

Etude de faisabilité pour la conception d'une base de données bibliographique open source

Denisa Eustasius

► **To cite this version:**

Denisa Eustasius. Etude de faisabilité pour la conception d'une base de données bibliographique open source. domain_shs.info.docu. 2005. <mem_00000280>

HAL Id: mem_00000280

https://memsic.ccsd.cnrs.fr/mem_00000280

Submitted on 24 Nov 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Remerciements

Je tiens à adresser mes plus chaleureux remerciements à Madame Elisabeth Noël, ma responsable de stage et l'initiatrice du projet, qui m'a aidée et soutenue pendant tous ces quatre mois. Ses conseils avisés et sa patience m'ont donné la possibilité de dépasser les moments d'hésitation et d'embarras et de rediriger le projet pour avoir les meilleurs résultats. Et c'est grâce à elle que j'ai retiré autant de satisfactions de ce stage.

Je la remercie aussi pour sa constante disponibilité à mon égard et pour m'avoir donné la chance d'effectuer un stage aussi riche d'enseignements.

Ma gratitude s'adresse également à Monsieur Mohamed Hassoun pour son écoute et sa compréhension et à tout l'ensemble des enseignants de l'Essib pour toutes les leçons qu'ils nous ont donné pendant cette année.

Résumé :

Le service Formist souhaite mettre en place une base de données bibliographique francophone rassemblant tous les documents sur la maîtrise de l'information. Ce mémoire décrit les analyses effectuées et les solutions apportées lors de la conception logique de cette base. Une étude de marché sur les logiciels open source de gestion des bibliographies a été aussi réalisée afin de choisir le plus adapté aux besoins de la base de données. Un thésaurus bilingue (français – anglais) spécifique au domaine de la formation documentaire a été également conçu pour servir à l'indexation et à la recherche ; il sera complémentaire à la base.

Descripteurs : base de données bibliographique, maîtrise de l'information, open source, logiciel de gestion de bibliographie, thésaurus

Abstract :

Formist wishes to set up a bibliographical database gathering all the documents written in French relating to the topic of information literacy. This thesis describes the analyses carried out and the solutions proposed throughout the logical development phase of this database. A market study was conducted about open source bibliographic management software in order to select the one most appropriated to the needs of the database. A bilingual thesaurus (French - English) dealing with the field of information literacy was also elaborated to be used for the indexing and research; it will be used to complement the database.

Keywords : bibliographic database, information literacy, open source, bibliographic management software, thesaurus

Toute reproduction sans accord express de l'auteur à des fins autres que strictement personnelles est prohibée.

Sommaire

INTRODUCTION.....	7
CONTEXTE DU STAGE.....	9
CONTEXTE GENERAL	9
PRESENTATION FORMIST	10
ANALYSE EXTERNE.....	13
L'IDENTIFICATION ET LE TRAITEMENT DES BESOINS	16
LES BESOINS DE L'ACTIVITE	16
LES BESOINS DU SYSTEME	17
CHOIX D'UN LOGICIEL.....	19
LES PROGRAMMES INFORMATIQUES EN CODE OUVERT	19
<i>Les SGBD open source</i>	21
<i>Les logiciels bibliographiques en code ouvert</i>	22
L'ETUDE DE MARCHE	24
FONCTIONNALITES DE REFBASE.....	31
GESTION DES UTILISATEURS : TYPES DE PROFILS.....	31
<i>Administrateur</i>	31
<i>L'utilisateur inscrit</i>	31
<i>L'utilisateur non-enregistré ou "en mode visualisation"</i>	32
AJOUT DES NOTICES	33
RECHERCHE.....	33
<i>Types de recherches</i>	33
<i>Affichage</i>	34
<i>Affiner la recherche</i>	34
EXPORT / IMPORT.....	35
CONSTRUCTION DU THESAURUS SUR LE THEME DE LA MAITRISE DE L'INFORMATION.....	36
PREMIERES DEMARCHES	36

LE CONTACT AVEC LES ETABLISSEMENTS INTERNATIONAUX	37
ANALYSES PRELIMINAIRES	38
<i>Thésaurus vs. langage naturel</i>	38
<i>Considérations générales</i>	39
LE SQUELETTE DU THESAURUS.....	41
<i>Relations</i>	41
<i>Réalisation et graphique</i>	42
CONCLUSION	45
BIBLIOGRAPHIE.....	46
TABLE DES ANNEXES.....	48

Introduction

Mon stage s'est déroulé au sein du service FORMIST (FORMation à l'Information Scientifique et Technique) et a porté sur l'étude de faisabilité d'une base de données bibliographique dans le domaine de la maîtrise de l'information : BOSIL (Base Ouverte Sur Information Literacy). Ce stage intervient dans un contexte plus général : celui de la modernisation du service Formist. La base de données sera complémentaire au nouveau site, entièrement restructuré, disponible au public à partir de Juin 2005.

Pour réussir ma mission, j'ai travaillé en plusieurs étapes en fonction des tâches qui ont été tracées au début du stage et des contraintes qui ont surgi au fur et à mesure.

- Date de début du stage : 23 mai 2005

23-31 mai :

- analyse des bases de données scientifiques existantes et définition des champs et des services de la future base ;
- analyse externe : veille concurrentielle sur les bases de données dédiées au domaine.

1-20 juin :

- analyse des besoins ;
- création d'un « portrait idéal » du logiciel souhaité ;
- conception du squelette et du cahier des charges du thésaurus.

21-30 juin :

- étude de marché sur les logiciels existants.

1-10 juillet :

- définition de la structure logique de la base (MCD, MLD, MCTs) ;
- circulation des données à l'intérieur de la base.

11-22 juillet :

- analyse des services possibles et à offrir aux utilisateurs, aux partenaires (import/export, impression, flux rss, ...)

- recherche de partenaires.

23 juillet-21 août : vacances (fermeture de l'établissement) :
rédaction du cahier des charges et du mémoire de stage.

22-31 août :

- analyse des conséquences techniques, en fonction des
choix ;

- proposition de plusieurs scénarii ;

1-22 sept :

- choix du logiciel ;

- analyse « économique » : besoins de gestion, compétences
requis pour la mise en place physique et pour la maintenance de la
base de données ;

- calendrier.

- Date de fin du stage : 22 septembre 2005

Ce mémoire présente en détail les phases de déroulement du projet énoncées ci-dessus et expose les réalisations faites et les solutions proposées. Il est composé de quatre chapitres chacun étant consacré à une grande étape du projet : l'analyse externe, l'analyse des besoins, le choix du logiciel et la conception du thésaurus.

Contexte du stage

Contexte général

Selon le [granddictionnaire.com](http://www.granddictionnaire.com) la « maîtrise de l'information » ou « information literacy » (terme utilisé en anglais) est :

« L'ensemble de compétences permettant de reconnaître l'existence d'un besoin d'information, d'identifier l'information adéquate, de la trouver, de l'évaluer et de l'exploiter en relation avec une situation donnée, dans une perspective de résolution de problème. » (<http://www.granddictionnaire.com>)

Depuis 1974, époque où on a commencé à parler d'information literacy, un nombre toujours croissant de bibliothèques, d'institutions d'enseignement, d'organisations publiques et privées, de professionnels de l'information s'intéresse au domaine. La littérature spécifique est riche et provient, pour la plupart, de pays industrialisés anglophones, surtout des Etats Unis et de l'Australie. D'où le nombre important d'institutions et d'études en langue anglaise.

En ce qui concerne la partie francophone, le domaine n'a pas encore une appellation précise. « Formation documentaire », « compétence informationnelle », « usage de l'information » etc. sont tous utilisés pour définir cette discipline ; mais, parmi tous ces termes, l'expression « maîtrise de l'information » semble s'imposer. Les efforts dans ce domaine sont notable : programmes universitaires, études scientifiques, services et structures dédiées etc.

Le paysage français de la maîtrise de l'information comprend aujourd'hui trois niveaux :

- local : les SCD (Services Communs de la Documentation) et les instituts d'enseignement, où les acteurs sont les bibliothécaires et les enseignants ;
- régional : les URFISTs (Unités Régionales de Formation à l'Information Scientifique et Technique). Créées en 1982, les URFISTs desservent sept grands régions de France : Paris, Rennes, Strasbourg, Bordeaux, Lyon, Toulouse et Nice.

- national : FORMIST (FORMation à l'Information Scientifique et Technique). Hébergé par Enssib (Ecole nationale supérieure des sciences de l'information et des bibliothèques), Formist a été créé en 1999 pour répondre aux besoins des formateurs, enseignants et étudiants du supérieur concernant la maîtrise de l'information. Le site, qui est ouvert en fait à tous sans formalités, propose des différents outils et documents (cours méthodologiques de recherche de l'information et d'utilisation des outils de recherche, articles scientifiques, annonces de colloques etc.) pour apprendre à chercher l'information.

Présentation FORMIST

Le service Formist a plusieurs missions :

- développer et animer le réseau de formateurs : collaborer avec les autres institutions centrées sur la maîtrise de l'information et former les futurs bibliothécaires à devenir eux-mêmes des formateurs à la maîtrise de l'information ;
- concevoir et diffuser des documents pédagogiques dans le domaine de l'IST ;
- maintenir une veille technologique sur les sources du domaine de la formation documentaire.

Ainsi, parmi les activités essentielles du service Formist, les plus importants sont :

1) la maintenance et le développement du site web Formist.

A travers son site, Formist signale des documents scientifiques et pédagogiques concernant le domaine de la formation documentaire. Les thèmes abordés couvrent, le plus possible, l'intégralité du domaine :

- recherche d'information sur Internet ;
- outils de recherche : typologie (moteurs de recherche, bases de données, catalogues de bibliothèque, ...) et usage ;
- traitement, diffusion, exploitation et évaluation de l'information ;
- comment écrire et lire une bibliographie ;
- technologies de l'information.

Le site s'adresse particulièrement aux étudiants de l'enseignement supérieur et aux formateurs. En conséquence, il recense plusieurs types de documents pour répondre aux besoins de chacun :

- articles et autres publications scientifiques ;
- cours : s'adressant exclusivement aux étudiants. Les cours requièrent la validation de CES (Comité Editorial et Scientifique) avant que leur fiche descriptive soit publiée sur le site;
- supports techniques d'utilisation d'outils ;
- annonces de colloques et de formation ;
- comptes rendus de rencontres professionnels, de conférences etc.

La base de documents est le résultat de deux activités réalisées de façon continue : la veille documentaire faite par le service et les propositions faites par les utilisateurs (souvent par les auteurs eux-même). Après validation (pour les ressources pédagogiques), elles sont référencées via une fiche descriptive, conforme à la norme LOM (Learning Object Metadata). Cette norme a été adopté pour l'indexation des ressources depuis la restructuration du site en début 2005.

Le site propose aussi plusieurs services à ses visiteurs ; outre les services classique comme la recherche (différents types de recherche) et l'accès aux documents, l'utilisateur peut s'abonner aux fils RSS selon ses besoins ou à un système d'alerte personnalisé, peut sélectionner les documents qui l'intéressent et les mettre dans son panier, suivre des parcours autour de thèmes précis etc.

2) formations

Le service Formist assure des formations dans le domaine de la maîtrise de l'information à l'Enssib : pour DCB (Diplôme de Conservateur de Bibliothèque) et FIB (Formation Initiale des Bibliothécaires).

Les deux conservateurs des bibliothèques qui animent ce service offrent des enseignements aux étudiants (et futurs professionnels de l'information) et publient des brochures et d'autres documents pédagogiques (support papier et en ligne). Elles s'impliquent aussi dans la direction des études de recherche et des mémoires.

3) participations aux et organisation des colloques, conférences, ateliers etc.

Les « Rencontres Formist » sont des réunions internationales annuelles consacrées à la formation documentaire. Elles sont organisées chaque année sur des thèmes différents. Les rencontres Formist de 2005 ont eu pour sujet « Parcours de formation documentaire des étudiants : à qui de jouer ? Développer les compétences informationnelles dans un cursus disciplinaire ». La journée a réuni plus de cent cinquante formateurs et spécialistes en IST venus de toute la France, de Belgique, Canada, Suisse et Grande Bretagne. La publication en ligne, sur le site de l'ENSSIB, des Actes des Rencontres est prévue pour l'automne 2005.

Outre cette conférence, Formist organise des journées d'étude (LOM) et des colloques (dans le cadre de l'Enssib ou à l'extérieur) et anime des tables rondes, comme celle du Salon du Livre ("Collaborer entre réseaux pour former les usagers à la maîtrise de l'information").

