



HAL
open science

**Réalisation d'un serveur d'investigation iconographique
et bibliographique sur le thème de "l'Eau en Allemagne"
pour le Festival International de Géographie 2003 de
Saint-Dié-des-Vosges (02-05 octobre 2003)**

Brigitte Parant

► **To cite this version:**

Brigitte Parant. Réalisation d'un serveur d'investigation iconographique et bibliographique sur le thème de "l'Eau en Allemagne" pour le Festival International de Géographie 2003 de Saint-Dié-des-Vosges (02-05 octobre 2003). domain_shs.info.bibl. 2003. mem_00000055

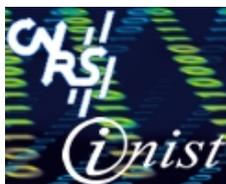
HAL Id: mem_00000055

https://memic.ccsd.cnrs.fr/mem_00000055

Submitted on 21 Jan 2004

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.



Université Henri Poincaré
Université Nancy 2
Institut National Polytechnique de Lorraine

DESS Information Scientifique et Technique - Intelligence Economique
Promotion 2002-2003

RAPPORT DE STAGE

**Réalisation d'un serveur d'investigation
iconographique et bibliographique
sur le thème de « l'Eau en Allemagne »
pour le Festival International de Géographie 2003
de Saint-Dié-des-Vosges**

(02-05 octobre 2003)

Stage effectué à l'INIST-CNRS du 5 mai au 5 septembre 2003

par

Brigitte PARANT

Maître de Stage :
Laurence RAGEOT

Rapport soutenu :
le 25 septembre 2003

REMERCIEMENTS

Mes remerciements sont adressés aux différentes personnes de *l'équipe-projet (DPS-SHS) de l'INIST* qui ont accepté de m'entourer pendant toute la durée de mon stage :

- Jacques DUCLOY, responsable du DPS (Département Produits et Services) ;
- Laurence RAGEOT, responsable du service SHS (Sciences Humaines et Sociales) ;
- Bernard SAMPITE, ingénieur-documentaliste ;
- Philippe HOUDRY, chef de projet ;
- François PARMENTIER, ingénieur de recherche ;
- Solveig VIDAL, ingénieur-documentaliste.

et Dolorès DARDAINE, responsable du site portail INIST (ConnectSciences), qui a apporté un regard critique sur la navigation et l'ergonomie des serveurs d'investigation.

Je remercie également *l'équipe de PRODIG*, et plus particulièrement :

- Claire LIST, responsable du pôle de documentation de PRODIG ;
- Gaëlle HALLAIR, ingénieur-documentaliste.

Enfin, je tiens à adresser tous mes remerciements à ma famille, mon époux et mes trois grands garçons qui depuis quelques mois m'ont soutenue et encouragée.

SOMMAIRE

| | |
|---|----|
| REMERCIEMENTS | 2 |
| SOMMAIRE | 3 |
| TABLE DES ILLUSTRATIONS | 5 |
| INTRODUCTION..... | 6 |
| 1. PRESENTATION DE L'INIST..... | 7 |
| 1.1. Présentation de l'organisme d'accueil..... | 7 |
| 1.1.1 Qui ? | 7 |
| 1.1.2 Fait quoi ? Avec quoi ? | 7 |
| 1.1.2.1 Les services INIST sur Internet..... | 8 |
| 1.1.2.2 Les services de veille scientifique et technologique | 9 |
| 1.1.3 Pour qui ? | 10 |
| 1.1.4 Avec qui ?..... | 10 |
| 1.1.4.1 Les partenaires..... | 10 |
| 1.1.4.2 Le personnel de l'INIST..... | 10 |
| 1.1.5 Avec quels moyens ?..... | 11 |
| 1.1.6 Avec quels effets ? | 11 |
| 1.2. Présentation des équipes du CNRS impliquées dans la mission..... | 11 |
| 2. PRESENTATION DE LA MISSION | 13 |
| 2.1 Origine et définition de la mission | 13 |
| 2.1.1 Origine de la mission..... | 13 |
| 2.1.1 Définition de la mission | 13 |
| 2.2 Notion de « serveur d'investigation » | 13 |
| 2.2.1 Etymologie de l'expression | 13 |
| 2.2.2 Définitions spécialisées | 14 |
| 2.3 Objectifs de la mission | 15 |
| 2.4 Etapes de la mission | 15 |
| 2.4.1 Analyse de l'existant | 15 |
| 2.4.2 Analyse des besoins | 16 |
| 2.4.3 Appropriation des outils et technologies..... | 16 |
| 2.4.4 Déroulement de la mission..... | 17 |
| 3. CONCEPTION DU SERVEUR D'INVESTIGATION : « FIG 2003 »..... | 18 |
| 3.1 Phase expérimentale..... | 18 |
| 3.1.1 Prise de connaissance de l'environnement DILIB | 18 |
| 3.1.2 Première approche du serveur d'investigation BIBAN..... | 19 |
| 3.2 Serveur iconographique..... | 20 |
| 3.2.1 L'indexation des images..... | 20 |
| 3.2.2 Le droit de l'image | 20 |
| 3.2.3 Création d'un modèle de notice | 21 |

