, les facettes. Tous ces fichiers sont indexés. PROTERM est capable d'assumer cette fonction dans un module additionnel.

Il gère les homonymes dans un fichier spécial.

Il ne limite pas la longueur des scopes notes, l'introduction d'une scope note consiste à l'ouverture d'un fichier texte ou on peut ajouter ou supprimer une ligne, on a la possibilité de rechercher un mot, une phrase ou plus. L'utilisateur peut définir le genre de la note (définitions, notes historique etc...).

Il offre à l'utilisateur la possibilité de choisir entre 24 sortes de relations.

La réorganisation automatique des relations après un changement d'un terme descripteur à non-descripteur fonctionne parfaitement et confortablement avec INDEX, alors que les autres logiciels ne peuvent pas le faire.

L'utilisation de certaines couleurs facilitent la lecture et la compréhension du fonctionnement du logiciel.

Les menus d'aide, les messages d'erreur, les messages du système facilitent la manipulation du logiciel.

L'import et l'export de données, le fonctionnement en réseau, l'utilisation par plus d'un utilisateur, la gestion de mots de passe, la gestion de plus d'un thésaurus multilingue, l'organisation d'un/des enregistrement(s), les statistiques sur le vocabulaire, les possibilités diverses au niveau de produits de sortie, le changement de paramètres d'impression, sont des caractéristiques d'INDEX.

Enfin, il n'accepte pas la date de création et suppression d'un terme.

Malgré ce fait, la gestion de 24 relations strictement contrôlées donne à INDEX une valeur inégalée¹⁸.

18. Référence N° 20.

Voir annexes N° 12 et 13 et 14 et 15.

CONCLUSION :

Les progrès de la technologie des ordinateurs qui conduisent mènent à des prix décroissants, à une manipulation plus facile et en même temps à des performances meilleures : ont énormément facilité la tâche fastidieuse de développement et d'utilisation des thésaurus.

On a vu que chaque logiciel a ses points forts, la plupart répondent aux besoins fondamentaux de logiciel de thesaurus. Mais dans plusieurs logiciels quelques besoins fondamentaux ne sont pas satisfaits :

** le nombre de niveaux hiérarchiques*

* Le nombre maximum des caractères est limité à 40 ou même moins.

* La relation associative est soit complètement oubliée ou ne peut être définie par l'utilisateur.

* Les champs des notes, les champs d'étendu d'un terme.

* Les contrôles de la cohérence relative au vocabulaire et aux relations sont soit complètement oubliés ou incomplets seulement.

* La présentation de l'exposition systématique soit sur l'écran ou sur papier n'est pas possible.

Cependant, même le logiciel de thesaurus qui satisfait tous les besoins fondamentaux ne sera pas suffisant dans certains cas.

Finalement, c'est l'utilisateur final qui doit décider lequel des logiciels convient le mieux à ses besoins.

BIBLIOGRAPHIE

1. BURKART M. PROTERM ein softwarepaket fuer aufbau pflege, handling von thesauri und anderen wortgutsammlungen. **Nachrichten fuer Dokumentation**, 1988, vol 39, n 4, ISSN 0027-7436, p 249-252.
2. CHARTRON G., DALBIN S., MONTEIL M.G., VERILLON M. Indexation manuelle et indexation automatique : Dépasser les oppositions. **Documentaliste**, 1989, vol 26, n 4-5, ISSN 0012-4508, p 181-187.
3. CHAUMIER J., FOURTEAU P. Le traitement des thésaurus à schémas fléchés par l'informatique graphique interactive. **Documentaliste**, 1979, vol 16, n 1, ISSN 0012-4508, p 9-14.
4. CREPOUX E. Les logiciels documentaires MISTRAL et STAIRS dans la perspective du texte intégral. **Documentaliste**, 1982, vol 19, n 3, ISSN 0012-4508, p 87-91.
5. DEWEZE A. HYPERDOC : Base de données relationnelles pour application documentaire. **Documentaliste**, 1987, vol 24, n 4-5, ISSN 0012-4508, p 190-191.
6. DEWEZE A. OMNIDOC : Utilisation d'un logiciel polyvalent de gestion multifichiers pour le traitement de l'ensemble des fonctions documentaires. **Documentaliste**, 1986, vol 23, n 3, ISSN 0012-4508, p 109-116.
7. DORE D. Logiciel documentaire BESTSLER. **Archimag**, 1987, n 14, ISSN 531-014, p 48-52.
8. FUGMAN R. An interactive classaurus on the PC. **International Classification**, 1990, vol 17, n 3-4, ISSN 0340-0050, p 133-137.