Il est aussi intervenu dans d'autres réunions et conférences sur le thème de la maîtrise de l'information ou connexe en France ou à l'étranger. Au cours de l'année 2004-2005, Formist a participé activement au Séminaire IVIG, organisé par l'association des bibliothèques universitaires tchèques à Prague, aux quatrièmes Rencontres Internationales DHI à Université de Juárez, Mexique, au Congrès LOEX, Louisville, Etats-Unis et au groupe de travail sur la formation des formateurs de l'URFIST de Lyon.

Analyse externe

L'analyse externe, précédant la constitution de la base, s'est réalisée en deux étapes avec des objectifs différents :

- étude des bases de données scientifiques, pour avoir un exemple en ce qui concerne leur organisation : structures et services ;
- veille concurrentielle sur les bases de données du domaine.

1. Les bases de données scientifiques sur lesquelles nous avons porté notre recherche sont : LISA, ERIC, PASCAL et FRANCIS.

Les points pris en considération :

- la structure : champs, données ;
- les types de documents ;
- la forme de saisie des données : casses, ponctuation ;
- les facilités d'interrogation et de re-interrogation (pour affiner la recherche) ;
- la couverture linguistique de la base et de l'indexation ;
- les services offertes : types de recherche et facilités (opérateurs booléens, troncature, mots vides, ...), outils complémentaires (vocabulaire contrôlé, index, ...), impression, historique, liens hypertexte internes et externes (vers les maisons d'édition, autres bases etc.).
- affichage des résultats : types et choix, nombre de notices par page, champs affichés, format du libellé des champs, ordre, tri.

Cette étude nous a permis de tracer à la fin tous les détails à prendre en compte pour BOSIL :

- les types de documents avec leurs caractéristiques spécifiques ;
- la définition des champs et de la nature des données pour chaque champ (c.f. Annexe 1) ;
- les services à envisager : import de données, export de notices dans différents formats, impression, sauvegarde, historique, envoi par mail des résultats, alertes DSI, abonnements aux fils RSS;

- facilités : navigation, recherche (booléenne, multi-champs, expression exacte, troncature, regroupement des termes, sauvegarde de la stratégie de recherche, respect ou non des casses), styles bibliographiques, affichage des résultats (par défaut ou par choix, nombre de notices par page, formats, champs, tri, ordre), sélection des références, consultation du thésaurus et des index alphabétiques.

2. La veille concurrentielle a eu comme but de trouver les autres bases de données sur le thème de la maîtrise de l'information. Elle s'est effectuée sur l'espace francophone et anglophone. Dans la première partie nous sommes orientée vers les sites des institutions (associations, bibliothèques, instituts etc.) susceptibles de s'occuper du domaine de l'information literacy. Les forums et les listes de discussions, les annonces de conférences et les comptes rendus des réunions, colloques, congrès, séminaires ont été aussi pris en compte.

Bien qu'il existe un nombre impressionnant d'institutions intéressées par le domaine, qui collectent et publient sur leurs sites des documents relatifs à la maîtrise de l'information - soit des cours et des autres documents pédagogiques, soit des résultats des recherches, soit tout simplement des essais relatifs au domaine, à l'adresse des formateurs et des étudiants - il existe peu de bases de données spécialisées. Nous avons pu trouver des projets mort-nés lancés depuis quelques années, des initiatives plus récentes pas encore mises en pratique et des bases déjà conçues mais pas finies.

Finalement, quatre bases de données (viables) et un annuaire dédiés à la maîtrise de l'information ont été découverts :

- Boston Library Consortium Information Literacy Database

<http://webapps.lib.uconn.edu/BLC/>

- Information Literacy database

(besoin de FileMaker Pro v.6 pour lancer la base et d'un mot de passe)

<http://www.iowa-city.k12.ia.us/Library/LibAdmin/InfoLitDB.htm>

- University of Washington Libraries : base de données.

Objectif : aider les étudiants dans leurs recherches d'information. Besoin d'y être étudiant pour recevoir un mot de passe.

<http://www.lib.washington.edu/uwill/>

- Information Literacy Online

<http://webapps.lib.uconn.edu/InfoLit/default.cfm>

- DORIL : Directory of Online Resources for Information Literacy

<http://bulldogs.tlu.edu/mdibble/doril/>

Une fiche descriptive de chacun a été réalisé, portant sur :

- type de base de donnée ;
- la localisation (URL) ;
- maintenance : développeur / hébergeur, dates (de création, de mise à jour ; périodicité) ;
- types de documents ;
- le volume de la base (nombre total de références) ;
- types de recherche ;
- services offertes ;
- types d'affichage des résultats.

L'identification et le traitement des besoins

Avant de concevoir une base de données, il est préférable sinon essentiel d'identifier précisément tous les besoins, afin de pouvoir y répondre convenablement.

Il y a deux types différents de besoins à identifier et à traiter pour avoir le meilleur résultat : les besoins de l'activité et les besoins du système. Pour les premiers, on doit prendre en compte le fonctionnement « quotidien » de l'activité. Les besoins du système sont étroitement liés à la conception effective de la base. Les deux types de besoins sont complémentaires.

Les besoins de l'activité

Dans un premier temps, nous avons établi une liste de critères d'analyse, critères généraux pour toute base de données bibliographique ; ensuite, nous avons noté les fonctionnalités souhaitées qui devrait être assuré par le logiciel ou ajoutées dans la phase de conception physique et technique de la base. Pour le résultat en détail de notre analyse voir l'Annexe 2.

Bref, nous avons identifié les besoins suivants:

- interopérabilité et compatibilité technique avec le site Formist (construit sous SPIP Agora, un système open source pour la création des sites web fonctionnant sur un serveur Apache, une base MySQL et PostgreSQL, le système d'exploitation Linux et les langages de programmation PHP et XML) ;
- autonomie technique : ce qu'autorise un logiciel en code ouvert et modifiable, qui permet de formater la structure de la base selon ses propres intérêts ;
- un outil de travail collaboratif et, donc, multi utilisateurs, pour pouvoir impliquer un certain nombre de partenaires dans le processus d'alimentation de la base ;
- donner accès libre et gratuit aux utilisateurs / aux internautes de consulter les ressources, mais, en même temps, avoir un outil qui gère les comptes utilisateur

(pour enregistrer les partenaires et pouvoir leur donner accès au back-office de la base) ;

- indexer les documents à l'aide d'un thésaurus dédié au domaine et complémentaire à la base de données ;
- permettre une interrogation souple avec toutes les fonctions nécessaires à une recherche documentaire complexe ;
- faire des statistiques sur le volume et la consultation de la base.

Les besoins du système

Les informations rassemblées lors de l'analyse des besoins de l'activité sont converties pour pouvoir déduire et définir les besoins du système. Cette phase se décline en plusieurs sous étapes : l'identification et le regroupement des données, la définition des champs et des attributs, la définition des relations, le choix du type de modèle et d'une implémentation et, à la fin, l'évaluation de l'analyse qui a comme résultat l'établissement des MCD (Modèle conceptuel des données) et MLD (Modèle logique des données).

Pour BOSIL, étant une base de données bibliographique, l'élément principal est le document ; c'est le document qui est référencé et c'est toujours lui qui est recherché. La table DOCUMENT constitue, donc, la seule entité majeure du MCD ; les entités mineurs sont : les auteurs, les tables désignant les caractéristiques de chaque type de document, le créateurs de la notice et la table dédiée aux personnes qui proposent des documents (table qui servira plutôt à la fidélisation du public).

Pour le modèle conceptuel de données, voir l'Annexe 3.

Il faut peut-être mentionner que pour la réalisation du MCD nous avons utilisé la méthode de la généralisation qui permet de classer les entités en génériques et spécifiques et de mettre en évidence les propriétés particulières de chaque sous-type d'entité. Dans notre cas, l'entité générique est le DOCUMENT et ses entités spécifiques sont : OUVRAGE, PERIODIQUE, COLLOQUE, THESES et

RAPPORTS OFFICIELS. A leur tour, les entités OUVRAGE et PERIODIQUE sont génériques pour CHAPITRE D'OUVRAGE et, respectivement, ARTICLE. La technique utilisée est « la partition » (symbolisée par le signe + encerclé) où les attributs d'une entité spécialisée sont propres à cette entité-là et uniquement à elle. Afin de ne pas charger le MCD, nous avons noté uniquement les clés primaires en ajoutant à la fin du schéma, séparés, les attributs spécifiques pour chaque table. Ils sont aussi présents de manière plus explicite dans le MLD. Pour le Modèle logique de données voir l'Annexe 4.

Comme type de modèle, nous avons choisi celui qui est le plus approprié pour une base de donnée bibliographique et qui satisfait le mieux les besoins de l'activité : le modèle relationnel ; il est, de toute manière, le modèle de base de donnée le plus utilisé à l'heure actuelle. L'implémentation sera implicitement sur un SGBDr (Système de Gestion de Bases de Données relationnelles). Celui-ci a plusieurs avantages :

- facilité d'accéder aux données ;
- modification facile de la structure interne de la base ;
- représentation de manière logique des données ;
- on peut développer facilement et rapidement des requêtes pour récupérer les notices ;
- interrogation de la base à l'aide d'un langage standard (SQL).

Choix d'un logiciel

En raison des besoins définis à l'étape précédente nous nous sommes dirigée vers deux types de solutions informatiques pour l'implémentation de la base de données :

- les SGBDr (Systèmes de Gestion de Bases de Données relationnels)
- les logiciels bibliographiques.

L'étude de marché s'est limitée aux solutions open source, imposées lors de l'analyse des besoins. Deux raisons principales nous ont fait choisir les OSS (Open Source Software) :

- le besoin d'interopérabilité avec le site Formist ;
- leur nombreuses avantages (détaillés plus bas).

Les programmes informatiques en code ouvert

Les OSS sont souvent assimilés aux logiciels libres, mais entre « libre » et « open source » il y a plusieurs différences. Les logiciels libres sont des solutions informatiques propriétaires, commerciales, distribués gratuitement. Les logiciels open source sont des solutions souvent gratuites, qui permettent à l'utilisateur de consulter et modifier le code selon ses besoins. Il peut être ainsi transformé en un tout autre logiciel. Il y a également des logiciels open source commerciaux, peu nombreux ; la modification du code source est encore possible, mais sous contrôle.

La différence entre « libre » et « open source » réside principalement dans la licence qui accompagne le produit. Si les logiciels libres ont des licences commerciales, les OSS sont distribués sous une licence de type GPL (« General Public Licence ») ou voisine : LGPL (« Lesser General Public Licence »), MPL (« Mozilla Public Licence »), BSD (« Berkeley Standard Distribution »).

Ces licences se divisent en deux types : copyleft (GNU : « GNU's not Unix ») et non-copyleft (tous les autres : BSD, Apache, ...). Gratuits au départ, les produits à licence non-copyleft peuvent devenir, par des adjonctions ultérieures, des logiciels propriétaires. Les licences copyleft ne permettent pas cette évolution :

le code, indifféremment du nombre de modifications survenues, reste libre à la disposition de toute personne qui veut le consulter.

La plupart des logiciels open source ont une licence GNU. Celle-ci a été développée par Richard Stallman en 1989 ; la version qui est utilisée aujourd'hui est 2.0 sortie en 1991. Elle prévoit la liberté de distribuer des copies des logiciels, de consulter et modifier le code source, de publier les formes modifiées (avec l'obligation de laisser visible et consultable le nouveau code). Malgré les termes de cette licence, d'habitude il y a une seule personne (ou un nombre très petit de volontaires) qui crée et développe le programme. Il existe souvent une ou plusieurs listes de discussions autour d'un produit. Une de ces listes est généralement dédiée aux développeurs, aux personnes qui discutent sur leurs idées et contribuent à l'ajout de nouvelles fonctionnalités ou à l'enrichissement de celles existantes. Les autres listes sont des espaces où les utilisateurs qui ne sont pas nécessairement à l'aise techniquement demandent de l'aide pour les problèmes rencontrés lors de l'installation ou de l'utilisation du logiciel.

Le tableau ci-dessous révèle les principales différences entre un logiciel sorti sous une licence commerciale et un logiciel en code source ouvert.

	Logiciels à licence commerciale	Logiciels open source (à licence GNU)	
Inconvénients	Payants (sauf les logiciels « free »)	Gratuits pour la plupart	Avantages
	Pas d'accès au code source.	Code source visible et modifiable.	
	Droit de copie et d'usage limité.	Possibilité de copier le logiciel un nombre illimité de fois, de l'installer sur tous les postes que l'on a besoin.	
	Pas de droit de revente.	Une fois amélioré ou complété, le code peut être republié avec l'obligation de mettre à disposition les modifications vis-à-vis d'autres utilisateurs.	