| | | |
|-------|--|----|
| 3.2.4 | Reformatage des notices « images » | 22 |
| 3.2.5 | Traitement des photographies | 25 |
| 3.2.6 | Création d'un serveur d'investigation de photographies..... | 26 |
| 3.3 | Serveurs bibliographiques | 26 |
| 3.3.1 | Constitution d'un corpus de notices bibliographiques avec MIRIAD | 27 |
| 3.3.2 | Déchargement des corpus et constitution des serveurs bibliographiques | 28 |
| 4. | REALISATION DU SERVEUR D'INVESTIGATION « FIG 2003 » | 29 |
| 4.1 | Arborescence du serveur d'investigation « FIG 2003 » | 29 |
| 4.2 | Mise en place d'un plan de classement | 30 |
| 4.3 | Création d'un thésaurus..... | 32 |
| 4.4 | Création du serveur d'investigation « BIBALEau » | 33 |
| 4.5 | Création d'un dictionnaire..... | 33 |
| 5. | FONCTIONNEMENT DU SERVEUR « BIBALEau » | 34 |
| 5.1 | Présentation des photographies | 34 |
| 5.2 | Fonctionnement des serveurs d'investigation | 35 |
| 5.2.1 | Le serveur iconographique | 35 |
| 5.2.2 | Le serveur de références bibliographiques..... | 38 |
| 5.2.3 | Le serveur commun - multibase | 39 |
| 6. | BILAN DE LA MISSION..... | 40 |
| 6.1 | Difficultés rencontrées | 40 |
| 6.2 | Les résultats obtenus | 41 |
| 6.3 | Les limites du projet..... | 41 |
| 6.4 | Les améliorations possibles..... | 41 |
| 6.5 | Les perspectives à court terme | 42 |
| | CONCLUSION | 43 |
| | GLOSSAIRE..... | 46 |
| | BIBLIOGRAPHIE | 49 |
| | TABLE DES ANNEXES : | 51 |
| | ANNEXE 1 : Bibliographie sélective sur le classement et l'indexation des | 52 |
| | ANNEXE 2 : Etapes de création du serveur DILIB..... | 56 |
| | ANNEXE 3 : Echange de courrier avec le partenaire PRODIG | 57 |
| | ANNEXE 4 : Page d'Accueil « BIBALEau »..... | 58 |
| | ANNEXE 5 : Problèmes rencontrés sur les mots-clés les plus fréquents | 59 |

TABLE DES ILLUSTRATIONS

| | |
|--|----|
| Fig. 1 Répartition par domaine des 26 000 périodiques du fonds documentaire de l'INIST.... | 8 |
| Fig. 2 Constitution d'un serveur d'investigation | 14 |
| Fig. 3 Architecture générale d'un serveur | 18 |
| Fig. 4 Fichier modèle Indexation | 21 |
| Fig. 5 Exemple : Notice d'indexation de la photographie n°1 – phot01 | 21 |
| Fig. 6 Programme en LEX qui prépare le fichier indexation pour sa conversion en RDF/XML | 22 |
| Fig. 7 Construction d'arbre en XML..... | 23 |
| Fig. 8 Programme en C qui convertit en RDF/XML..... | 25 |
| Fig. 9 Plate-forme MIRIAD | 27 |
| Fig. 10 Descripteurs français les plus fréquents..... | 28 |
| Fig. 11 Arborescence des applications..... | 29 |
| Fig. 12 Plan de classement élaboré à partir du plan de classement de la base de données FRANCIS pour l'application geo..... | 30 |
| Fig. 13 Extrait du fichier <i>bgi.class.sgml</i> utilisé pour insérer le plan de classement dans le serveur | 31 |
| Fig. 14 Extrait du fichier <i>geo.FD.thes.ed</i> utilisé pour insérer le thésaurus dans le serveur.... | 32 |
| Fig. 15 Extrait du dictionnaire français : fichier « bibaleau.FR.dic » | 33 |
| Fig. 16 Carte d'Allemagne, d'une partie de la République Tchèque et des cours d'eau avec des pages HTML | 34 |

INTRODUCTION

Dans le cadre de ma formation en DESS IST-IE, j'ai eu l'opportunité d'effectuer mon stage de fin d'année d'études universitaires à l'INIST - CNRS (INstitut d'Information Scientifique et Technique - Centre National de la Recherche Scientifique). Cet institut de renommée mondiale dans le domaine de l'Information Scientifique et Technique (IST), collecte, traite et diffuse de l'information de la recherche scientifique, sous forme de copies de documents scientifiques ou de notices bibliographiques. Il propose également ses produits et services sous forme électronique via Internet et permet d'accéder de plus en plus aux documents primaires (articles en texte intégral). Ses différents portails participent au développement de la diffusion de l'information.

L'élargissement de la gamme de produits et services proposée aux clients de l'INIST depuis ces dernières années et les avancées technologiques, en matière de gestion et de diffusion d'un corpus documentaire, ont permis à Jacques DUCLOY, responsable du Département Produits et Services, de mettre en place et d'exploiter la plate-forme DILIB (Document & Information LIBrary). Les avantages de cette plate-forme, par rapport aux produits existants, sa capacité à évoluer selon les besoins propres de l'Institut, font de DILIB un outil puissant et parfaitement adapté aux besoins d'une investigation documentaire qui intègre des fonctions d'analyse infométrique.

C'est avec cette plate-forme logicielle utilisant l'ingénierie SGML-XML que j'ai eu la possibilité de participer à un projet de réalisation d'un serveur d'investigation iconographique et bibliographique pour le Festival International de Géographie de Saint-Dié-des-Vosges de 2003 sur le thème de l'Eau en Allemagne en partenariat avec l'UMR 8586 (Unité Mixte de Recherche) PRODIG (Pôle de Recherche pour l'Organisation et la Diffusion de l'Information Géographique).

Dans un premier temps, je présenterai la mission qui m'a été confiée par l'organisme d'accueil : l'INIST. Puis, je détaillerai les différentes étapes qui m'ont amenées à concevoir un prototype de serveur en vue d'une démonstration lors d'une manifestation de professionnels spécialisés en géographie. Dans un second temps, j'expliquerai la réalisation et le fonctionnement de celui-ci et je terminerai cette étude spécialisée par un bilan exhaustif de la mission en y incluant une perspective d'avenir pour ce nouveau produit INIST issu de la recherche et du développement.

1. PRESENTATION DE L'INIST

1.1. Présentation de l'organisme d'accueil

1.1.1 Qui ?

L'[INIST](http://www.inist.fr)¹ est une Unité Propre de Service (UPS 76) du CNRS et le premier centre intégré européen d'information scientifique et technique. Implanté en Lorraine depuis 1988, il constitue un ensemble de compétences unique en Europe qui collecte et diffuse les résultats de la recherche scientifique mondiale. L'INIST est reconnu comme étant « une source française essentielle d'information mémorisée, l'INIST est aussi le pôle majeur de la fourniture électronique de documents en France » [Liv. 3].