9. GANZMANN J. Criteria for the evaluation of thesaurus software. *International Classification*, 1990, vol 17, n 3-4, ISSN 0340-0050, p 148-154.

10. ILJON A. Le logiciel ASTUTE : Un outil informatique pour la création, la mise à jour, l'édition et l'impression de thésauri mono et multilingues. *Documentaliste*, 1977, vol 14, n 5-6, ISSN 0012-4508, p 25-33.

11. KLEINBART P. Prolegomenon to intelligent thesaurus software. *Journal of Information Science, Principales & Practice*, 1985, vol 11, n 2, ISSN 0165-5515, p 45-53.

12. KIBBEY M. MARCON II : Information control on a micro computer. *Library Hi Tech*, 1986, vol 4, n 2, ISSN 0737-8831, p 11-21.

13. MARCHALOT M. A la recherche d'un outil d'interrogation d'un système de documentation automatique, une expérience de thésaurus d'interrogation. *Documentaliste*, 1982, vol 19, n 6, ISSN 0012-4508, p 203-206.

14. MCLELLAND J. Computers, data bases and thesauri. *Aslib Proceedings*, 1990, vol 42, n 7-8, ISSN 0001-253X, p 201-205.

15. MILSTEAD J. Specifications for thesaurus software. *Information Processing & Management*, 1991, vol 27, n 2-3, ISSN 0306-4573, p 165- 175.

16. MILSTEAD J. Thesaurus software packages for personal computers. *Data Base*, 1990, vol 13, n 6, ISSN 0162-4105, p 61-65.

17. MOUREAU M., DELAUNAY J. Principe et développement d'un thésaurus. Exemple d'application : Le thésaurus pétrol. *Bulletin des Bibliothèques de France*, 1973, vol 18, n 1, p 5-29.

18. RADA R. Maintening thesauri and metathesauri. **International Classification**, 1990, vol 17, n 3-4, ISSN 0340-0050, p 158-164.
19. REVERSE A. Le logiciel MISTRAL 3/4. **Documentaliste**, 1979, vol 16, n 1, ISSN 0012-4508, p 15-18.
20. RITZLER C. Comparative study of PC-supported thesaurus software. **International Classification**, 1990, vol 17, n 3-4, ISSN 0340-0050, p 138-147.
21. ROBERTS N. The pre-history of the information retrieval thesaurus. **Journal of Documentation**, 1984 , vol 40, n 4, p 271-285.
22. ROHOU C. La gestion automatisée des thésaurus. Etude comarative de logiciels. **Documentaliste**, 1987, vol 24, n 3, ISSN 0012-4508, p 103-108.
23. ROUX-FOUILLET J. Informatique documentaire : Comment et pourquoi. **Archimag**, 1990, n 35, ISSN 531-014, p 30-37.
24. SCHMITZ-ESSER W. Thesauri facing new challenges. **Intrenational Classification**, 1990, vol 17, n 3-4, ISSN 0340-0050, p 129-132.
25. SLATTA R. Memory-resident thesaurus (software package). **Business Software**, 1986, vol 4, n 6, ISSN 0742-1214, p 76-77.

26. BACON N. Etude de l'évolution de thésaurus du BRGM et début d'adaptation au logiciel MISTRAL. Mémoire de stage, Université Claude Bernard, Lyon 1, 1979.

27. LAUREILHE M.T. Le thesaurus, son rôle, sa structure, son élaboration. Lyon : Presse de l'ENSB, 1977, 48 p.

28. MILSTEAD J. Thesaurus software packages. In Information in the Year 2000 : From Research Cooperation, 1990, p 3-15.

29. SABRI KAMEL E. Etude de faisabilité d'un thésaurus pluridisciplinaire. Paris : Mémoire de stage, ENSB, 1977.

30. Collection CXP Catalogues de progiciels, gestion électronique de documents, gestion documentaire, 1991.

31. Logiciel de gestion de Bibliothèque.

LISTE DE CONTROLE DE LOGICIEL DE THESAURUS :

Il est bien connu que les capacités des logiciels varient de l'un à l'autre, de ce point de vue, il est important de définir les besoins d'une façon assez précise pour déterminer quel logiciel peut y répondre. Les besoins qu'il faut définir sont :

- * Les types de relations spéciales.
- * Les champs de notes.
- * Les dates.
- * Les procédures d'approbation d'un terme.
- * Les formats d'affichage et les produits de sortie.

Il faut prendre en considération que les besoins varient avec le temps ; voilà quelques exemples de changement :

- * Les listings nouveaux et spéciaux.
- * Les procédures d'approbation d'un terme.
- * Les utilisateurs autorisés et non autorisés à l'introduction des données.