Avantages	L'existence d'une garantie et la possibilité de demander du support auprès de l'éditeur.	Pas de garantie. Il peut être difficile pour une entreprise de trouver du service autour de logiciels open source. Il peut s'avérer compliqué de trouver des consultants, des spécialistes, des personnels assurant le support. Il y a des entreprises qui offrent les logiciels OS gratuitement, mais avec un support payant. L'article 1 de la GPL stipule que : « Vous pouvez demander une rétribution financière pour la réalisation de la copie et demeurez libre de proposer une garantie assurée par vos soins, moyennant finances ».	Inconvénients
	Documentation complète et mise à jour.	Documentation souvent inexistante.	
	Performances notables.	D'habitude, un logiciel en code source ouvert n'a pas toutes les fonctionnalités d'un logiciel propriétaire, mais on observe que chaque jour les OSS montent en puissance, les versions évoluent, les trous de fonctionnalités sont comblés et les performances améliorées.	

Les SGBD open source

Bien que le marché des SGBDs soit dominé par les solutions proposées par les éditeurs commerciaux : IBM (DB2), Oracle (Oracle9i) et Microsoft (Access, SQL Server), les OSS sont aussi reconnus et assez largement utilisés. Parmi les SGBDs open source, les plus populaires sont :

- MySQL (<http://www.mysql.com>),
- MaxDB (<http://www.mysql.com/products/maxdb/>) - l'ancien SAP DB,
- Ingres (<http://opensource.ca.com/projects/ingres/>),
- PostgreSQL (<http://www.postgresql.org>),
- Firebird (<http://firebird.sourceforge.net/>).

Les avantages de ces systèmes sont : le faible coût (coût nul de licence, mais assistance souvent payante), la fiabilité, la rapidité de l'installation, la simplicité d'utilisation. PostgreSQL et MySQL ont des fonctionnalités similaires ; PostgreSQL comporte plus de fonctions évoluées, mais MySQL dispose d'une base plus importante et offre un support plus affirmé.

De loin MySQL est le plus connu et le plus utilisé des systèmes de gestion de base de donnée en code ouvert : il est installé sur plus de 4 millions de serveurs dans le monde et continue d'être téléchargé plus de 30 000 fois par jour. Yahoo ! l'utilise pour la section de petites annonces, mais aussi pour les rubriques : Yahoo Sport, Yahoo News et Yahoo Finance. Google, Cisco, NASA, Lucent Technologies, Motorola, HP, Xerox, Sony Pictures l'utilisent aussi.

Les logiciels bibliographiques en code ouvert

Un logiciel bibliographique est un programme conçu pour gérer des listes de références bibliographiques. Il permet, en même temps, de stocker et gérer de manière personnelle des notices, de les retrouver facilement dans la base, d'automatiser la mise en forme des citations et de la bibliographie.

A la différence de logiciels propriétaires (comme EndNote), les logiciels bibliographiques open source sont souvent des outils de travail collaboratif et, en conséquence, multi-utilisateurs, qui permettent de gérer de grosses bases de données (bibliographiques). D'habitude, on a trois types de profils : l'administrateur, l'utilisateur enregistré et l'utilisateur non enregistré (« en mode visualisation »). Ces logiciels sont conçus pour la plupart avec l'idée d'un laboratoire de recherche où l'équipe partage une bibliographie et chaque membre ajoute de nouvelles entrées à la base commune, en ayant en même temps des listes bibliographiques avec des annotations personnelles. La philosophie de ces logiciels repose donc sur l'idée que la base de donnée est constituée d'un corpus central de listes bibliographiques thématiques créées et alimentées par le(s) administrateur(s). Il existe un nombre d'utilisateurs autorisés, avec un mot de passe, à accéder à la base; ils ont le droit d'ajouter et de modifier les notices. Le droit de suppression revient uniquement à l'administrateur. Pour les usagers authentifiés, les logiciels proposent un nombre de services spécifiques, comme la possibilité d'avoir leurs

propres listes bibliographiques, un style personnel d'écriture des références bibliographiques, de faire un import des données depuis une base externe etc. Au cas où la base est consultable en ligne, les autres utilisateurs (soit les internautes) accèdent en mode visualisation; ce qui veut dire qu'ils peuvent faire des recherches, visualiser les notices qui les intéressent et exporter les résultats. Mais ils n'ont pas le droit de modifier la base centrale ou d'avoir des listes personnelles.

Pour la plupart, les logiciels bibliographiques open source sont construits sous un SGBD (d'habitude MySQL) ce qui facilite la gestion d'une base de données bibliographique. Un SGBDr offre les fonctions essentielles pour concevoir la base :

- structurer la base de données : tables, attributs, champs, possibilité de paramétrage ;
- gestion des utilisateurs ;
- recherche des données (en langage SQL).

Les logiciels bibliographiques ajoutent un nombre important d'éléments :

- une interface graphique plus conviviale ;
- gestion de plusieurs listes bibliographiques à la fois ;
- formulaires de saisie différents en fonction du type de document ;
- possibilité de créer des styles bibliographiques ;
- éventuellement, un système d'indexation des ressources ;
- import depuis une autre base de données locale ou distante ;
- export de notices en formats divers : même dans un document word déjà créé (les notices bibliographiques sont exportées directement et automatiquement à l'endroit indiqué par l'utilisateur, dans le document souhaité).

En conséquence, un premier choix s'est fait au niveau du type du logiciel. La solution la plus sage était un logiciel bibliographique conçu sous un SGBDr, pour avoir toutes les fonctionnalités dans un seul produit.

L'étude de marché

L'étude de marché s'est portée sur les logiciels de gestion de références bibliographiques en code source ouvert. Pour cette étude, nous avons mené une recherche de type documentaire. L'essentiel de la documentation que nous avons trouvée sur ce type de logiciels provient du web; une source riche en informations a été le site SourceForge, site dédié aux programmes open source. Ensuite nous avons visité le site officiel de chaque produit ou de son concepteur. Il y avait des produits sans site officiel ; dans ce cas nous nous sommes dirigée vers d'autres sites qui recensent ce type d'outils en ajoutant des petites descriptions. Les listes de discussions nous ont fourni aussi des données utiles.

Quoique les logiciels bibliographiques ne soient pas très bien connus en général, le monde de développeurs de ces outils est assez large ; nous avons pu trouver plus de trente produits :

Nom logiciel	Adresse web
ALLBIB	(allbib.sourceforge.net)
BASILIC	(http://artis.imag.fr/Software/Basilic/)
BIBCURRED	(http://sourceforge.net/projects/bibcursed)
BIBEDIT	(http://www.iui.se/staff/jonasb/bibedit/)
BIBLIOGRAPH	(http://www.bibliograph.org/)
BIBLIOPERA	(http://bibliopera.mom.fr/)
BIBLIOSHARE	(http://www.bibliograph.org/links/biblioshare)
BIBLIO X	(http://www.silmaril.ie/bibliox/biblioxdoc.html)
BIBLIOZ	(sans site officiel)
BIBORB	(http://biborb.glymn.net/doku.php)
BIBULUS	(http://www.nongnu.org/bibulus/)
BIBUS	(http://bibus-biblio.sourceforge.net/)
BIBSTER	(http://bibster.semanticweb.org/)
CMF BIBLIOGRAPHY AT	(sans site officiel)
CTAN DIRECTORY: BIBLIO / BIBTEX	(sans site officiel)
DOCUMENT ARCHIVE	(http://docarc.sourceforge.net/)
DOCUMENT DATADASE	(http://docdb.sourceforge.net/)
GBIB	(http://gbib.seul.org/)
JABREF	(http://jabref.sourceforge.net/)
LIT W 3 [logiciel	(sans site officiel)

Nom logiciel	Adresse web
allemand]	
LIBDB	(http://www.libdb.com/)
MAMBO BIBLIOGRAPHY COMPONENT	(http://bibliography.mamboforge.net/)
PYBLOGRAPHER	(http://pybliographer.org/)
PHPBIBMAN	(http://phpbibman.sourceforge.net/)
PHPBIBTEX	(http://www.rennes.supelec.fr/ren/perso/etotel/PhpBibtexDbMng/)
REFBASE	(http://refbase.sourceforge.net/)
REFDB	(http://refdb.sourceforge.net/)
TKBIBTEX	(http://www.cat.csiro.au/cmst/staff/pic/tkbibtex.html)
WIKINDX	(http://wikindx.sourceforge.net/)
YAPBIB	(http://www.cs.jhu.edu/~jcorso/yapbib/)
Z NOTE	(sans site officiel)

Chacun a été décrit dans la limite du volume des informations disponibles. Tous ne possédaient pas une présentation sur leur site ou sur un site connexe ; en fait, la majorité n'avait pas du tout de documentation.

Globalement, il y a quelques fonctionnalités minimales que nous avons retrouvées dans la plupart de ces logiciels :

- gestion des comptes utilisateurs ;
- référencement standard, mais possibilité d'ajouter ses propres champs ;
- quelques styles bibliographiques prédéfinis, mais possibilité d'avoir un style personnel ;
- niveaux différents de recherche : simple et avancée (recherche booléenne et multi-champs) ;
- import / export des données (en différents formats) ;
- ajout du texte intégral ;
- fonctionnement en réseau.

A la fin, les logiciels les plus performants et qui semblaient les plus adaptés à nos besoins ont été choisis et intégrés dans le tableau comparatif suivant :

	BIBLIOPERA	BIBUS	JABREF	REFBASE	REFDB	WIKINDX
VERSION	La version actuelle : 0.7 Les versions prochaines : 0.8 : (mineure) Correction de bugs éventuels sur la version 0.7 ; finalisation de l'aide utilisateur, traduction de l'interface en anglais. 0.9 : Stabilisation de l'architecture d'import depuis des catalogues 1.0 : Stabilisation des formats d'import / export par fichier	1.0.0 La première version publique, 0.8, est apparu en Juin 2004	1.7.1 JabRef est le résultat de la fusion de deux logiciels bibliographiques : Bibkeeper et JBibtexManager	0.8 (sortie le 5 avril 2005)	0.9.5	3.0.7
DEVELOPPEUR	Maison de l'Orient et de la Méditerranée. Reponsable de projet : Marjorie Burghart Marjorie.Burghart@mom.fr	Pierre Martineau pmartino@users.sourceforge.net	Morten Omholt Alver: mortenalver@users.sourceforge.net Nizar Batada: nbatada@users.sourceforge.net	Matthias Steffens : msteffens@refbase.net msteffens@ipoe.uni-kiel.de	Markus Hoenicka: mhoenicka@users.sourceforge.net	Mark Grimshaw : sirfragalot@users.sourceforge.net
SITE OFFICIEL	http://bibliopera.mom.fr	http://bibus-biblio.sourceforge.net/	http://jabref.sourceforge.net/	http://refbase.sourceforge.net/	http://refdb.sourceforge.net/	http://wikindx.sourceforge.net/
DOCUMENTATION	Assez sommaire : elle explique comment installer le programme, mais les derniers chapitres sont encore en cours de rédaction.	Succincte, mais elle pointe les aspects importants. Disponible sur le site, elle explique comment installer et utiliser le logiciel.	N'existe pas ; elle est en construction. Les éditeurs nous invitent à consulter le menu Help.	En construction. Il existe un petit manuel sur l'installation.	Présence d'une documentation très fournie, dont la lecture peut se révéler fastidieuse : le manuel a 261 pages et il est complété par un didacticiel de 52 p. pour les utilisateurs non-administrateurs.	Inexistante. Test démo disponible sur le site, mais pas avec toutes les fonctionnalités. Documentation relative à l'installation, disponible sur le site.
BASE DE FONCTIONNEMENT	Serveur web ; Apache (mais tout serveur http supportant PHP devrait pouvoir être utilisé) ; PHP 4 ; module	Une base MySQL ou SQLite ; le paquet OpenOffice ;	Java 1.4. ou +	Apache, PHP, MySQL	MySQL / PostgreSQL / SQLite. Jade/Open Jade et SP/OSP ; processeurs	MySQL, PHP, Apache