1.1.2 Fait quoi ? Avec quoi ?

Avec son fonds documentaire qui couvre des domaines tels que :

- Les Sciences Humaines et Sociales ;
- La Médecine, la Pharmacologie ;
- La Biologie ;
- Les Sciences de l'Ingénieur ;
- La Physique ;
- La Chimie ;
- La Psychologie ;
- La Terre, l'Océan, l'Espace ;
- Les Sciences de l'Information.

et compte 26 000 titres de périodiques et 242 000 monographies (60 000 rapports scientifiques, 62 000 comptes-rendus de congrès, 110 000 thèses et 10 000 ouvrages), l'INIST propose un gamme complète de services d'accès à l'information scientifique et technique. Premier fournisseur français de copies de documents à caractère scientifique et technique, l'INIST produit également 2 bases de données multilingues et multidisciplinaires recensant l'essentiel de la littérature scientifique internationale avec près de 17 millions de références.

Ses deux bases :

- **PASCAL** avec 14 millions de références bibliographiques multidisciplinaires et multilingues depuis 1973 en Sciences, Technologie et Médecine
- et
- **FRANCIS** avec près de 3 millions de références bibliographiques multidisciplinaires et multilingues depuis 1972 en Sciences Humaines et Sociales

sont accessibles sur cédéroms, serveurs ou Internet.

¹ <http://www.inist.fr>

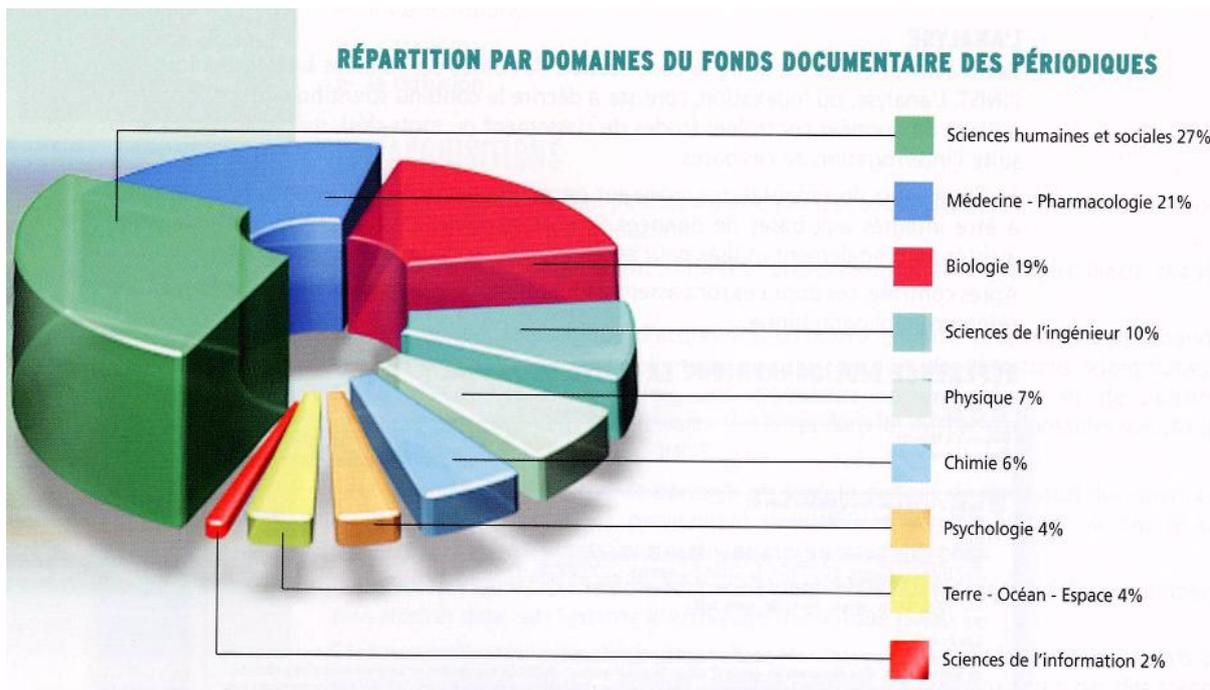


Fig. 1 Répartition par domaine des 26 000 périodiques du fonds documentaire de l'INIST

1.1.2.1 Les services INIST sur Internet

Les technologies de pointe utilisées en matière de stockage de l'information permettent à l'INIST d'offrir des services sur Internet, tels que :

- [Articlesciences](#)² : qui est un moteur de recherche et de commande en ligne d'articles scientifiques et techniques de l'INIST associé à un module de paiement sécurisé en ligne ;
- [BiblioSciences](#)³ : actuellement réservé aux laboratoires du CNRS, BiblioSciences est un accès mutualisé à l'information scientifique et technique regroupant plusieurs bases de données scientifiques et techniques reconnues internationalement ;
- [ConnectSciences](#)⁴ : le portail CNRS d'information scientifique et technique qui permet d'offrir, dans un environnement personnalisé, un accès unique et gratuit à toute une gamme de services spécialisés ;
- [ISRN](#)⁵ : l'International Standard Report Number qui permet d'obtenir un code d'identification international destiné à faciliter l'identification de l'accès aux rapports scientifiques ou administratifs.
- [Services@inist](#)⁶ : Ensemble de services documentaires comprenant :

² <http://articlesciences.inist.fr>

³ <http://bibliosciences.inist.fr>

⁴ http://connectsciences.inist.fr/services/accueil/accueil_invite.php

⁵ <http://www.inist.fr/ISRN/enregistrer.html>

⁶ <http://services.inist.fr/public/fr/conslt.htm>

- ARTICLE@INIST le catalogue des articles et monographies du fonds INIST depuis 1990 ;
 - FORM@INIST un formulaire de commande en ligne pour les copies d'articles ;
 - COMPTE@INIST un module de suivi et de gestion des documents et des commandes de copies d'articles.
- [Thesa](http://thesa.inist.fr)⁷ : la banque de données française qui signale les sujets de thèses en préparation dans les établissements de la conférence des Grandes Ecoles.

1.1.2.2 Les services de veille scientifique et technologique

En outre, l'INIST s'investit dans le développement de services de veille scientifique et technologique. Par ce biais, l'INIST collabore avec plusieurs organismes de recherche à différents projets nationaux, européens et internationaux. L'INIST conçoit, développe et utilise différentes technologies pour l'analyse de l'information ce qui lui permet de proposer différentes prestations en fonction de besoins spécifiques selon un profil personnalisé, des analyses statistiques, des traitements infométriques.