Les critères pour choisir un logiciel de thésaurus sont¹⁹ :

A. Critères généraux :

...1. Données techniques

.....Compatibilité du matériel
.....Ordinateurs sur lesquels le logiciel fonctionne
.....Sauvegarde de données
.....Système d'exploitation

19. Référence N° 9.

-Un seul utilisateur
-Plusieurs utilisateurs

-Logiciel
-Langage de programmation
-Un utilisateur
-Plusieurs utilisateurs

...2. Données de développement

-Développeur

-Version
-Version récente
-Première version
-Nombre total des versions

...3. Prix

...4. Service après-vente

-Soutien d'institutions

-Formes de soutien
-Téléphone de ligne directe
-Consultation
-Entraînement
-Lettres nouvelles
-Soutien actif
-Installation
-Actualisation
-Modification

...5. Documentation

-Types de manuels
-Manuels d'Opérations
-Manuels d'Utilisateurs

-Parts inclus
-Documentations sur :
-Les spécifications techniques
-L'installation
-L'application
-Les messages d'erreurs
-Sauvegarde
-Table de Matières

-Index
-Structure de manuel
-Achèvement de l'information
-Correction de l'information
-Clarté
-Exemples
-Version d'entraînement
-Travaux dirigés

...6. Ergonomies du logiciel

-Formes de dialogues
-Commandes
-Menu
-Hybride
-Souris

-Fonctions d'aide

-Messages
 -Auto-explicatif
 -Expliqué dans ce manuel
 -Messages d'erreurs
 -Messages de réaction
 -Messages d'alerte
 -Messages de confirmation

-Contrôle d'accès
 -Mot de passe
 -Restrictions pour utilisateurs individuels
 -Restrictions pour données spécifiques
 -Restrictions pour fonctions spécifiques
 -Procédures de sauvegarde
 -Automatique
 -Appareil de sauvegarde
 -Caractéristiques de réorganisation

B. Critères relatifs aux fonctions des logiciels

...1. Définitions structurales

.....1.1. Terme et attributs relatifs du terme

-Champ prédéfinis du :
 -Terme
 -Nombre maximal de caractères
 -Etendue Note/Texte
 -Nombre maximal de caractères
 -Notation
 -Nombre maximal de caractères
 -Source du terme
 -Nombre maximal de caractères
 -Longueur variable
 -Information du langage du terme
 -Nombre maximal de caractères
 -Champs supplémentaires
 -Nombre maximal de caractères

.....Définitions de l'utilisateur

-Nombre de champs
-Longueur des Champs
-Séquence des champs

.....1.2. Relations

.....1.2.1. Entre des termes d'un seul vocabulaire

-Définitions des relations
-Relations prédéfinies
-Relations définissables par l'utilisateur

.....Nombre de relations prédéfinies

-Types de relations
-Relations d'équivalence
-Synonymie normale (Non-descripteur-->descripteur)
-Factorisations sémantiques (ditto)
-Alternatives (Non-descripteur-->descripteurs alternatifs)
-Relations hiérarchiques
-Pas de différenciation
-Différenciations de la relation partitive et générique
-Définitions des principes de division (catégories)
-Relations associatives

-Pas de différenciation
-Différenciation de divers types
-(prédécesseurs-successeurs,
-relations d'appartenance)

-Nombre de relations entre termes individuels
-Relations d'équivalence
-Synonymie normale
-Factorisation sémantique
-Alternative
-Relations hiérarchique.
-Nombre de termes inférieurs par
-terme élargie
-Nombre de termes élargis par terme inférieur
-(Polyhiérarchie)
-Nombre de niveaux hiérarchiques
-Relation associative

-1.2.2. Entre des termes de différents vocabulaires
-Type de vocabulaires
-Thesauri multilingues
-Vocabulaire compatibles

-Connection entre différents langage naturels
-Nombre maximal de langues différents
-Statut du/des langage(s) individuel(s)
-Langages égaux
-Dominance d'un langage

-Connection entre différents langage d'indexation
-Nombre maximal de langage d'indexation
-Type de langage d'indexation
-Classifications
-Thesauri

...2. Introduction

.....2.1. Saisie des données

-Mode de saisie
-Introduction en lots à partir d'autres systèmes
-Clavier
-Mode d'introduction des termes et des attributs
-Mode d'introduction des relations

-Facilité des saisie
-Complexité d'introduction des termes
-et des relations
-Pas séparés
-Séquence fixée des routines d'introduction