	BIBLIOPERA	BIBUS	JABREF	REFBASE	REFDB	WIKINDX
	permettant les transformations XSLT (Sablotron, DOM XSLT) ; MySQL	Python2.2x et MySQL-phyton 2.5.x ou PySQLite1.1.x.			XSLT et FO ; TEX (pour les bibliographies en LaTeX), un éditeur SGML/XML ; un navigateur Internet ; un serveur Web (Apache), un interpréteur PERL et le client Z39.05	
SYSTEMES D'EXPLOITATION	Tous (mais il a été testé uniquement sur Windows et Linux)	Windows / Linux	Tous	Tous le plates-formes	GNU/Linux, FreeBSD, NetBSD, OSX, Solaris/SunOS, Windows/Cygwin	Tous
LANGUE DE L'INTERFACE	Français – BibliOpéra est, en effet, un produit français. (Anglais prochainement)	Anglais, Français, Portugais, Danois	Anglais par défaut, mais on peut choisir le Français dans le menu « Option »	Anglais	Anglais	Anglais , français, italien
GESTION UTILISATEURS	Trois profils.	Tous les droits pour l'administration ; droits limités pour l'utilisateur. Pas d'informations sur les comptes.	?	Trois profils	Trois profils	Trois profils.
STYLE DE BIBLIOGRAPHIE	Plusieurs disponibles ; possibilité de créer un style personnel.	3 formats bibliographiques disponibles : Medline, Refer et RIS ; possibilité de créer un style personnel.	?	Possibilité d'avoir un style personnel.	Le style principal : RIS, mais on peut créer un style personnel avec XML.	Plusieurs disponibles, selon le modèle anglo-saxon ; possibilité de créer un style personnel.
GESTION DOUBLONS	Ne permet pas l'entrée des doublons.	?	Trouve et enlève les doublons	?	Ne permet pas l'entrée des doublons	Contrôle à la saisie: pour éviter les doublons. Message

	BIBLIOPERA	BIBUS	JABREF	REFBASE	REFDB	WIKINDX
						d'alarme lorsqu'on est en train d'introduire un doublon.
UNICODE	Utilisation de l'unicode (encodage des caractères), permettant d'afficher tous les types d'alphabets sur une même page.	?	?	Oui	Oui	?
INDEX ALPHABETIQUES	Oui	?	Non	?	Non	Oui
GESTION VOCABULAIRE CONTRÔLE	Oui : liste d'autorités.	Non	Non	Non	Non	Non
RECHERCHE D'INFORMATION	Possibilité d'afficher l'ensemble des notices enregistrées, celles enregistrées par l'utilisateur en cours, et les notices dont l'enregistrement est à valider. Possibilité de visualiser l'indexation matière d'une notice. Module pas trop développé !	Recherche simple et complexe. La dernière est la traduction en SQL de la recherche normale donnant la possibilité de l'enrichir.	Multi champs L'option "Fetch" sert aux recherches sur serveurs distants, dans les bases Medline et CiteSeer.	Booléenne, multi champs. Parenthésage (regrouper les termes de la requête). Troncature. Sauvegarde de la stratégie de recherche dans un historique et possibilité de combiner deux ou plusieurs requêtes précédemment faites.	Plusieurs mots-clés, plusieurs champs, recherches dans des listes thématiques différentes, parenthésage etc., possibilité de choisir la manière de visualisation des résultats. Il est possible de sauvegarder les requêtes et les résultats dans un historique et, ensuite, les combiner.	Recherche simple et complexe; affichage général et tri en fonction d'un set de critères.
IMPORT DE DONNEES	Fonction en construction ; on peut toutefois faire l'import grâce à la norme Z39-50; d'autres formats d'import sont en cours de développement.	De PubMed et EndNote	Depuis Web of Science, CAB, Francis, Medline, Science Direct, EndNote,	Basé sur copier / coller. De CSA et EndNote	De Reference Manager, EndNote. Il existe des filtres d'importation pour Medline (en XML),	D'une autre bibliographie BibTex; pourtant cette fonction n'est pas très ergonomique

	BIBLIOPERA	BIBUS	JABREF	REFBASE	REFDB	WIKINDX
			Reference Manager		BibTeX, MARC et DocBook.	(système copier / coller des notices).
EXPORT DE DONNEES	Export prévu par l'application, sous format xml. Fonction en construction.	?	Sous différents formats: Html, Xml, Harvard RTF, EndNote	Sous différents formats : EndNote, RIS, BibTeX et MODS (un standard de format bibliographique XML établi par Library of Congress).	En plusieurs formats : en HTML, PDF, XML etc.	En format RTF, HTML, RIS, EndNote et BibTex. Le programme garde le nouveau document une heure.
GESTION STATISTIQUES	Non	Non	Non	Non	Non	Non
ERGONOMIE ET ESTHETIQUE	- Interface web peu pratique à utiliser. Elle est néanmoins totalement paramétrable grâce aux fichiers xsl (mise en forme de données en langage xml). - L'interface en html est très épurée, presque sans couleurs. Il est cependant possible de la modifier complètement grâce aux fichiers CSS (langage web de mise en forme).	Interface facile à utiliser qui ressemble à une boîte postale électronique. Tout utilisateur d'e-mails devrait bien se débrouiller avec Bibus. Les références peuvent être déplacées par « drag & drop ».	Interface plutôt orientée technique, mais suffisamment esthétique. Des raccourcis ont été créés pour faciliter l'utilisation.	Interface simple, intuitive et ergonomique.	Interface conviviale et facile à utiliser.	Interface conviviale. Impossible d'utiliser les boutons standard d'une page web : il faut passer par la barre de navigation interne.
DISPONIBILITE EN LIGNE	Oui	Oui	?	Oui	Oui	Oui

Fonctionnalités de REFBASE

Rebase est un logiciel open source multi-utilisateurs de gestion de références bibliographiques. La version actuelle, 0.8, est sortie le 5 avril 2005 et la documentation est en construction. Toutefois, une version démo est disponible sur le site officiel du logiciel et on a accès à sept autres applications gratuites en ligne (sept bases de données déjà construites à l'aide de ce produit). L'environnement informatique pour pouvoir le faire fonctionner est composé par le paquet: PHP, Apache et MySQL, composantes qui sont utilisées aussi par le site Formist. L'interface (en anglais) est conviviale et facile à utiliser.

Rebase s'est démarqué lors de l'étude de marché essentiellement par ses multiples facilités créées pour les partenaires de la base de données, que les autres logiciels ne présentent pas, et par le module recherche qui est bien développé permettant interroger la base et affiner la recherche en fonction de plusieurs paramètres. Il a été considéré comme répondant le mieux aux besoins de la future base de données Bosil.

Gestion des utilisateurs : types de profils

En ce qui concerne la gestion des comptes utilisateur Rebase prévoit trois types de profils: l'administrateur, l'utilisateur inscrit et l'utilisateur non enregistré.

Administrateur

L'administrateur est la personne (il peut y en avoir deux ou plusieurs) qui gère(nt) la base de données en ayant tous les droits : sur l'ensemble des notices, sur les utilisateurs autorisés (en ajouter un, le supprimer, changer le mot de passe, ...) etc.

L'utilisateur inscrit

Etant un produit multi-utilisateur, Rebase permet de gérer une macro-base centrale, visible pour tous, mais pour chaque usager authentifié, il réserve des services à part. L'utilisateur inscrit est, ainsi, une personne autorisée avec un identifiant et un mot de passe à accéder à la base. Une fois enregistré, il peut

contribuer à l'alimentation de la base en ajoutant des nouvelles notices ou en enrichissant les autres.

Rebase propose pour les utilisateurs enregistrés (aussi appelés dans le cas de Bosil « les partenaires ») des droits et des services spécifiques : "Marked", "Copy", "Selected", "User Keys", "User Notes", "User File", "User Groups", "Cite Key" et "Related". Ces services permettent tous de stocker des informations personnelles relatives à un enregistrement particulier, par exemple ses mots-clés ou notes; ces informations ne sont pas visibles aux autres utilisateurs ou visiteurs. L'utilisateur enregistré a aussi le droit de créer ses propres bibliographies dans des dossiers séparés. En sélectionnant des ressources qui l'intéressent dans la macro-base ou en rajoutant d'autres, externes, il a la possibilité de construire une bibliographie personnelle. Celle-ci peut être visualisée et téléchargée par les autres utilisateurs (inscrits ou non), mais elle n'est pas modifiable; seul, son auteur, peut ajouter, modifier ou supprimer les items.

Toujours pour les utilisateurs inscrits avec mot de passe, il y a la possibilité de faire des recherches dans leurs propres listes et notices. Un autre service intéressant offert aux usagers enregistrés est de pouvoir sauvegarder de façon permanente et partageable l'historique des recherches faites par chacun d'entre eux. Ils peuvent former des groupes (thématiques ou de travail) pour partager leurs ressources. RefBase leur propose aussi un système d'alerte qui les informera chaque fois qu'une nouvelle référence a été ajoutée ou modifiée. Un autre service offert aux utilisateurs est l'abonnement aux flux RSS.

L'utilisateur non-enregistré ou "en mode visualisation"

Si la base est publique (en ligne) elle peut être consulté par n'importe quel utilisateur connecté à l'Internet. Il a le droit de faire des recherches, de voir leurs résultats, de visualiser les notices qui l'intéressent et de les exporter. En revanche, il ne peut pas modifier les notices ou ajouter d'autres. Il n'a pas non plus le droit de constituer des bibliographies personnelles ou d'avoir un historique de leurs interrogations. Tout ce qu'il est autorisé à faire est de visualiser les ressources existantes dans la macro-base.

Ajout des notices

Le formulaire de saisie prévoit également les informations classiques et communes à toute base de données bibliographique, mais aussi des champs différents qui ne font qu'augmenter la chance de retrouver l'item souhaité lors d'une recherche : le titre original d'un document traduit, le type de support (disquette, CD, diapositive etc.), le type de travail universitaire (thèse, mémoire, ...), le titre abrégé de journal/revue, les numéros DOI, ISSN et ISBN. A la différence de ce que les autres logiciels présentent dans leurs formulaires, Refbase inclut la plupart des données prévues dans le Cahier des charges de Bosil : tous les types de document qui seront décrits dans la base de données y sont retrouvables, la majorité des champs bibliographiques (mêmes ceux spécifiques aux ressources électroniques) sont également présentes, les champs non bibliographiques comme l'affiliation de l'auteur ou son rôle (sa contribution) sont eux aussi prévus.

L'édition et la reprise de la saisie d'une notice est tout à fait possible : tous les utilisateurs peuvent éditer une référence, mais seuls ceux authentifiés et l'administrateur ont le droit de la modifier.

Parmi les autres fonctionnalités du RefBase on doit noter aussi la possibilité d'indiquer l'URL de la ressource ou d'attacher à une notice le document en texte intégral (en format PDF ou .doc). Cette dernière fonction ne sera pas utilisée par Bosil, d'abord parce que la base de données sera bibliographique, et ensuite, parce que donner accès au texte intégral impliquera les formalités pour le respect du droit d'auteur.

Recherche

Types de recherches

Refbase permet trois types de recherche : simple, avancée et dans le catalogue de la bibliothèque qui héberge la base de donnée concernée. Le dernier type n'est pas utilisé par les autres applications disponibles en ligne.

En utilisant la recherche simple, on peut interroger la base selon plusieurs critères : auteur, titre, année de publication, volume et nombre de pages. La recherche

avancée propose des nombreuses options: plusieurs mots-clés, utilisation des opérateurs booléens, plusieurs champs (en fait, tous les champs sont pris en calcul), etc. Pour les utilisateurs enregistrés Refbase offre un nombre de fonctionnalités que les autres logiciels bibliographiques open source ne possèdent pas : possibilité de sauvegarder la stratégie de recherche, de partager l'historique, de combiner les requêtes précédemment faites.

De plus, Refbase permet d'afficher toutes les notices (alphabétiquement selon le nom d'auteur) pour avoir une image sur l'ensemble de références de la base.

L'utilisateur a également la possibilité de consulter les dernières notices ajoutées : pour le jour courant, le jour avant et pour les sept derniers jours. Cette fonction – disponible depuis la page d'accueil de la base de données - est aussi utile pour les utilisateurs que pour le gestionnaire de la base en lui servant pour faire des statistiques. Refbase permet aussi de consulter (toujours depuis la page d'accueil) les références selon l'année de publication des documents sans avoir besoin de passer par le formulaire de recherche.

Affichage

Avant de soumettre la requête, l'utilisateur a la possibilité de choisir la manière de visualiser les résultats : trois critères (selon un champ précis, choisi par l'utilisateur) et l'ordre (croissant ou décroissant).