Les outils de veille utilisés à l'INIST sont les suivants :

- **la Plate-forme STANALYST** regroupe différents outils :
 - le serveur **MIRIAD** (Module Intranet de Recherche d'Information et d'Analyse Documentaire) pour l'analyse statistique descriptive des données issues de bases de données ;
 - la plate-forme **ILC** (Infométrie, Langage, Connaissance) pour l'extraction terminologique à partir du texte intégral et des mots-clés ;
 - les logiciels **NEURODOC** et **SDOC**, pour la classification automatique et la représentation graphique des données sous forme de cartes en hypertexte ;
 - le système **HENOCH**, qui organise les résultats et les met à disposition des utilisateurs sur Internet.
- **la Plate-Forme DILIB** développée avec le LORIA (Laboratoire Lorrain de Recherche en Informatique et ses Applications) permet l'exploration dynamique de plusieurs centaines de références bibliographiques sur un sujet donné.
- **le Logiciel LEXIQUEST MINE**, développé par la société Lexiquet, est un logiciel d'exploration terminologique permettant la visualisation de termes associés dans un corpus donné.

⁷ <http://thesa.inist.fr>

1.1.3 Pour qui ?

L'INIST met à disposition des grands organismes publics de recherche, de l'enseignement supérieur et du milieu socioéconomique, un ensemble de ressources et de services destinés à améliorer la diffusion et l'accès à l'information scientifique et technique internationale. Ainsi, l'INIST est présent sur le plan local au travers de partenariats initiés avec les principaux acteurs universitaires, politiques et économiques de la région Lorraine. Son implication dans le plan Etat-Région et sa participation au programme d'intelligence économique en Lorraine permettent l'accueil et la formation d'étudiants des universités lorraines aux techniques de la recherche documentaire.

1.1.4 Avec qui ?

1.1.4.1 Les partenaires

Voici quelques partenaires de l'INIST qui participent à la diffusion de l'information scientifique et technique :

- *des organismes gouvernementaux et internationaux* : ABES (Agence Bibliographique de l'Enseignement Supérieur), l'INPI (Institut National de la Propriété), **PRODIG** (Pôle de Recherche pour l'Organisation et la Diffusion de l'Information Géographique) – le partenaire avec lequel j'ai travaillé tout au long de ma mission, ...
- *des serveurs* : Dialog, Questel-Orbit, Qwam system, ...
- *des éditeurs* : Ovid Technologies, ...
- *des bibliothèques et centres de documentation internationaux* : BNF (Bibliothèque Nationale de France), CISTI (Canada Institute for Scientific et Technical Information)...
- *des associations et organismes professionnels* : GFII (Groupement Français de l'Industrie de l'Information), IFLA (International Federation of Library Associations),...

1.1.4.2 Le personnel de l'INIST

Le personnel de l'INIST se répartit dans différents départements :

- ↪ le Département Prospective et Relations extérieures (**DPRE**)
- ↪ le Département Ressources Documentaires (**DRD**)
- ↪ le Département Produits et Services (**DPS**) où se situe le service des Sciences Humaines et Sociales (**SHS**) où j'ai effectué mon stage.
- ↪ le Département Prototypes et Moyens Informatiques (**DPMI**)

ainsi qu' une Unité de Recherche et d'Innovation (**URI**), un service « Affaires Générales » et un service « Affaires INIST- Diffusion ».

Une multitude de métiers et de compétences représentées par des spécialistes de l'information, veilleurs, ingénieurs, documentalistes, ingénieurs informaticiens et techniciens gèrent les flux d'informations, de la collecte à la diffusion sélective la mieux adaptée aux besoins des différentes communautés d'utilisateurs.

1.1.5 Avec quels moyens ?

Les nouvelles technologies utilisées en matière de stockage, de traitement et de diffusion de l'information et les moyens humains contribuent au bon fonctionnement de cet institut. L'infrastructure informatique de l'INIST est en constante évolution.

- **L'informatique centralisée** concentre toutes les applications et toutes les ressources sur un même ordinateur afin de simplifier et concentrer des tâches d'exploitation.
- **L'informatique dédiée** utilise, pour un type de tâches données, un ordinateur très compétent dans un domaine (IBM 9121, Serveur NOVELL, Machines UNIX,...).
- **L'informatique distribuée** automatise entièrement les échanges d'information d'un ordinateur à un autre.⁸

1.1.6 Avec quels effets ?

En développant une politique active de partenariats avec les communautés scientifiques nationales et internationales ainsi qu'avec les différents acteurs du marché de l'Information Scientifique et Technique, l'INIST est reconnu comme étant « la source principale à utiliser pour se procurer des documents complets » [LIV 1].

Avec une gamme complète de services en ligne, l'INIST satisfait toutes les personnes soucieuses d'obtenir une information pertinente en rapport avec la recherche scientifique internationale et ses résultats. Ainsi l'INIST favorise et développe l'accès à l'information électronique en utilisant des méthodes et des outils de veille scientifique et technique.

1.2. Présentation des équipes du CNRS impliquées dans la mission

➤ L'INIST (Vandoeuvre-lès-Nancy)

C'est le service SHS qui dépend du DPS qui a été chargé de réaliser le serveur d'investigation en Géographie pour l'année 2003.

L'équipe du service Sciences Humaines et Sociales participe à la constitution et à la promotion de la base de données FRANCIS « afin d'assurer sa visibilité auprès des différents acteurs de la recherche »⁹. La base de données FRANCIS est l'œuvre de différents acteurs : les ingénieurs documentalistes nancéens et parisiens (28%), les analystes documentaires (56%) et les partenaires tels que PRODIG et DOGE (Documentation Automatisée en Gestion

⁸ Intranet INIST : Département Prototypes et Moyens Informatiques

⁹ Intranet INIST : Service Sciences Humaines et Sociales (SHS) - Missions

des Entreprises) pour 15%. C'est au sein de ce service dont la responsable est Laurence RAGEOT que j'ai pu effectuer mon stage et produire un serveur en Géographie de références iconographiques et bibliographiques.

J'associe à cette production, Jacques DUCLOY qui a été à l'initiative de ce projet, Bernard SAMPITE (le correspondant de PRODIG à l'INIST), Philippe HOUDRY, François PARMENTIER et Solveig VIDAL pour leurs aides respectives.

➤ **PRODIG (Paris)**

C'est avec le pôle documentaire de l'UMR 8586 [PRODIG](http://prodig.univ-paris1.fr/umr/index.htm)¹⁰ de Paris que l'INIST a été amené à travailler en étroite collaboration.