-Affichage des termes introduits et des relations
-sur écran

.....Dérivation automatique des relations implicites

.....2.2. Modification

.....Mode de modification

.....Changements globaux possibles (Des codes
..... de langages)

.....Facilité de modification

.....Complexité de modification

.....Facilité des changements affectant le statut
.....des termes (descripteurs non descripteurs)

.....Affichage sur écran des termes et des relations

.....2.3. Annulation

.....Mode d'annulation

.....Annulations globales des termes et des relations

.....Facilité d'annulation

.....Complexité de l'annulation

.....Annulation automatique des relations d'un terme
.....annulé

.....2.4. Contrôle de la cohérence

.....Définitions

.....Prédéfinie

.....Définie par l'utilisateur

.....Terme et attributs du terme

.....Rejet des entrées multiples du même terme

.....Relations

.....Contrôle de la réciprocité des relations

.....Rejet de plus d'un type de relation entre 2 termes

.....Rejet des relations incomplètes

.....Rejet d'une relation double entre deux termes

.....Rejet de la relation hiérarchique ou associative
.....entre descripteur et non-descripteur

.....Contrôle des relations illogiques à travers
.....les niveaux hiérarchiques

.....Autres contrôles

...3. Récupération

.....3.1. Affichage sur écran

.....Mode de recherche d'un terme

.....Balyage

.....Défilement

.....Autres possibilités

-Affichage d'un terme individuel
-Avec attributions
-Avec relations

-Affichage des listes de mot
-Critères pour la sélection des termes
-Section alphabétique
-Chaînes
-Attributs (langage, notation, source)
-Types de relations
-Texte marqué pour des buts spécifiques
-Combinaison des critères

-Formats d'affichage des listes de mots
-Tableau alphabétique
-Liste de mots
-Liste de mots plus relations et attributs
-Autres variations
-Affichage KWIC
-Affichage hiérarchique
-Présentation systématique (Sortie par notation)
-Système détaillé
-Sans référence aux relations
-Avec référence aux relations
-Catégories élargies (groupes du sujet/
-facettes)
-Affichage graphique

-Interaction possible en thesauri sur écran
-Défilement /Balayage
-Navigation des termes relatifs sémantiquement
-Sélection des termes d'édition et d'annulation
-Modifications directes et annulations en listes

-3.2. Sortie vers l'imprimante

-Définitions des formats d'extraction
-Format standard prédéfinis
-Format définissables par l'utilisateur
-Sauvegarde des format définis par l'utilisateur

-Critères de la sélection des termes
-Section alphabétique
-Chaînes
-Attributs (notation, facette)
-Type de relation
-Combinaison des critères

-Formats d'impression
-Tableau alphabétique
-Sans informations ultérieure
-Avec des relations
-Avec des attributs
-Index KWOC
-Index hiérarchique
-Sans relations

-Avec relations
-Présentation systématique (Sortie par notation)
-Système détaillé
-Sans relations
-Avec relations
-Sans attributs
-Avec labels de noeud
-Catégories élargies (groupes du sujet/
.....facettes)
-Affichage graphique
-Exposition en colonnes des vocabulaires
.....multilingues/compatible
-Aspects définissables par l'utilisateur
-Information ajoutée aux termes
-Relations
-Attributs
-Présentation des relations
-Suppression de certaines relations
-Séquence des relation en imprimerie
-Codes de référence définissables par
.....l'utilisateur pour l'extraction
-Décalage
-Pagination
-Densité de la ligne
-Sous-titre
-Différenciation typographique des descripteurs/
.....non descripteurs
-Autres aspects

...4. Indexation

-Orientation
-Formes d'affichage de thesaurus sur écran
-alphabétique
-systématique
-Autres formes d'exposition
-Mode de recherche de termes
-Navigation à travers la structure sémantique
-Mode d'introduction
-Entrée des termes
-Sélection directe des termes du thesaurus d'écran
-Contrôle de l'introduction
-Rejet des termes inconnus
-Définition par l'utilisateur de l'utilisation
.....des termes candidats
-Remplacement des termes de thesaurus non admis
.....en indexation
-Remplacement des termes composés par des facteurs
.....sémantiques
-Remplacement des non-descripteurs par des
.....descripteurs (pour les thesaurus à termes préférés)

-Remplacement des termes en langages secondaires par
.....des termes du langage dominant
.....en thesaurus multilingues

-Représentation des concepts
 -Terme préféré
 -Terme non-préféré

-Actualisation
 -Changements globaux en index
 -Statistique à propos de l'utilisation des descripteurs

...5. Recherche

-Orientation
 -Formes d'affichage du thesaurus sur écran
 -Exposition alphabétique
 -Exposition systématique
 -Autres formes d'expositions
 -Mode de recherche des termes
 -Navigation à travers la structure sémantique