Les résultats sont alors affichés sous forme de tableau : chaque colonne correspond à un champ différent, chaque ligne représente une notice. On peut sélectionner une ou plusieurs références dans la liste des résultats et les afficher avec tous leurs détails de description ou en choisissant les champs que l'on veut visualiser.

Affiner la recherche

Aussi bien dans le cas des recherches que du feuilletage, Refbase permet d'affiner la recherche directement à partir de l'écran d'affichage des résultats : si le nombre de occurrences est trop grand, on peut faire une autre interrogation concernant uniquement les résultats obtenus précédemment.

Export / import

Les options d'export et import sont tout à fait fonctionnelles. EndNote est la source et le format principaux. On peut importer, aussi, depuis CSA (Cambridge Scientific Abstracts Internet Database Service).

Pour l'export l'utilisateur doit choisir entre quatre options: EndNote, RIS, BibTeX et MODS (un standard de format bibliographique XML établi par Library of Congress). Le format du document d'export peut être choisi entre HTML et .txt.

Construction du thésaurus sur le thème de la maîtrise de l'information

Un thésaurus est un ensemble organisé de termes représentatifs, appelés « descripteurs », termes qui désignent les concepts propres à un domaine scientifique. Il est à la fois un outil d'indexation et de recherche. Le but principal d'un thésaurus est d'aider l'utilisateur d'une base de données à retrouver l'information.

Premières démarches

Pour la construction d'un thésaurus multilingue (l'intention est d'avoir un thésaurus bilingue : français – anglais), il faut tenir compte de plusieurs choses :

- dans une première instance, voir s'il y a un autre thésaurus ou liste d'autorités dans le domaine concerné ; s'il existe, il faut alors l'adopter et l'adapter à nos besoins ;
- s'il n'y a pas un tel outil, il faut procéder alors à sa conception.

Suite à notre recherche documentaire, nous avons constaté qu'il n'existe pas de thésaurus, ni en français ni dans une autre langue, dédié à la maîtrise de l'information. Cependant, il y a plusieurs vocabulaires contrôlés dans les domaines de l'éducation et des sciences de l'information. Nous les avons retenus pour pouvoir les utiliser lors de la conception proprement dite.

Ensuite la toute première démarche quand on démarre un tel travail est de se procurer les normes et les standards correspondantes, s'ils existaient. Ainsi, nous nous sommes servis de deux standards internationaux (ISO) :

- 2788 : 1986 - pour la construction et la gestion des thésauri monolingues ;
- 5964 : 1985 – pour la construction et la gestion des thésauri multilingues ;

et de la norme AFNOR NF Z47-101 pour « l'établissement des thésaurus multilingues » .

Le contact avec les établissements internationaux

Conformément aux normes nommées ci-dessus, il faut contacter deux institutions internationales pour s'assurer qu'il n'y aura pas une duplication de travaux. Ainsi, avant de commencer à concevoir un thésaurus multilingue, il est obligatoire qu'on annonce son intention aux centres d'information internationaux compétents :

- pour les thésaurus en langue anglaise : Bibliographic Systems Center, School for library Science, Case Western Reserve University, Cleveland, Ohio, USA;
- pour toutes les autres langues : Centralny Instytut Informacji Naukowej i Technicznej Ekonomicznej (CIINTE), Varsovie, Pologne.

Ces deux organismes ont la responsabilité de collecter tous les thésauri dans tous les domaines. Il faudra donc leur envoyer une copie de la première édition et des éditions suivantes de notre thésaurus.

Un premier problème rencontré a été le contact même avec ces établissements. Les normes qui les indiquent sont vieilles respectivement de 19 et 15 ans (il existe une version nouvelle du standard ISO 2788 : 1986, sortie en 2003, mais qui reprend les mêmes erreurs). Entre temps, les choses ont changé. Bibliographic Systems Center (USA) n'a plus cette responsabilité ; aujourd'hui c'est « **Subject Analysis Systems Collection** » - **Faculty of Library and Information Science, University of Toronto, Canada** qui s'en occupe. Egalement, CIINTE (Pologne) n'existe plus ; il a été détruit et ses tâches et ses responsabilités sont assumées aujourd'hui par d'autres organismes, principalement par **Osrodek Przetwarzania Informacji** (« Information Processing Centre »), mais personne ne sait exactement quel organisme s'occupe maintenant des thésauri. Cette nouvelle institution, bien qu'elle ait été contactée plusieurs fois, n'a jamais répondu à nos courriels. En revanche, la Faculté de Science de l'information et des Bibliothèques du Canada nous a fourni les informations dont nous avons eu besoin : ils nous ont confirmé qu'il n'existe pas des thésauri en anglais sur le thème de la maîtrise de l'information et nous en ont indiqué d'autres dans des domaines proches.

Analyses préliminaires

Thésaurus vs. langage naturel

Avant de commencer la conception d'un vocabulaire contrôlé il faut voir s'il est vraiment nécessaire d'en créer un ou, bien, si l'utilisation du langage naturel est préférable. Les deux solutions présentent des avantages et des inconvénients.

Le langage naturel est toujours plus spécifique et précis, plus exhaustif et mis à jour en temps réel ; en plus, il n'implique aucun coût de construction. L'utilisation du langage naturel enlève, aussi, le problème de l'incompatibilité entre deux ou plusieurs bases de données. Mais, il faut tenir compte qu'à tout moment il peut apparaître des erreurs de syntaxe et qu'il y a un grand effort intellectuel placé sur l'utilisateur. D'ailleurs, en utilisant le langage naturel on risque d'avoir un important manque de pertinence lors d'une recherche.

Le langage contrôlé facilite la recherche dans la base en guidant l'utilisateur vers les descripteurs les plus précis à employer, puis vers les meilleurs résultats. Il peut présenter un relative manque de spécificité et d'exhaustivité et il n'est pas immédiatement mis à jour (il faut du temps jusqu'à ce qu'un nouveau terme soit validé et inclus dans le thésaurus). Il peuvent exister, également, des erreurs d'indexation : deux documents avec le même sujet décrits avec des descripteurs différents. Le langage artificiel doit être appris par l'indexeur et par l'utilisateur. C'est, un frein pour un échange facile avec d'autres bases de données : la non utilisation du même thésaurus implique l'incompatibilité entre les bases. En plus, si le taux de consultation de la base de données n'est pas élevé, l'effort d'indexer les documents est inutile ; il vaut mieux alors que les recherches soient faites en langage naturel. Toutefois, si on veut mettre en place un service de push d'informations (DSI, alertes), un thésaurus est absolument nécessaire.

En considérant que le but principal d'une base de données est de donner accès aux ressources et que l'important, dans ce cas, est que l'utilisateur retrouve les documents qui lui sont nécessaires, l'utilisation d'un thésaurus s'est imposée.

Considérations générales

Le thésaurus est bilingue : français – anglais, les deux langues utilisant l’alphabet latin et s’écrivant de gauche à droite. Ces deux langues ont un statut identique dans le thésaurus. Aucune n’est considérée « source » ou « cible », les descripteurs employés proviennent des deux langues en essayant d’établir une relation biunivoque entre les classes d’équivalence.

Domaine

Bien déterminé en avance, le champ disciplinaire (« maîtrise de l’information ») est exact et étroit. « Information literacy » est le terme anglais qui définit le concept ; il n’est pas, en fait, le seul terme existant pour ce domaine, mais le plus utilisé et celui qui s’est imposé dans la littérature. En français, il a plusieurs équivalents, dont aucun ne s’est révélé comme le plus approprié pour couvrir le concept en question ; en général, l’expression « maîtrise de l’information » semble s’imposer.

Les disciplines les plus proches sont les sciences de l’information, la pédagogie, l’informatique.

Le noyau conceptuel est formé par les termes liés aux compétences informationnelles : consultation des bases de données, diffusion de l’information, évaluation de l’information et de la démarche documentaire, instruments et lieux documentaires, présentation et production de l’information, etc.

La périphérie est constituée par les concepts utilisés dans les disciplines proches avec application à la maîtrise de l’information : traitement de l’information, typologie des documents et des sources de documentation, Internet, compétences informatiques, etc.

Types de documents de la base de données

Le type de littérature concernée par la base de données est à prendre en considération, car les périodiques nécessitent une indexation plus approfondie que les thèses ou les actes de congrès. En conséquence, comme la base comprendra aussi des articles de périodiques, le thésaurus devra également être adapté à une indexation complexe.

Quantité de données dans la base

BOSIL ne sera pas une base de données de grandes proportions ; de plus, un thésaurus trop détaillé est coûteux à construire et à maintenir et il faut bien considérer si on en a réellement besoin. Un tel thésaurus sera aussi inefficace dans le mesure où la précision de l'indexation peut être perdue : plus le nombre de termes recouvrant des concepts proches est grand, plus l'indexeur a tendance à ne pas utiliser les mêmes descripteurs pour les documents qui parlent du même sujet ; d'où la manque de pertinence en terme de « silence » lorsqu'on fait une recherche dans la base.

Langue

Bien qu'il s'agisse d'une base de données francophone par excellence, le thésaurus sera bilingue : français et anglais, pour faciliter la recherche dans la base.

Utilisateurs finaux

Les utilisateurs de la base de données sont les usagers finaux - donc les étudiants, les formateurs - ou un intermédiaire, un spécialiste en recherche documentaire délégué à faire les requêtes. Les deux catégories doivent être prises en compte : d'abord parce que les spécialistes en recherche d'information sont eux-mêmes, parfois, mis dans la situation de former les usagers à la maîtrise de l'information et ensuite parce que la base sera en libre accès, et n'importe qui pourra la consulter. Ainsi, il faut penser à un thésaurus bien adapté aux deux catégories.

En conséquence, la terminologie utilisée devrait être la plus proche du langage commun, familière à l'utilisateur qui est peu habitué avec les termes spécifiques documentaires ; mais, dans le cas des concepts spécialisés on garde le terme existant, même s'il n'est pas connu par le grand public ; exemple : « opérateur booléen » ou « OPAC ». Particulièrement pour ce type de cas, il faut prévoir, aussi, la conception d'un glossaire destiné à définir et à expliquer les descripteurs présents dans le thésaurus. Un tel outil est déjà en construction ; il sera disponible depuis la page d'accueil du site FORMIST.

Le squelette du thésaurus

Pour la conception proprement dite du thésaurus nous nous sommes servis de plusieurs outils : d'autres vocabulaires contrôlés dans les disciplines connexes, des dictionnaires, glossaires et tout autre outil terminologique dans les domaines des sciences de l'information, de l'éducation et de l'informatique.

Relations

Le thésaurus a été développé, en ayant comme point de départ les descripteurs déjà utilisés par Formist pour indexer les documents référencés sur le site. Trois types de relations ont été utilisés :

La relation d'équivalence

C'est la relation entre le descripteur choisi et le terme non-préférentiel, marquée par EP (employé pour).

Exemple :

« ANNUAIRE »

EP : « Guide », « Répertoire ».

Mais : « GUIDE »

E : « Annuaire » - pour indiquer le terme correct. « E » = « emploie » (« voir »).

La relation hiérarchique

C'est la relation entre le descripteur et les termes qui lui sont supérieurs ou inférieurs du point de vue conceptuel. Ainsi, il y a deux types de termes : génériques et spécifiques.

Exemple : « DOCUMENTATION »

TG : « Science de l'information »

TS : « Indexation »

La relation associative

C'est la relation entre deux termes proches, mais qui ne sont pas équivalents ou synonymes ; il est important d'indiquer aux utilisateurs les termes associés qu'ils peuvent employer pour élargir leurs requêtes.

Exemple : « CENTRE DE DOCUMENTATION »

TA : « Bibliothèque »

Réalisation et graphique

Une présentation de type index n'est pas utile, parce qu'elle ne montre pas les relations entre les termes et ne peut servir ni à l'indexation, ni à la recherche. C'est pour cette raison que nous avons opté pour le type hiérarchique de présentation : on cherche un terme et le système nous affiche tous les termes connexes (équivalents, associés, génériques et spécifiques).

Ainsi, pour la version imprimée, étant donné qu'il s'agit d'un thésaurus multilingue, nous avons choisi le type alphabétique, pour présenter les descripteurs français en mettant en parallèle leurs équivalents anglais, et en indiquant leurs relations.