Cet organisme est une Unité Mixte de Recherche, qui associe le CNRS, l'université (Paris 1, Paris 4, Paris 7) et l'Ecole Pratique des Hautes-Etudes (E.P.H.E.). Cette UMR regroupe des chercheurs, des enseignants-chercheurs et du personnel administratif, elle est située dans les locaux de l'Institut de Géographie de Paris, 191 rue Saint Jacques, dans le 5^{ème} arrondissement. L'UMR (6 équipes) travaille sur des programmes de recherche et dispose également de 3 pôles de mise en valeur des travaux produits, dont le pôle documentation PRODIG qui est le producteur de la « Bibliographie Géographie Internationale » (BGI).

L'INIST acquiert les données bibliographiques de BGI produites par les membres du pôle documentaire ou avec l'aide de correspondants et les intègre dans ses produits (dont FRANCIS). Afin de simplifier la chaîne de traitement, l'INIST a fourni à PRODIG, le logiciel PSILOG pour le catalogage et l'analyse des notices bibliographiques. Grâce à cette coopération, l'INIST peut proposer une base de données en géographie dans ses produits.

Par ailleurs, chaque année la BGI est éditée sous format papier et, est commercialisée par PRODIG. Cette publication couvre l'essentiel du champ géographique mondial. Elle facilite les échanges avec les autres laboratoires de géographie des pays étrangers. Elle est aussi un instrument de diffusion de l'information géographique de tout premier ordre en couvrant l'essentiel du champ géographique mondial.

C'est sur une proposition commune de Jacques DUCLOY, le responsable du DPS à l'INIST et de Claire LIST, la responsable du pôle documentaire de PRODIG qu'est née la décision de créer un serveur iconographique et bibliographique pour l'édition du 14^{ème} Festival International de Géographie de Saint-Dié-des-Vosges.

L'INIST devait apporter sa technologie et PRODIG devait fournir ses matières premières dans le choix des notices et images. A cet effet, l'ingénieur documentaliste de PRODIG, Gaëlle HALLAIR a effectué des prises de vue, sélectionné et indexé toutes les photographies, puis nous a assuré l'envoi de celles-ci. En outre, elle a été la personne ressource en matière de renseignements géographiques.

¹⁰ <http://prodig.univ-paris1.fr/umr/index.htm>

2. PRESENTATION DE LA MISSION

2.1 Origine et définition de la mission

2.1.1 Origine de la mission

Les bonnes relations entretenues entre l'INIST et PRODIG depuis plusieurs années ont favorisé l'émergence d'un projet commun. Ainsi, lors d'une réunion en décembre 2002, il a été décidé de créer un prototype de serveur d'investigation rassemblant des ressources documentaires (iconographiques et bibliographiques) en vue de présenter celui-ci au Festival International de Géographie de Saint-Dié-des-Vosges sur le thème de l'année : « *Eau et Géographie : source de vie, source de conflits, trait d'union entre les hommes* » avec comme pays invité « *l'Allemagne* ».

La réalisation de ce serveur m'a été confiée par Jacques DUCLOY, responsable du DPS et c'est Laurence RAGEOT, responsable du service SHS, qui a mis en place les conditions nécessaires à la réussite de ma mission.

2.1.1 Définition de la mission

Ma mission a consisté à organiser des ressources bibliographiques et iconographiques, respectivement extraites de la base de données FRANCIS et du fonds iconographique de PRODIG afin de constituer *un serveur d'investigation* sur le thème de l'eau en Allemagne à l'aide de la plate-forme DILIB¹¹ développée par l'INIST et le LORIA.

A l'issue de cette mission, ce serveur sera présenté au 14^{ème} Festival International de Saint-Dié-des-Vosges qui se déroulera du 2 au 5 octobre 2003.

2.2 Notion de « serveur d'investigation »

2.2.1 Etymologie de l'expression

D'après le dictionnaire de l'Informatique et de l'Internet le terme *serveur* « vient de l'anglais *server*, on-line data service » et signifie qu'il s'agit « d'un système informatique destiné à fournir des services à des utilisateurs connectés et, par extension, organisme qui exploite un tel système ». Il est précisé qu'un *serveur* peut, par exemple, permettre la consultation et l'exploitation directe de banques de données.¹²

Sous la rubrique informatique du Petit Robert, un *serveur* « est un système informatique qui héberge un ou des services télématiques [Dic. 1].

D'après ce même dictionnaire, le terme *investigation* vient du latin *investigatio* et signifie qu'il s'agit « d'une recherche suivie, systématique sur quelque objet ».

¹¹ Plate-forme pour l'Ingénierie du document et l'information scientifique et technique.

¹² Journal officiel du 16 mars 1999 "Vocabulaire de l'informatique et de l'Internet"

2.2.2 Définitions spécialisées

Ici, l'expression « *un serveur d'investigation* » signifie qu'il s'agit d'un serveur qui permet de naviguer dans un corpus de notices que celles-ci soient de nature bibliographiques et/ou iconographiques. Grâce à *un serveur d'investigation*, généré automatiquement à l'aide de la plate-forme DILIB développée par l'INIST et le LORIA, l'on peut (par exemple, pour un corpus de notices bibliographiques) :

- ◆ Naviguer et retrouver les notices pertinentes à partir de plusieurs index (descripteurs, mots du titre, auteurs) ;
- ◆ Explorer rapidement et de manière récursive les sous-thématiques (« clusters ») mises en évidence grâce à des méthodes de calculs infométriques liées à la méthode des mots associés (co-occurrences) ;
- ◆ Découvrir pour chaque cluster (*agrégat*) :
 - ◆ la liste des termes qui le constituent ;
 - ◆ la liste des associations internes (relations entre les termes du cluster) ;
 - ◆ la liste des documents pertinents ;
 - ◆ la liste des auteurs associés au cluster ;
 - ◆ la liste des périodiques associés au cluster ;
 - ◆ la liste des affiliations associées au cluster ;
 - ◆ la liste des pays d'affiliation associés au cluster ;
 - ◆ un histogramme par année ;
- ◆ Envoyer une requête sur Internet après analyse des mots des notices du cluster.