-Mode d'introduction
 -Entrée des termes
 -Sélection direct des termes

-Contrôle de l'introduction
 -Rejet des termes inconnus
 -Remplacement des termes de thesauri non admis
.....à la présentation des concepts :
 -Remplacement des termes composés par des
.....facteurs sémantiques
 -Remplacement des non-descripteurs par
.....descripteurs
 -Remplacement des termes d'un langage secondaire
.....par des termes du langage dominant dans
.....un thesaurus multilingue
 -Inclusion automatique des synonymes

-Formulation des stratégies des recherches
 -Option de recherche générique et automatique
 -Recherche automatique des termes relatifs
 -Inclusion automatique des prédécesseurs du terme
.....de recherche

-Actualisation
 -Statistiques à propos des termes de recherche

PAPRIKA

CONSTITUTION DU THESAURUS

Après saisie du mot, le système vérifie que celui-ci n'existe pas déjà au thésaurus.

4 utilisations possibles de cet écran en modifiant le caractère dans le cadre mode: création <C>, modification <M>, annulation <A>, visualisation <V>.

Paprika v 4.2		Saisie des mots du thésaurus		15/09/89	
Nombre de mots du thésaurus : 16.233				Mode : C	
Code	Terme générique	Nb util.	Prof.		
98	INFORMATION	0	2		
Code	Terme	Nb util.	Niv.		
1212	AGENCE DE PRESSE	0	2		
Type : A		Renvois		Mode : C	
Code	Terme associé	Nb de renvois			
1502	COMMUNICATION	2			
Sortie :			Validation		

Le nouveau terme créé peut faire l'objet d'un renvoi définitif (voir) ou de plusieurs renvois d'orientation (voir aussi). Les termes objet du renvoi sont alors considérés comme des termes associés.

PAPRIKA

CATALOGAGE

5ème ETAPE : Indexation Matière

A chaque mot-clé, le système vérifie son existence au thésaurus et affiche le nombre d'utilisations du mot.

Paprika v 4.2	Indexation matière	15/09/89
Document		
Code type: 1 LIVRES		Indice: 346.036
No de notice: I4058		
Titre : Nid du coucou (Le)		
Auteur: STOLL	Clifford	

N° d'indexation matière : 1	Mode : C
-----------------------------	----------

Code	N° d'indexation matière	N.U.
10102	1° mot : PIRATAGE INFORMATIQUE	3
103	2° mot : ARMEE	28
7111	3° mot : UNIVERSITE	16
808	4° mot : U.S.A.	154
15112	5° mot : ESPIONNAGE	12
	6° mot :	
	7° mot :	
	8° mot :	

Mot : SERVICES SECRETS

Ce terme ne figure pas au thésaurus. Voulez vous le créer ? 0

Sortie :	Validation
----------	------------

Lorsque le terme saisi ne figure pas au thésaurus, le système affiche une liste constituée des termes commençant par les mêmes lettres et de tous les termes contenant un des mots entrant dans la composition du terme. Si aucun des termes proposés ne convient, l'opérateur peut demander à passer en création du mot. Il devra positionner le nouveau mot dans la hiérarchie avant de poursuivre l'indexation.

L i u - P a l m e r T h e s a u r u s C o n s t r u c t i o n S y s t e m
HIERARCHY DISPLAY (IN) DEPTH: 2

INFOR
LOG

HIERARCHY LIST HELP WINDOW

- | | | |
|-----|---|----------------------------------|
| L | <F1> - Help | <F2> - Select new hierarchy |
| P | <F3> - Select new Current Term | <F4> - RT display |
| S | <F5> - Sibling display | <F6> - UF display |
| U | <F7> - Thesaurus display | <F8> - Facet indicator display |
| MAT | <F9> - Hierarchy display | <F10> - System Menu |
| M | | |
| P | <^A> - Add a hierarchy | <^F> - Add FI to Current Term |
| U | <^N> - Add NT to Current Term | <^R> - Add RT to Current Term |
| | <^U> - Add UF to Current Term | |
| | <^D> - Delete a term | <^E> - Edit a term or scope note |
| | <^T> - Transfer term [and children] | |
| | Move cursor to the term whose BT is to be changed and
press <^T>. Then select the new BT and press <Enter> | |
| | <^PgUp> - Increase Hierarchy Depth Limit by one | |
| | <^PgDn> - Decrease Hierarchy Depth Limit by one | |