Exemple :

MAITRISE DE L'INFORMATION / INFORMATION LITERACY

	Français	Anglais
EP	Compétence documentaire	Bibliographic instruction
	Compétence informationnelle	Computer literacy
	Formation à la maîtrise de l'information	Digital information literacy
	Formation à l'information	Digital literacy
	Formation documentaire	Electronic literacy
	Formation des usagers	Electronic information literacy
	Informer	Hyper-literacy
	IST	Infoliteracy
	Médiation	Informacy
	Modélisation de l'utilisateur	Information competence
	Pédagogie de l'information	Information competence skills
	S'informer	Information competencies
		Information competency
		Information empowerment
		Information fluency
		Information handling skills
		Information literacy competence
		Information literacy competencies
		Information literacy skills
		Information mastery
		Information mediacy
		Information problem solving

		Information problem solving skills Information skills Information technology literacy Internet literacy Library literacy Mediacy Media literacy Network literacy Skills-based literacies Skills of information literacy Research skills
TG	Education Formation Information Recherche	Education Formation Information Research
TS	Analyse documentaire Diffusion de l'information Evaluation de l'information Evaluation de la démarche documentaire Exploitation de l'information Lieux documentaires Recherche d'information Technologies de l'information	Documentary analysis Information dissemination Information evaluation Documentary process evaluation Information management Information places Information retrieval Information technology
TA	Documentation Science de l'information	Documentation Information science

Pour la version électronique, nous avons pensé à deux types de représentation graphique : par arbre et par flèches.

La représentation de type arbre se concrétise dans un système de liens : chaque descripteur a sa fiche où on trouve tous ses relations et ses termes connexes. Ceux-ci renvoient, à l'aide des hyperliens, vers leurs propres fiches. Cette représentation est très pratique à manipuler aussi bien pour l'indexeur que pour l'utilisateur.

La représentation par flèches (Topic map) est plus profonde du point de vue conceptuel et, donc, plus efficace lors de la recherche. A partir d'un descripteur, on navigue à l'intérieur du thésaurus en glissant d'un terme à un autre pouvant visualiser les relations, les connexions, mais aussi les voisinages. La figure 1 sert

d'exemplification. Le type de flèche indique le type de relation. D'habitude, la relation d'équivalence n'est pas présente dans ces graphiques, mais nous avons pensé l'indiquer par le système « mouse over ».

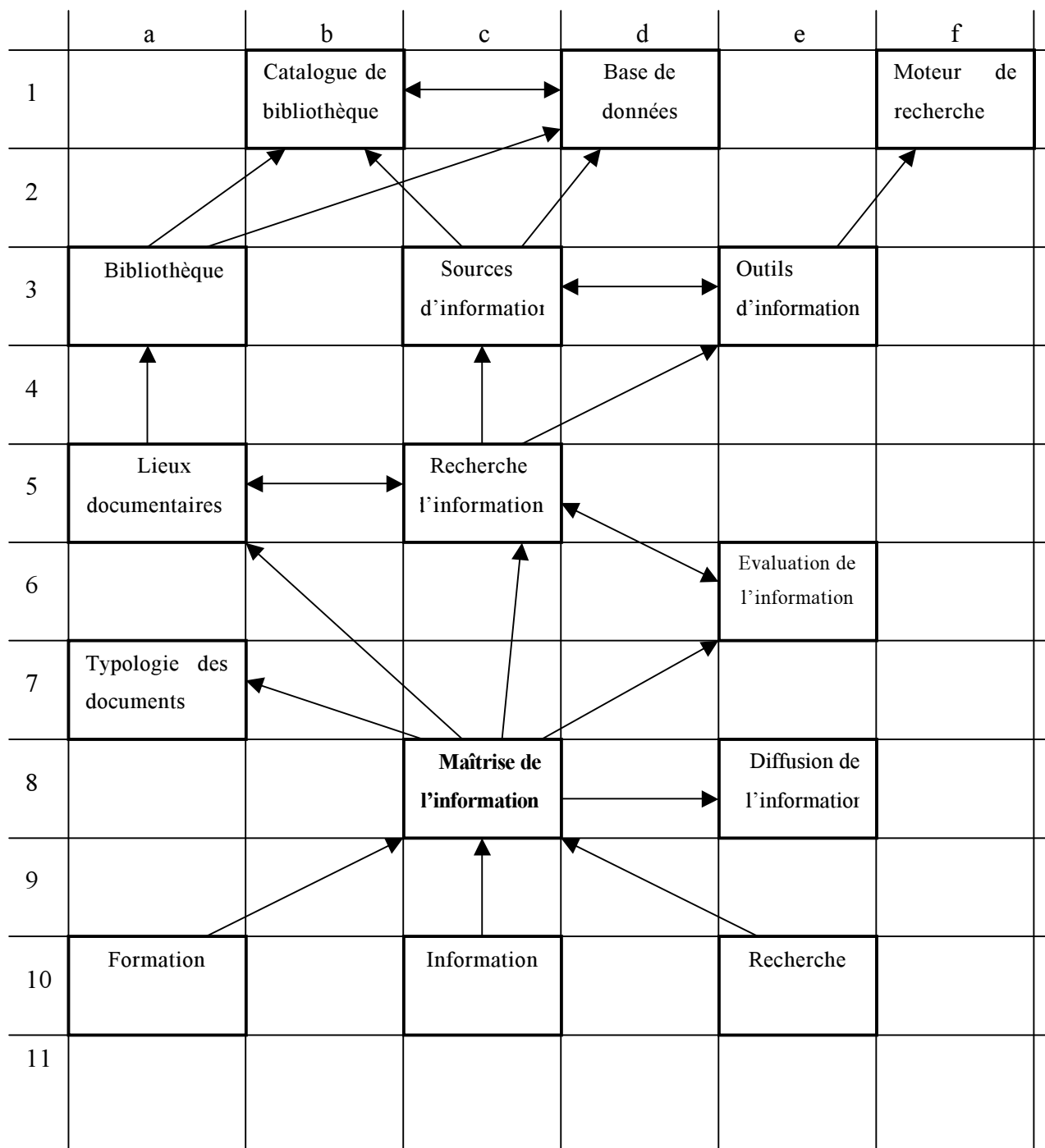


Figure 1 : représentation graphique de type flèche

Conclusion

Il est pour l'instant prématuré de s'exprimer sur l'évolution de la base de données. Une fois la conception physique et l'implémentation terminées la base deviendra tout à fait opérationnelle. Conformément au calendrier établi, ce sera probablement au début de l'été prochain que les utilisateurs auront la chance de la consulter et de tester, en fait, le fruit de nos efforts.

A la fin de ce stage et, en même temps, au début de ma carrière professionnelle, le résultat s'avère positif et nous avons l'espoir que le travail effectué nous servira à l'avenir dans d'autres contextes. Pendant ces quatre mois, nous avons eu la possibilité de mettre en valeur et d'approfondir les connaissances et les compétences acquises au cours de la formation de Master 2 Ride, d'un côté, et de découvrir des nouveaux aspects du segment particulier des bases de données. Tout un autre monde s'est ouvert devant nous : le domaine lui-même de la formation documentaire, l'univers des logiciels libres et open source, les bases de données bibliographiques, les thésauri etc. Et tout ce monde a attiré notre attention et notre curiosité et a eu le rôle bienfaisant de nous former.

Il faut nécessairement évoquer la bonne et enrichissante collaboration avec ma tutrice de stage qui m'a aidée et encouragée tout au long de ce projet. S'il y a eu des embarras, c'est avec son appui qu'ils ont été dépassés. Et malgré ces petits ennuis, dont nous gardons le goût de l'aspect formateur plus que de difficulté, la mission et les objectifs généraux du stage ont été réalisés.

Bibliographie

INTRODUCTION A LA MAÎTRISE DE L'INFORMATION

Association of College & Research Libraries (ACRL). *Norme sur les compétences informationnelles dans l'enseignement supérieur*. Montréal : CREPUQ (Conférence des recteurs et des principaux des universitaires du Québec), 19 août 2004. 11 p.

Cinquièmes Rencontres FORMIST (5 ; 2005 ; Lyon). *Parcours de formation documentaire des étudiants : à qui de jouer ? Développer les compétences informationnelles dans un cursus disciplinaire*.

Former les étudiants à la maîtrise de l'information : repères pour l'élaboration d'un programme. [en ligne] Paris : Ministère de l'Éducation Nationale, de la Recherche et de la Technologie, 2005. [consulté le 21 juillet 2005] Disponible sur : <http://www.enssib.fr/repere>

Site FORMIST. [en ligne] Villeurbanne : Enssib, 2005. [consulté le 20 septembre 2005] Disponible sur : <http://formist.enssib.fr>

WEBBER, Sheila ; BOON, Stuart. *British academics from different disciplines : comparing their conception of information literacy*. Cinquièmes Rencontres FORMIST. Lyon, 9 juin 2005.

ETUDES DE FAISABILITE

CHAPMAN, Ann. *Developing a National Database of Resources in Accessible Formats : a feasibility study*. Londres : British Library Cooperation and Partnership Programme, 2000. 117 p. [consulté le 20 mai 2005] Disponible sur : <http://www.ukoln.ac.uk/services/lic/sharethevision/fsed.pdf>

DALLOT-BEFIO, Yalita Y. *Etude de faisabilité d'une base de données des publications des chercheurs de l'INSA de Lyon : Rapport de stage*. [en ligne] Villeurbanne : Enssib, 1998. 80 p. [consulté le 20 mai 2005] Disponible sur : <http://www.enssib.fr/bibliotheque/documents/dessid/dallotbefio.pdf>

GLOSSIOTI, Maria. *Feasibility study concerning the Internet alternative for the database*. Information Society Technologies & Statistical Atlas of the European Union, 2002. 26 p. [consulté le 20 mai] Disponible sur : http://www.statlas.org/project/deliverables/wp2/STL_GSDB_%20D2.3.pdf

CONCEPTION D'UNE BASE DE DONNEES

BAUER, Marie-José. *Création d'une base de données de notices bibliographiques pour les Filières Universitaires Francophones. Formation des chercheurs à l'utilisation des Nouvelles Technologies d'Information et de Communication : Rapport de stage*. Villeurbanne : Enssib, 2000. 59 p. [consulté le 20 mai 2005] Disponible sur : <http://www.enssib.fr/bibliotheque/documents/dessid/rsbauer.pdf>

GARDARIN, Georges. *Internet / Intranet et bases de données : Data Web, Data Media, Data Warehouse, Data Mining*. Paris : Eyrolles, 1999. 249 p.

GARDARIN, Georges. *Bases de données : objet & relationnel*. Paris : Eyrolles, 2001. 790 p.

STEPHENS, Ryan K ; PLEW, Ronald R. *Conception de bases de données*. Paris : CampusPress, 2003. 526 p.

VAN-HUNG, Rayane. *Conception et mise en place d'une base de données : Rapport de stage*. Villeurbanne : Enssib, 2000. 84 p. . [consulté le 20 mai 2005]
Disponible sur : <http://www.enssib.fr/bibliotheque/documents/dessid/rsrayane.pdf>

OPEN SOURCE

BRETTHAUER, David. "Open source : a history". In: *Information Technology and Libraries*. Vol. 21, n°1, 2002, p 3-10. ISSN 0730-9295

COYLE, Karen. "Open source, Open standards". In: *Information Technology and Libraries*. Vol. 21, n°1, 2002, p 33-36. ISSN 0730-9295

FONTAINE, Alain. *Développement open source avec Windows*. Paris : Dunod, 2003. 322 p.

CONSTRUCTION THESAURUS

AITCHISON, Jean ; GILCHRIST, Alan ; BAWDEN, David. *Thesaurus construction and use: a practical manual*. London: The Association for Information management (Aslib), 1997. 212 p.

Association française pour la normalisation. *Principes directeurs pour l'établissement des thésaurus multilingues : NF Z 47-101. Norme française homologuée par décision du Directeur Général de l'AFNOR le 20 novembre 1990 pour prendre effet le 20 décembre 1990*. Paris : AFNOR, décembre 1990 ISSN 0335-3931

CACALY, Serge. *Dictionnaire de l'information*. Paris : Armand Colin, 2004. Collection « Coursus ». 278 p.

HUDON, Michèle. *Le thésaurus : conception, élaboration, gestion*. Montréal : Editions ASTED, 1995. Collection « Clé en main ». 222 p.

International Standards Organisation. *Guidelines for the Construction, Format, and Management of Monolingual Thesauri : Z 39.19*. (Revision of Z39.19-1980) Bethesda, Maryland: NISO Press, 28 August 2003. ISSN: 1041-5653.

International Standards Organisation. *Guidelines for the Construction, Format, and Management of Multilingual Thesauri : 5964 :1985*. Bethesda, Maryland: NISO Press, 1985.