Il s'agit encore d'un prototype dont certaines fonctionnalités sont en cours d'amélioration [Res. 8]. Les serveurs d'investigation représentent une manière originale d'exploiter un corpus de données pour en dégager les principaux thèmes. A l'occasion de manifestations professionnelles et scientifiques, l'INIST élabore des serveurs expérimentaux autour des sujets abordés.

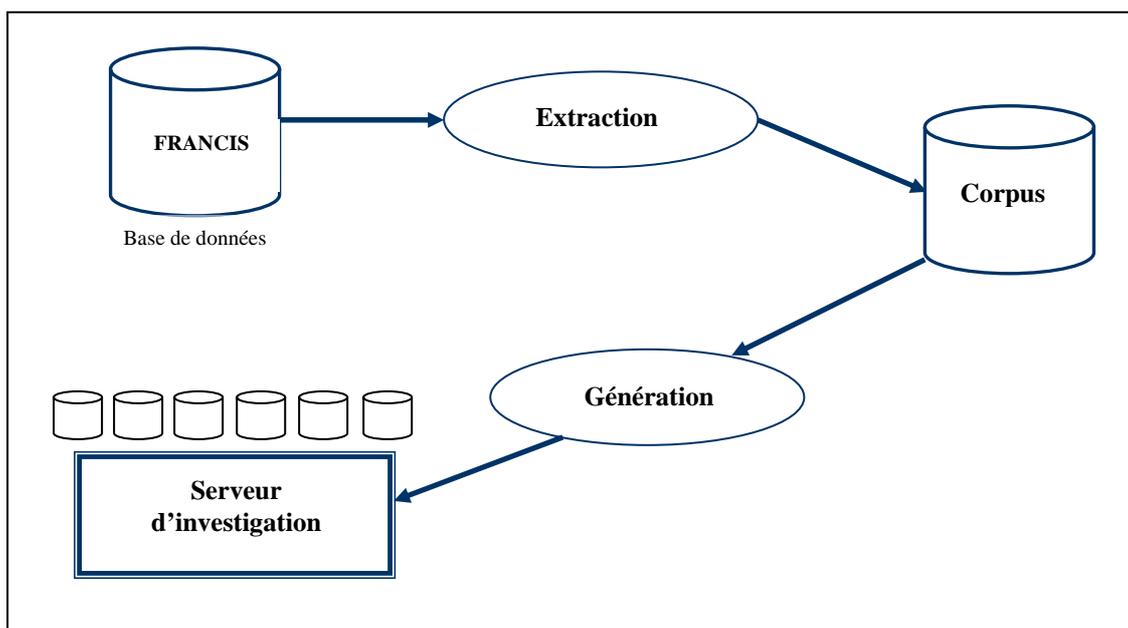


Fig. 2 Constitution d'un serveur d'investigation [Res. 5]

A l'aide des bases de données FRANCIS et PASCAL, l'on extrait un corpus de ressources documentaires. C'est avec la plate-forme bibliométrique DILIB que l'on peut générer un serveur d'investigation.

2.3 Objectifs de la mission

Pour réaliser ce serveur d'investigation, en vue d'une présentation au Festival International de Géographie de Saint-Dié-des-Vosges en octobre 2003, il m'a fallu :

- développer un nouveau produit à base de photographies (numérisées) à l'aide de la plate-forme DILIB avec l'application pilote qui avait été utilisée pour la constitution du serveur BIBAN.
- puis, réfléchir sur la méthodologie de travail à adopter en vue d'une utilisation ultérieure par des acteurs de l'INIST.

2.4 Etapes de la mission

2.4.1 Analyse de l'existant

Pour concrétiser un tel projet qui s'intègre dans les travaux du DPS, j'ai dû prendre en compte un certain nombre de paramètres, notamment les travaux antérieurs à ma mission.

Lors d'un précédent Festival (*le 13^{ème} Festival International de Géographie de 2002*), un serveur d'investigation de notices bibliographiques avait déjà été réalisé par l'INIST sur le thème : « [Géographie et Religions](#) ».

Afin de proposer un nouveau produit lors de cette manifestation de l'année 2003 nommée « **FIG 2003** »¹³, Jacques DUCLOY et Claire LIST ont souhaité adjoindre à un serveur de références bibliographiques des *ressources iconographiques*.

A l'aide d'une application pilote commencée au LORIA en 1995 par l'équipe LORIASI (LORIA dans la Société de l'Information), un serveur nommé BIBAN (Base Iconographique et Bibliographique Art Nouveau) a été conçu à l'occasion du centenaire de l'Art Nouveau et de l'Ecole de Nancy. Le serveur BIBAN a été le premier serveur regroupant en un seul endroit des ressources bibliographiques et iconographiques [Rap.1].

Il est « le premier prototype de bibliothèque électronique qui contient un ensemble de « livres électroniques » en HTML, une base iconographique, une base bibliographique (extrait de la BHA¹⁴, coproduite par l'INIST et J.P. Getty Trust¹⁵) et des références de sites Internet. Une indexation homogène a permis d'introduire plusieurs dispositifs de navigation iconographique, depuis la simple liste d'images associées à un mot-clé, jusqu'à une classification neuronale en passant par des mosaïques d'images proches. BIBAN intègre également une dimension multilingue (français, anglais, allemand).

¹³ FIG : Festival International de Géographie.

¹⁴ Bibliographie de l'Histoire de l'Art – <http://www.inist.fr/BHA>

¹⁵ Pour en savoir plus : <http://www.getty.edu/>

2.4.2 Analyse des besoins

Le serveur en géographie destiné à un public de professionnels et de scientifiques présents lors d'une manifestation internationale devait permettre de présenter une réalisation innovante en matière de diffusion d'informations bibliographiques et iconographiques. A cet effet, il a fallu que j'étudie :

d'une part :

- ◇ les exigences du demandeur en l'occurrence Jacques DUCLOY qui a décidé de mettre en place à l'INIST et pour la manifestation de géographie de Saint-Dié-des-Vosges, un nouveau serveur combinant plusieurs types de données ;
- ◇ les exigences de notre partenaire PRODIG ;
- ◇ les circonstances qui ont amené ceux-ci à décider une telle initiative ;
- ◇ les contraintes de réalisation du projet :
 - durée de la mission : 4 mois,
 - dates du Festival : du 02 au 05 octobre 2003,
 - disponibilité des partenaires : équipe de PRODIG, personnel INIST - en tenant compte des périodes de congés de chacun ;
- ◇ les moyens mis à ma disposition pour concrétiser ce projet (matériels et logiciels disponibles) ;
- ◇ la possibilité d'obtenir des aides avec du personnel qualifié (tel que celles des ingénieurs documentalistes de l'INIST et de PRODIG) ;

et d'autre part, que je prévois l'évolution possible du projet, en terme :

- ◇ de volume de travail ;
- ◇ d'approches techniques ;
- ◇ d'implication en nombre et qualification du personnel ;
- ◇ de continuité du projet.