Press <Esc> to Exit Help Window

Annexe 4

F1 aide

INFORMATIQUE
LOGICIEL
LANGAGE
PROGICIEL...
SYSTEME D'EXPLOITATION
UTILITAIRE
MATRIEL
MEMOIRE...
PERIPHERIQUE
UNITE CENTRALE

Hierarchy List

ANALYSE DOCUMENTAIRE	AN
INFORMATIQUE	IN

PRESS <F1> FOR HELP

Annexe 5

F2

INFORMATIQUE
LOGICIEL
LANGAGE
PROGICIEL...
SYSTEME D'EXPLOITATION
UTILITAIRE
MATRIEL
MEMOIRE...
PERIPHERIQUE
UNITE CENTRALE

Current Term Selection Window

CURRENT TERM:
PRESS <Esc> TO EXIT.

PRESS <F1> FOR HELP

Annexe 6

F3

INFORMATIQUE
LOGICIEL
LANGAGE
PROGICIEL...
SYSTEME D'EXPLOITATION
UTILITAIRE
MATERIEL
MEMOIRE...
PERIPHERIQUE
UNITE CENTRALE

SI

SIBLING LIST HELP WINDOW

<Enter> - Make a sibling Current Term

<Esc> - Return to Hierarchy List

Press <Esc> to Exit Help Window

INFORMATIQUE
*LOGICIEL
MATERIEL

PRESS <F1> FOR HELP

Annexe 7

F5

INFORMATIQUE
LOGICIEL
LANGAGE
PROGICIEL...
SYSTEME D'EXPLOITATION
UTILITAIRE
MATERIEL
MEMOIRE...
PERIPHERIQUE
UNITE CENTRALE

THESAURUS DISPLAY WINDOW

Current Term: PROGICIEL
SN: Tout programme applicatif
BT: LOGICIEL
NT: PROGRAMME

PRESS <F1> FOR HELP

Annexe 8

F7

INFORMATIQUE
MATERIEL
 MEMOIRE
 MEMOIRE CENTRALE
 MEMOIRE DE MASSE
 TECHNOLOGIE MAGNETIQUE
 TECHNOLOGIE OPTIQUE

PRESS <F1> FOR HELP

Annexe 9

f9

INFORMATIQUE
LOGICIEL
 LANGAGE
 PROGICIEL...

System Menu

Maintenance Submenu

- C) CHANGE HIERARCHY
- F) Pack Files & Rebuild Indexes
- B) Backup Current Hierarchy
- R) Restore Selected Hierarchy
- X) Destroy Current Hierarchy

Press <Esc> to Exit this Window.

Annexe 10

f10

INFORMATIQUE
 LOGICIEL
 LANGAGE
 PROGICIEL...
 SYSTEME D'EXPLOITATION
 UTILITAIRE
MATERIEL
 MEMOIRE...
 PERIPHERIQUE
 UNITE CENTRALE

Message Window

There is no related term for the Current Term.
Press any key to continue...

Annexe 14

message d'erreur.

Thesaurus development

- | Construction | Review |
|--|---------------------------------|
| 10. Terms, relationships and scope notes | 50. View thesaurus |
| 11. Change or delete terms | |
| 12. Facets | |
| 13. Language alternatives | 60. Term list |
| 14. Indentation levels | |
| | 62. Single level term list |
| 20. Re - index | 63. Single term systematic list |
| | 64. Alphabetic systematic list |
| | 65. Alphabetic NT1, NT2 list |
| | 70. Create language index |
| | 80. Change print parameters |

Please enter your choice :