REITZ, Joan M. *ODLIS - Online Dictionary for Library and Information Science*. [En ligne] Libraries Unlimites, 2004. 8000 p.
Disponible sur : <http://lu.com/odlis>

Table des annexes

ANNEXE 1 : DEFINITION DES CHAMPS.....	49
ANNEXE 2 : BESOINS DE L'ACTIVITE	57
ANNEXE 3 : MODELE CONCEPTUEL DE DONNEES	63
ANNEXE 4 : MODELE LOGIQUE DE DONNEES ... ERREUR ! SIGNET NON DEFINI.	

Annexe 1 : Définition des champs

La table DOCUMENT

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	ID_doc	1	Texte (code: chiffres et éventuellement lettres)	Incrémentation automatique	Numéro de la notice : identificateur global unique.	Clé primaire qui lie la table DOCUMENT des tables désignant les types de document.
2	Type de document	1	Vocabulaire	Choix dans liste	<ul style="list-style-type: none"> • Ouvrage (monographie) • Chapitre d'un ouvrage • Périodique (toute publication en série) • Article • Thèses et mémoires • Rapport officiel • Acte de colloque 	
3	Titre propre	1	Texte	Saisie	Nom du document référencé	A saisir le titre dans son intégralité suivi du complément du titre s'il y a lieu.

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
4	Titre parallèle	1	Texte	Saisie	Nom du document au cas où celui ci est bilingue	Il ne sera affiché aux utilisateurs qu'au moment où il est rempli.
5	Titre original	1	Texte	Saisie	Nom du document au cas où celui ci est une traduction.	Il ne sera affiché aux utilisateurs qu'au moment où il est rempli.
6	Support	1	Texte	Choix dans liste	- [imprimé] - [en ligne]	Champ facultatif au cas des documents imprimés, mais obligatoire pour les ressources électroniques.
7	Num_auteur	1	Texte	Incrémementation automatique	La mention de responsabilité.	Clé primaire qui lie les tables DOCUMENT et AUTEUR. Pour la forme à respecter RAMEAU. Plusieurs auteurs : séparés par point virgule. Plus de trois : saisir le premier suivi de la mention « et al. »
8	Edition	1	Texte	Saisie		Il ne sera affiché aux utilisateurs qu'au moment où il est rempli.
9	Lieu d'édition	3	Texte	Saisie	La ville où le document a été publié.	
10	Pages	1	Texte (chiffres, lettres, caractères non-alphabétiques)	Saisie		A saisir le nombre total de pages. Pour un article ou chapitre d'ouvrage : les pages de début et de fin séparées par un tiret.
11	Bibliographie	1	Oui / Non	Case à cocher		Zone de la collation
12	Index	1	Oui / Non	Case à cocher		Zone de la collation
13	Illustrations	1	Oui / Non	Case à cocher		Zone de la collation
14	Glossaire	1	Oui / Non	Case à cocher		Zone de la collation
15	Tableaux	1	Oui / Non	Case à cocher		Zone de la collation
16	Tableaux chronologiques	1	Oui / Non	Case à cocher		Zone de la collation
17	Lexique	1	Oui / Non	Case à cocher		Zone de la collation
17	Graphiques	1	Oui / Non	Case à cocher		Zone de la collation
18	Annexes	1	Oui / Non	Case à cocher		Zone de la collation
19	Photos	1	Oui / Non	Case à cocher		Zone de la collation
20	Notes	1	Oui / Non	Case à cocher		Zone de la collation

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
21	Cartes		Oui / Non	Case à cocher		Zone de la collation
22	Figures	1	Oui / Non	Case à cocher		Zone de la collation
23	Planches	1	Oui / Non	Case à cocher		Zone de la collation
24	Langue	10	Texte	Choix dans liste Français par défaut Anglais, autres.	La langue du document	La langue va être saisie entièrement (pas d'abréviations arbitraires).
25	Date	1	Date/Time	Saisie	L'année de publication du document	Au cas de travaux universitaires : la date de soutenance, si le document n'est pas publié.
26	URL	1	Hyperlink	Saisie	Localisation d'un document électronique	Champ vide au cas des ressources non électroniques
27	URL Formist	1	Hyperlink	Saisie	Localisation de la fiche descriptive au sein du site Formist	Champ vide au cas d'un ressource qui n'est pas signalée sur le site.
28	Descripteurs français	25	Texte	Choix dans liste (lien vers le thésaurus : incrémentation directe par sélection)	Termes choisi d'un vocabulaire contrôlé (un thésaurus) décrivant le sujet du document.	Indexation
29	Descripteurs anglais	25	Texte	Choix dans liste (lien vers le thésaurus : incrémentation directe par sélection)	Termes choisi d'un vocabulaire contrôlé (un thésaurus) décrivant le sujet du document.	Indexation
30	Résumé	1	Texte long	Saisie	Texte à valeur ajouté reprenant les idées essentielles du document.	Indexation
31	Date création notice	1	Date/Time	Automatique	Date de la création de la notice	
32	Date MAJ	1	Date/Time	Automatique	Date de la mise à jour de la notice	
33	ID_Créateur	1	Texte	Incrémentation automatique	Numéro identificateur de la personne qui a saisi la notice.	Clé primaire qui lie les tables DOCUMENT et CREATEUR.
34	Num_proposition	10	Texte	Incrémentation automatique	Numéro de la proposition.	Clé primaire qui lie les tables DOCUMENT et PROPOSITION.

La table AUTEUR

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1.	Num_auteur	1	Texte	Incrémentation	La mention de responsabilité.	Clé primaire qui lie les tables

				automatique		DOCUMENT et AUTEUR. Pour la forme à respecter RAMEAU. Plusieurs auteurs : séparés par point virgule. Plus de trois : saisir le premier suivi de la mention « et al. »
2	Nom	3	Texte	Saisie/Choix dans liste (de l'index).		A saisir en majuscules, suivi d'une virgule.
3	Prénom	3	Texte	Saisie/Automatique		A saisir en minuscules.
4	Affiliation	3	Texte	Saisie/Automatique		
5	Contribution	10	Texte	Choix dans liste	Le rôle de chaque contributeur : - Traducteur - Réalisateur - Illustrateur - Préfacier - Postfacier - Editeur intellectuel - Editeur scientifique - Organisateur - Intervenant - Compileur - Rédacteur - Graveur - Coordinateur - Directeur de thèse - Responsable pédagogique - Tuteur - Photographe	Au cas de l'auteur proprement dit ce champ ne sera pas rempli.

La table OUVRAGE

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	ID_doc	1	Texte (code: chiffres et éventuellement lettres)	Incrémentation automatique	Numéro de la notice : identificateur global unique.	Clé primaire qui lie les tables DOCUMENT et OUVRAGE.
2	Maison d'édition	3	Texte	Saisie	L'éditeur du document	
3	ISBN	1	Numéro	Saisie	Le numéro ISBN de l'ouvrage.	

La table CHAPITRE OUVRAGE

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	ID_doc	1	Texte (code: chiffres et éventuellement lettres)	Incrémentation automatique	Numéro de la notice : identificateur global unique.	Clé primaire qui lie les tables CHAPITRE D'OUVRAGE et OUVRAGE.
2	Num_auteur	1	Texte	Incrémentation automatique	La mention de responsabilité.	Clé primaire qui lie les tables CHAPITRE D'OUVRAGE et AUTEUR. Pour la forme à respecter RAMEAU. Plusieurs auteurs : séparés par point virgule. Plus de trois : saisir le premier suivi de la mention « et al. »
3	Titre chapitre	1	Texte	Saisie	Titre du chapitre	

La table PERIODIQUE

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	ID_doc	1	Texte (code: chiffres et éventuellement lettres)	Incrémentation automatique	Numéro de la notice : identificateur global unique.	Clé primaire qui lie les tables DOCUMENT et PERIODIQUE.
2	Editeur	3	Texte	Saisie/Choix dans liste (de l'index).	L'éditeur de la revue.	
3	Périodicité	1	Texte	Choix dans liste prédéfinie	- quotidien - hebdomadaire - mensuelle - bimensuelle - trimestrielle - semestrielle - biannuelle - annuelle	
4	ISSN	1	Numéro	Saisie	Le numéro ISSN de la publication sériale.	

La table ARTICLE

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	ID_doc	1	Texte (code: chiffres et éventuellement lettres)	Incrémentation automatique	Numéro de la notice : identificateur global unique.	Clé primaire qui lie les tables ARTICLE et PERIODIQUE.
2	Titre article	1	Texte	Saisie	Titre de l'article	
3	Volume	1	Numéro	Saisie	Le volume de la revue dont l'article fait partie.	
4	Numéro	1	Numéro	Saisie	Le numéro de la revue dont l'article fait partie.	

La table COLLOQUE

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	ID_doc	1	Texte (code: chiffres et éventuellement lettres)	Incrémentation automatique	Numéro de la notice : identificateur global unique.	Clé primaire qui lie les tables DOCUMENT et COLLOQUE.
2	N° du colloque	1	Numéro	Saisie		
3	Date du colloque	1	Date/Time	Saisie	La date quand le colloque a été organisé.	Ce n'est pas la date de publication !
4	Lieu du colloque	1	Texte	Saisie	Lieu où le colloque a eu place.	Ce n'est pas le lieu d'édition !
5	Maison d'édition	1	Texte	Saisie	L'éditeur du document, si l'acte de colloque a été publié.	

La table THESES

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	ID_doc	1	Texte (code: chiffres et éventuellement lettres)	Incrémentation automatique	Numéro de la notice : identificateur global unique.	Clé primaire qui lie les tables DOCUMENT et THESES
2	Type de thèse	1	Texte	Choix dans liste		
3	Discipline	1	Texte	Saisie		

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
4	Nom_université	1	Texte	Saisie	Le nom de l'université de soutenance.	
5	N° officiel de la thèse	1	Numéro	Saisie		

La table RAPPORT OFFICIEL

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	ID_doc	1	Texte (code: chiffres et éventuellement lettres)	Incrémentation automatique	Numéro de la notice : identificateur global unique.	Clé primaire qui lie les tables DOCUMENT et THESES
2	Commanditaire	1	Texte	Saisie	Personne (physique ou juridique) pour laquelle ou au nom de laquelle le rapport a été réalisé	
3	Maison d'édition	1	Texte	Saisie		

La table CREATEUR

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	ID_Créateur	1	Texte	Incrémentation automatique	Numéro identificateur de la personne qui a saisi la notice.	Clé primaire qui lie les tables DOCUMENT et CREATEUR.
2	Nom	1	Texte	Choix dans liste	Nom de la personne qui a écrit la notice.	
3	Prénom	1	Texte	Automatique		
4	Affiliation	1	Texte	Automatique	La structure où il travaille (Formist, Urfist, Bibliothèque, ...)	

La table PROPOSITION

N°	Nom champ	Répétable	Type de donnée	Alimentation	Description	Remarques
1	Num_proposition	10	Texte	Incrémentation automatique	Numéro de la proposition.	Clé primaire qui lie les tables DOCUMENT et PROPOSITION.
2	Nom	1	Texte	Automatique	Nom de la personne qui a proposé le document.	
3	Prénom	1	Texte	Automatique		
4	Société	1	Texte	Automatique		Champ facultatif (si la personne qui a proposé le document l'a rempli).

Annexe 2 : Besoins de l'activité

A. Ergonomie :

- interopérabilité Formist : logiciel open source sous MySQL / PostgreSQL, PHP et Apache qui sont déjà installés pour le site web de Formist;
- disponibilité en ligne sur Internet de la base ;
- langue de l'interface : français ou anglais ;
- interface graphique conviviale et facile à utiliser ;
- système d'authentification sécurisé qui devra permettre de savoir qui a modifié la base, quand et quelles sont les modifications apportées ;
- naviguer soit avec les options du menu, soit avec des raccourcis des touches du clavier ;
- sauvegarde automatique des enregistrements une fois la saisie d'une notice terminée ;
- multi-fenêtrage.

B. Gestion utilisateurs :

Le logiciel choisi devra permettre le travail collaboratif et, en conséquence, qu'il soit multi-utilisateurs. On devrait avoir donc trois types de profils : l'administrateur, l'utilisateur enregistré et l'utilisateur non enregistré.

L'administrateur a tous les droits sur la base :

- gérer la base de données : avoir accès au code et pouvoir le modifier pour mieux structurer la base : ajouter des champs, modifier les particularités des tables etc.