2.4.3 Appropriation des outils et technologies

En premier lieu, j'ai dû m'approprier les outils nécessaires à la réalisation du serveur en utilisant :

- l'environnement UNIX ;
- l'analyseur lexical LEX ;
- le langage XML (eXtensible Markup Language) ;
- la codification des ressources iconographiques en RDF (Resource Description Framework) ;
- la plate-forme DILIB ;

puis :

- faire un état de l'art sur l'indexation d'images et leur classement ;
- m'informer sur le droit de l'image.

2.4.4 Déroulement de la mission

Ma mission a consisté à :

- travailler sur l'application pilote BIBAN afin d'assimiler la démarche souhaitée ;
- collaborer avec le partenaire PRODIG pour récupérer les données et définir la structure des notices d'indexation des images, mais aussi adapter ma stratégie en fonction des réflexions de notre partenaire ;
- créer une application qui permette de proposer un serveur en géographie à base de ressources bibliographiques et iconographiques ; tout d'abord, en créant un serveur d'images, puis un serveur bibliographique et regroupant les deux pour ne former qu'un seul serveur ;
- concevoir des pages HTML pour présenter les photographies ;
- construire un thésaurus en RDF/XML pour l'application ;
- prévoir un plan de classement pour cette application.

Au terme de la mission, ce nouveau serveur d'investigation alliant des ressources issues d'une base bibliographique et d'une base iconographique doit permettre de proposer un nouveau produit à destination des clients de l'INIST et de valoriser les travaux des chercheurs en géographie. A cet effet, un processus de travail a été mis en place afin de faciliter la réalisation de nouveaux serveurs, notamment en vue de proposer des produits INIST à base d'images.

3. CONCEPTION DU SERVEUR D'INVESTIGATION : « FIG 2003 »

3.1 Phase expérimentale

3.1.1 Prise de connaissance de l'environnement DILIB

L'utilisation de la *plate-forme DILIB* a nécessité de ma part une certaine adaptation à l'environnement UNIX. En effet, cette plate-forme qui contient des fichiers de données codées selon différents formats et des outils développés en langage C, est utilisable sous UNIX.

Il m'a fallu également assimiler ce que contenait DILIB. Ainsi, j'ai découvert que cette plate-forme pour l'Ingénierie du Document et de l'Information Scientifique et Technique rendait possible l'utilisation des techniques issues des normes SGML (Standard Generalized Markup Language) et XML sur des formats MARC (Machine Readable Cataloguing). De plus, ses applications visées sont particulièrement destinées à traiter et explorer l'information technique, documentaire et multimédia. Enfin, elle permet une construction de Systèmes de Recherche d'Information et produit des outils pour les bibliothèques électroniques. L'exploration des corpus documentaires est facilitée par une navigation hypertexte et leur exploitation par une analyse infométrique.

Conçue comme un moteur de recherche, l'architecture informatique de DILIB comprend :

- des notices ;
- des consignes ;
- des fichiers directs (copies de notices) ;
- des fichiers inverses (auteurs, mots-clés, mots du titre) ;
- des fichiers paramètres pour la navigation (mots-vides).

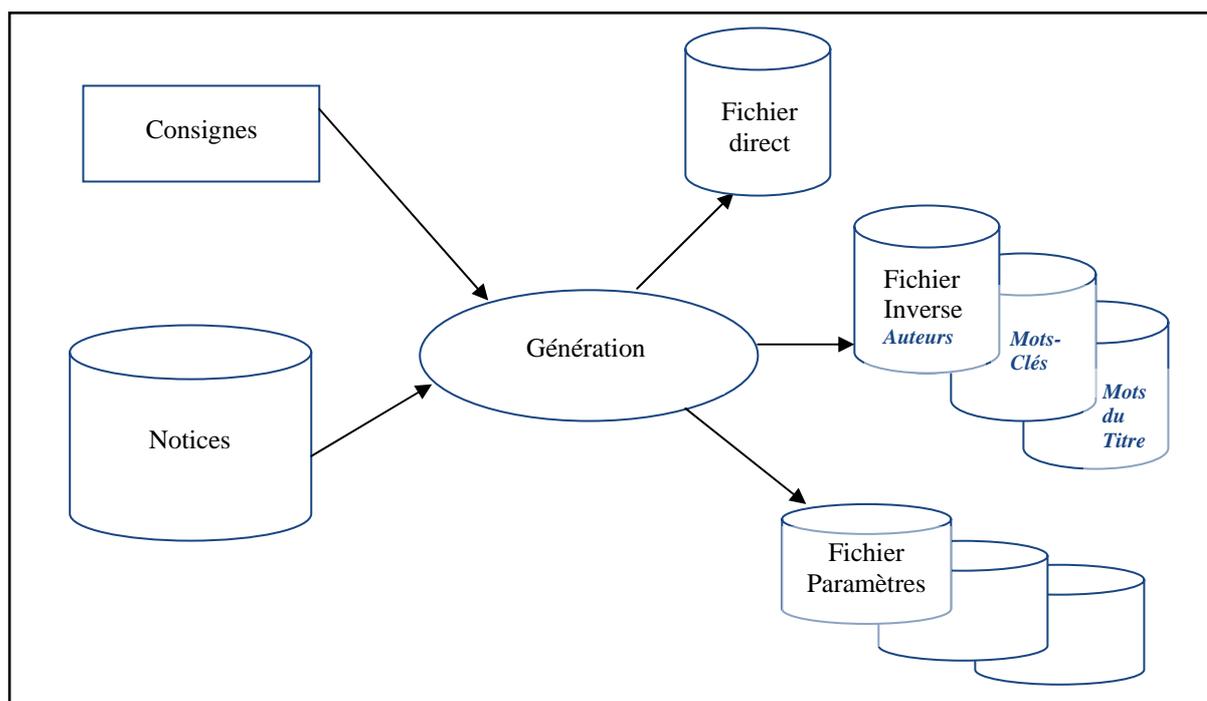


Fig. 3 Architecture générale d'un serveur [Res. 4]

L'accès à l'information s'effectue selon une architecture HFD (Hiérarchical File organisation for Documentation), organisation hiérarchisée qui permet de stocker des millions d'enregistrements avec un accès direct.