Fig. 1: The main Menu of TMS

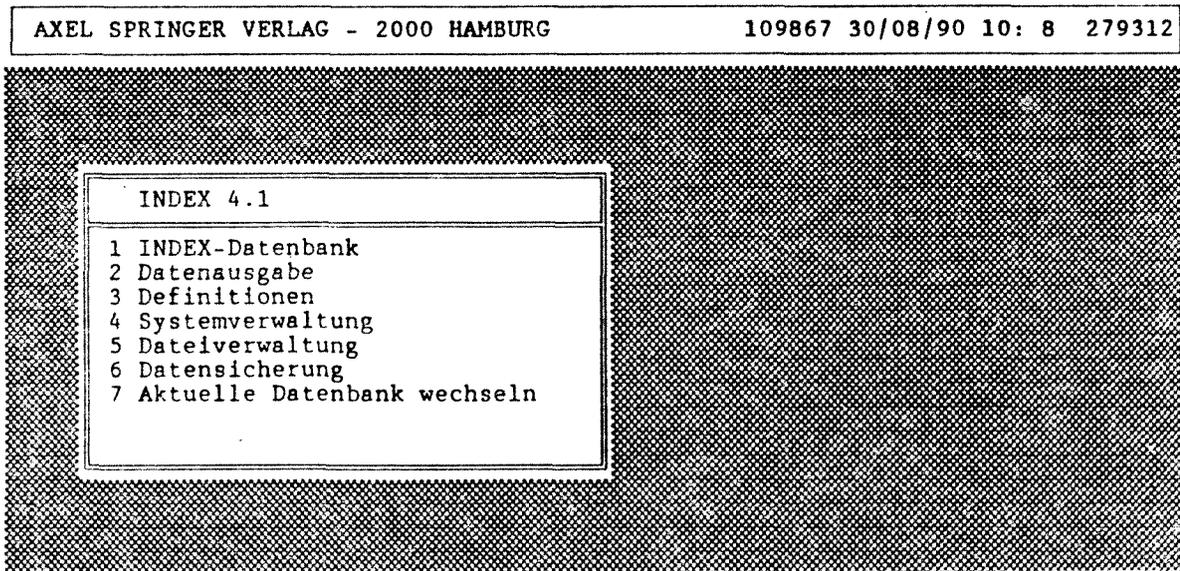
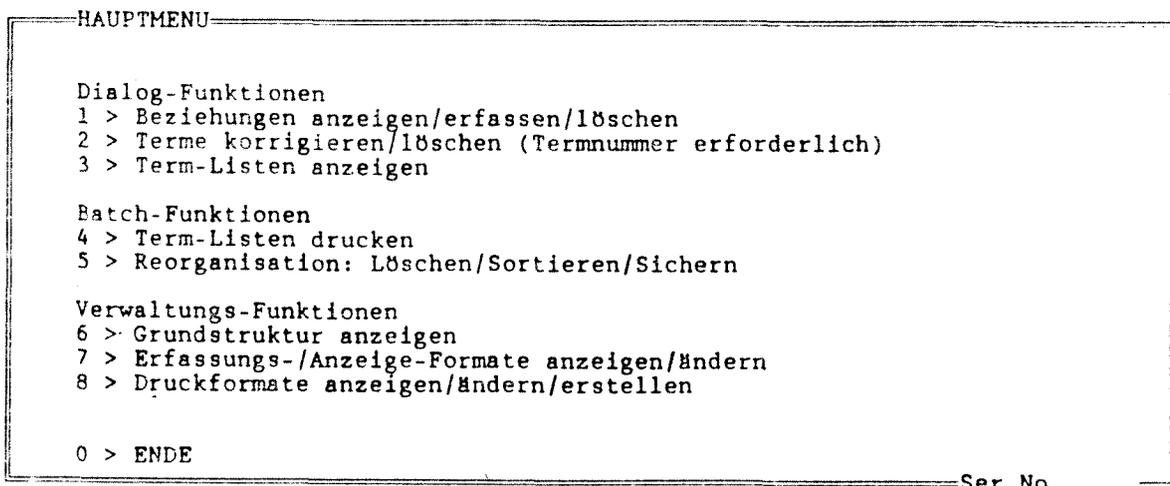


Fig. 2: The main menu of INDEX

PROTERM-T Vers. 2.4 (c) PROGRIS 1987, 1988

Dienstag, 07.08.90



Mit Cursor auswählen, dann <ENTER> oder Ziffer wählen

Fig. 3: The main Menu of PROTERM

Annex-12

les menus principaux

Wörter (G) alphabetisch	M	Sgr	Beziehungsarten					
Öffentliches Verkehrsmittel			1 Synonymie					
Öl			1.1 Abkürzung					
Öltanker		WAS	1.2 Identische Verwendung					
Paddel		WAS	1.3 Quasisynonymie					
Paddelboot		WAS	1.4 Alternative					
Pannent			ion					
Parkh	Deskriptor		g e Hierarchie e Hierarchie at gkeit ehung haft olge					
Parkp	Nr: 482 H: 0 S: M:							
Parks	Öffentliches Verkehrsmittel							
Parku	Sp	Sgr Fa		Wo G Gm Sch	1 2 3	Z1 Z2	F	T
Parkv		G						
Passa	Nichtdeskriptor			1 Synonymie	Asp:			
Passa	Kraftwagen							
Pedal	Sp	Sgr Fa		Wo G Gm Sch	1 2 3	Z1 Z2	F	T
Pendler		STR G						
Personen								
Personen								
Personen								

Wort VERBINDEN mit: text Wort Klasse Quelle

Fig. 8. Building a relationship between two terms in INDEX.