- gérer les listes bibliographiques thématiques : créer, alimenter, modifier, supprimer (droit exclusif de supprimer une notice) ;
 - gérer les comptes utilisateur : créer, supprimer, modifier ;
 - être alerté chaque fois qu'une notice a été créée ou modifiée par un utilisateur (enregistré);
 - créer des styles bibliographiques autres que celles offertes par défaut par le logiciel ;
- On devrait avoir une liste avec tous les partenaires de la base, visible uniquement par l'administrateur. Cette liste inclura une brève fiche descriptive personnelle (nom, prénom, affiliation) ; en cliquant sur l'un d'eux, on devrait avoir tous ses éléments d'identification : nom, prénom, fonction, société, date où il est devenu membre, identifiant et mot de passe et son activité dans le cadre de notre base de données : nombre de notices ajoutées ou modifiées dans la base centrale, nombre des listes personnelles ou de groupe.

Les utilisateurs enregistrés (aussi appelés « les partenaires ») ont plusieurs droits et privilèges:

- accéder à la base de données commune: consulter, rechercher, visualiser les résultats ;
- ajouter, éditer et modifier des notices ;
- créer ses propres bibliographies, dans des dossiers séparés personnels, en sélectionnant des ressources qui les intéressent dans la macro-base ou en rajoutant d'autres, externes;
- avoir un style bibliographique propre ;
- modifier le format de saisie des notices dans le cadre de ses listes personnelles, en ayant la possibilité d'ajouter d'autres champs et des annotations personnelles (ses propres mots-clés ou notes), qui ne vont pas apparaître dans la base centrale, commune ;
- faire des recherches dans ses propres listes ;
- visualiser et rechercher dans les listes personnelles d'un autre utilisateur (enregistré);
- faire un import des données depuis une base locale ou à distance ;
- exporter des données dans différents formats et avoir un format personnel d'export;

- imprimer les notices sélectionnées sans avoir besoin de les exporter avant ;
- envoyer les références sélectionnées par mail ;
- sauvegarder de façon permanente et partageable l'historique des recherches faites par chacun d'entre eux ;
- former des groupes (thématiques ou de travail) pour partager les ressources ;
- être alerté chaque fois qu'une nouvelle référence a été ajoutée ou modifiée dans une liste particulière de la base centrale ou dans une liste partagée avec d'autres utilisateurs (enregistrés) ;
- s'inscrire à la DSI (Dissémination sélective de l'information) pour recevoir des emails-alertes chaque fois quand une nouvelle notice correspondant à ses requêtes a été ajoutée à la base.

Ils n'ont pas le droit de :

- modifier le code source de la base;
- supprimer une notice ou une liste commune ;
- modifier ou supprimer les listes bibliographiques des autres utilisateurs (enregistrés) ;
- ajouter le texte intégral à une notice, ni dans la base centrale, ni dans ses propres listes ;
- voir l'activité des autres partenaires.

Les utilisateurs non-enregistrés ou "en mode visualisation" ont les droits suivants :

- faire des recherches dans la macro-base et visualiser les résultats ;
- sélectionner et visualiser les notices qui les intéressent ;
- exporter les résultats dans un format prédéfini ou personnel ;
- imprimer les notices sélectionnées sans avoir besoin de les exporter avant ;
- envoyer les références sélectionnées par mail ;
- avoir un historique de leurs recherches pendant une session ;
- s'inscrire à la DSI ;
- proposer un document.

Ils ne peuvent pas s'authentifier et n'ont pas le droit de :

- modifier la base centrale ;
- avoir des listes personnelles ;

- avoir accès aux listes personnelles des utilisateurs enregistrés ;
- éditer une notice ;
- avoir un style bibliographique personnel ;
- sauvegarder le format personnel d'export.

C. Alimentation de la base :

- formulaire de saisie différent en fonction du type de document ;
- éditer et reprendre la saisie d'une notice ;
- repérer les doublons et envoyer un message d'alerte lorsqu'on est en train d'introduire une notice déjà existante dans la base ;
- import de différentes bases de données locales (EndNote ...) ou distantes (CSA, Francis, Pascal ...) ;
- utilisation de l'Unicode ;
- gérer les propositions faites par les utilisateurs non-enregistrés :
 - formulaire de saisie de la proposition avec une brève description physique du document ;
 - alerte email à l'administrateur chaque fois qu'une proposition a été faite ;
 - possibilité de reprendre les données de la proposition (après réception et validation) en ajoutant les autres pour compléter la description ;
- pouvoir indexer les ressources.

D. Indexation des ressources

- avoir des index alphabétiques pour les noms d'auteur, les éditeurs, les titres de revue (tout ce qui est répétable) ;
- introduire automatiquement les données de l'index dans le formulaire de saisie sans avoir besoin de les taper encore une fois ;
- repérer les doublons (de l'index et du thésaurus) et les enlever ;
- gérer des listes hiérarchiques multilingues : intégrer un thésaurus bilingue accessible au moment de l'alimentation et de la recherche;

- pouvoir naviguer et faire des recherches dans la base à partir d'un ou plusieurs termes du thésaurus ;
- pouvoir faire des imports d'autorités et de termes.

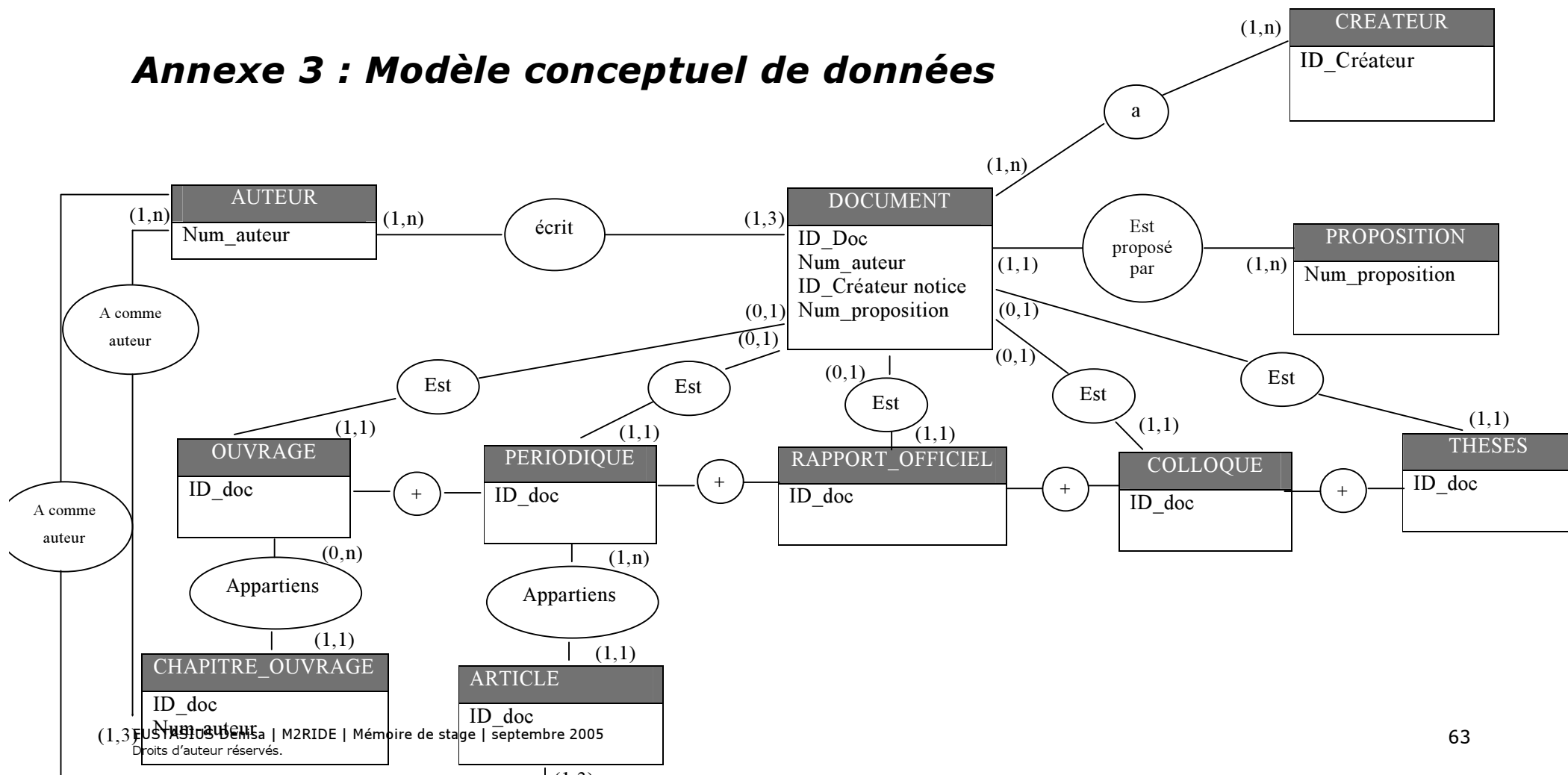
E. Recherche dans la base :

- affichage général des listes : naviguer dans les enregistrements selon plusieurs critères : par catégories thématiques, par auteurs, par type de document, par date de publication, par date de création de la notice ;
- au moins deux niveaux de recherche : simple, pour les utilisateurs moins habitués à la recherche d'information, et complexe, pour les professionnels ; le deuxième niveau devrait donner la possibilité d'utiliser tous les moyens de recherche :
 - recherche booléenne ;
 - multi-champs;
 - par type de document ;
 - chercher un ou plusieurs mots ou une expression exacte ;
 - permettre la troncature au début, à la fin et à l'intérieur d'un mot;
 - grouper les termes avec des parenthèses ;
 - combiner plusieurs requêtes précédemment faites ;
 - sauvegarde de la stratégie de recherche ;
 - possibilité de respecter ou non les casses.
- supporter les caractères accentués français ou d'autres caractères spéciaux en fonction de la langue utilisée ;
- afficher les résultats selon des critères établis par l'utilisateur :
 - afficher un nombre précis de notices par page ;
 - afficher uniquement les champs souhaités (à choisir dans une liste);
 - afficher en fonction de : nom auteur, titre, année de publication, date de création de la notice ou de mise à jour.
 - ordre croissant ou décroissant.
- visualiser une référence particulière sélectionnée dans la liste des résultats ;
- rechercher directement dans les bases de données locales ou distantes via l'interface du logiciel pour faciliter l'import de notices.

F. Statistiques

- gérer les statistiques concernant les ressources : taux de l'alimentation (nombre de notices saisies dans une période de temps donnée), nombre total des références dans la base centrale, taux de propositions (nombre total des propositions, nombre des propositions validés) ;
- gérer les statistiques concernant le taux de consultation de la base : nombre d'utilisateurs enregistrés, nombre d'utilisateurs inscrits à la DSI, nombre de listes personnelles ou de group, nombre de consultation par jour ou dans une période donnée de temps (pour les utilisateurs non enregistrés) ;
- gérer les statistiques concernant les recherches (ranking) : les requêtes les plus fréquentes, les auteurs / les titres les plus recherchés.

Annexe 3 : Modèle conceptuel de données



DOCUMENT (ID_doc, Type du document, Titre propre, Titre parallèle, Titre original, Support, Num_auteur, Edition, Lieu d'édition, Pages, Bibliographie, Index, Illustrations, Glossaire, Tableaux, Tableaux chronologiques, Lexique, Graphiques, Annexes, Photos, Notes, Cartes, Figures, Planches, Langue, Date, URL, URL Formist, Descripteurs français, Descripteurs anglais, Résumé, Date création notice, Date MAJ, ID_Créateur, Num_proposition)

AUTEUR (Num_auteur, Nom, Prénom, Affiliation (société), Affiliation (adresse), Contribution)

OUVRAGE (ID_doc, Maison d'édition, ISBN)

CHAPITRE_OUVRAGE (ID_doc, Nom auteur, Titre chapitre)

PERIODIQUE (ID_doc, Editeur, Périodicité, ISSN)

ARTICLE (ID_doc, Titre article, Volume, Numéro)

COLLOQUE (ID_doc, N° du colloque, Date du colloque, Lieu du colloque, Maison d'édition)

THESES (ID_doc, Type de thèse, Discipline, Nom_université, N° officiel de la thèse)

RAPPORTS OFFICIELS (ID_doc, Commanditaire, Maison d'édition)

CREATEUR (ID_Créateur, Nom, Prénom, Affiliation)

PROPOSITION (Num_proposition, Nom, Prénom, Société).

Annexe 4 : Modèle logique de données