KLASSIFIKATION		Nr: 5231 S: M:	
Notation	B T Eb U	1 2 3	F
AfAa0005 ::FA	K 1		
. * Aachener und Münchener Lebensversicherung AG allgemein			
WORT		Nr: 8428 H: 0 S: M:	
Aachener und Münchener Lebensversicherung AG			
Zusatz:		VEKO	

Wort VERBINDEN mit: text Wort Klasse Quelle

Fig. 9. Building a relationship between a classification and a term in INDEX.

WORT		Nr: 482 H: 0 S: M:	
Öffentliches Verkehrsmittel			
Sp	Sgr Fa	Wo G Gm Sch	1 2 3
	G		Z1 Z2
			F
			T
QUELLE		Nr: 5 S:x M:	
Meyers Enzyklopädisches Lexikon in 25 Bänden			
Zitcode	Jahr DT E Sp	Geg	L Sgb 1 2 3 Standort
MEYERS	1980 lx de		xyz

Wort VERBINDEN mit: text Wort Klasse Quelle

Fig. 10. Building a relation between a source and a term in INDEX.

INDEX 4.1			PROTERM 2.5	TMS 1.0
1. Synonymy	Descriptor	Non-Descriptor	X	X
2. Abbreviation	Full Descriptor	Abbreviation		
3. Identical Use	Term	Same Meaning		X
4. Quasisynonymy	Descriptor	Quasi-Synonym		
5. Alternative	Descriptor (alternative)	Same Meaning		
6. Combination	Single Descriptor	Combined Term	X	
7. Use	Basic Term	Example of Use		
8. Hierarchical Relation	Broader Term	Narrower Term	X	X
9. Generic Relation	Broader Term	Narrower Term		
10. Partitive Relation	Whole	Part		
11. Causality	Cause	Effect		
12. Appurtenance Relation	General Term	Instance		
13. Field Relation	Field	Element		
14. Associative Relation	Term	Related Term	X	X
15. Temporal Sequence	Predecessor	Successor		
16. Splitting	Predecessor	Successor		
17. Union	Predecessor	Successor		
18. Opposite	Term	Opposite Term		
19. Use Warning	Term	In Contrast To		
20. Production	Producer	Product		
21. Relation of Material	Material	Object		
22. Other Language	Language	Foreign Language		
23. Translation	Source Language	Target Language		
24. Systematic	Descriptor	Notation	X?	X?

one Notation/Facet per Term only

Table 3: The different kinds of relationships in INDEX, PROTERM and TMS

Annexe 14

LISTE DE DIFFERENTS TYPE DE SORTIE GERE PAR :

INDEX=I ; PROTERM=P ; TMS=T

Liste alphabétique des termes.....	I P T
Liste systématique des termes.....	I P
Minithesauri/Parthesauri.....	I P
Liste de termes sans relations.....	I P T
Liste de termes sélectionnée selon différents critères....	I
Liste de synonymie.....	I P
Thesaurus systématique.....	I P T
Thesaurus alphabétique.....	I P T
Thesaurus avec différent BT/NT.....	T
Listes Hiérarchique BT/NT.....	P
Listes de différente classifications.....	I
Listes de différente source register.....	I

ADRESSES DES VENDEURS DE QUELQUES LOGICIELS :

BNF Metals Technology Center
Wantage Business Park
Dechworth Rd., Wantage,
Oxfordshire OX12 9JB-England

BRS Software Products
8000 Westpark Drive
Mclean VA 22102
703/442-3870

CONCEPTEURS DISTRIBUTEURS
17, Rue de la Plaine
75020 Paris

DECALOG
Tour Europe
26000 Valence

G.B.C
Tour CFEM
6, bd Henri Sellier
92150 Suresnes

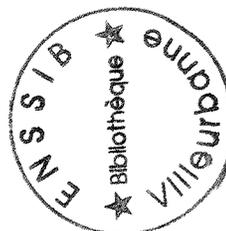
Information Management & Engineering
334 Commonwealth AVE.
Boston, MA 02115
671/247-3281

Liu-Palmer, 2272 Colorado Bd.
Los Angeles, Ca 90041, Suite 1280, USA.
213/664-7178

PLATINUM S.A.
90, Rue de Miromesnil
75008 Paris

PROGIPLUS
Tour Neptune
92086 Paris - La Défense

Pyramid Computer Systems Ltd.
59, Vastern Rd, Reading, Berks
RG1 8BY, ENGLAND



TAMIL
91, bd Edmond-Rasiand
92500 Rueil-Malmaison

Vernon Systems Ltd.
P.O.Box 6909
Auckland, NEW ZEALANDE
64/649/302-3147

VOLT
65, Rue Berthier
78000 Versailles



959065A