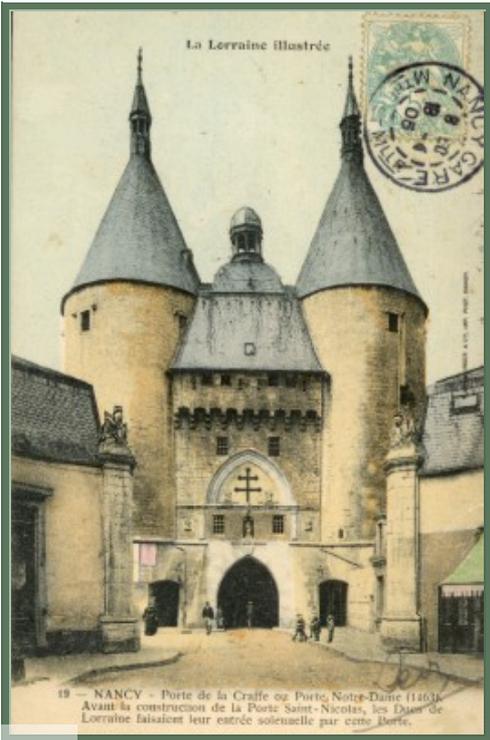
3.1.2 Première approche du serveur d'investigation BIBAN

La seconde étape de ma mission a consisté à m'imprégner du travail effectué sur le serveur [BIBAN](#) (*accès réservé*) - Base Iconographique et Bibliographique Art Nouveau – conçu et réalisé à partir de la plate forme logicielle DILIB [Rap.1].

A partir de cette application, j'ai tout d'abord étudié la manière dont étaient codifiées les ressources iconographiques en RDF (Resource Description Framework). A l'aide du fichier source, j'ai pu comprendre la description des images en RDF, un standard W3C basé sur XML. Ce n'est qu'après ce repérage que j'ai pu indexer des cartes postales de Nancy extraites de la collection personnelle de Jacques DUCLOY. Cette première approche m'a permis de me familiariser avec la technique d'indexation de documents iconographiques en utilisant un formalisme indépendant des systèmes d'exploitation qui fournit un cadre conceptuel pour définir et utiliser les métadonnées. Ce *modèle RDF* s'appuie sur des principes bien établis provenant des différentes communautés de représentation des données et permet l'interprétation des informations contenues dans les métadonnées. Les éléments en *Dublin Core* sont ainsi compris comme des propriétés de RDF qui leur donne un cadre ou une syntaxe particulière [Res. 3].

```

<rdf:Description about="Images/Ducloy/craffe"> } Elément de déclaration
  <dc:Title xml:lang="fre">Nancy : La porte de laCraffe</dc:Title> } Elément du titre avec l'attribut de langue
  <dc:Type>image</dc:Type> } Type de document : image
  <dc:Identifier>
    <inist:icon>Images/Ducloy/craffe.icon.jpg</inist:icon>
    <inist:inText>Images/Ducloy/craffe.text.jpg</inist:inText>
    <inist:big>Images/Ducloy/craffe.big.jpg</inist:big>
  } Les images sous 3 formats :
  petit format : icon - moyen format : inText
  grand format : big
  </dc:Identifier>
  <dc:Subject>
    </Bag>
    <inist:FD>
    <Bag>
      <li>Nancy</li>
      <li>B&acirc;timent militaire</li>
    </Bag>
    </inist:FD>
    <inist:image>
    <Bag>
      <li>Craffe</li>
      <li>Sculpture</li>
    </Bag>
  } Les mots-clés français
  </inist:image>
  <inist:cc>/003/531/162/183</inist:cc> } Les codes de classement INIST
  <inist:cc>/003/530</inist:cc>
  </dc:Subject>
</rdf:Description> } Elément de fin de déclaration
    
```



Notice carte postale : « La porte de la Craffe » à Nancy en description RDF

3.2 Serveur iconographique

3.2.1 L'indexation des images

Cette première approche d'indexation de photographies m'a tout de suite amenée à me documenter sur le domaine afin d'acquérir rapidement des compétences dans le traitement des documents iconographiques. A cet effet, j'ai dressé *une bibliographie sélective* sur le thème de : « l'indexation des images » ([Annexel](#)) en utilisant différentes sources de documentation :

- le SUDOC (Système Universitaire de Documentation) ;
- ARTICLE@INIST ;
- la plate-forme bibliométrique MIRIAD (avec PASCAL et FRANCIS);

puis, j'ai contacté par courriel un de mes anciens professeurs, Gérald DUFFING qui travaille au LORIA sur l'indexation des images [Res. 1] [Res. 3]

Le résultat de mes recherches a volontairement été limité à mes besoins immédiats, soit :

- la recherche dans la manière d'indexer des photographies [Liv. 5] ;
- la recherche de règles utilisées pour la description des documents (*Eléments définis par le Dublin Core*) [Liv. 2] [Liv. 5] ;
- la nécessité d'un plan de classement [Liv. 5] ;
- l'utilisation d'un thésaurus utilisable pour les ressources bibliographiques et iconographiques [Liv. 4].

3.2.2 Le droit de l'image

En ce qui concerne le droit de l'image, j'ai tout de suite été sollicitée pour faire des recherches sur le droit relatif au traitement et diffusion des images. Ma démarche de recherche de documentation a quelque peu été différente de la précédente, puisque je possédais déjà toute la documentation sur le droit de l'information (documentation personnelle issue de mon abonnement au périodique mensuel édité à l'ADBS - *Actualités du droit de l'information*). Ainsi, j'ai pu préciser la nature complexe du droit de l'image en matière de diffusion d'images tant sur le plan des textes de loi et que sur celui des jurisprudences [Art. 1] [Art. 2] [Art. 3].

Pour ce qui me concernait et en fonction des documents que je devais traiter pendant cette mission, aucun problème de droit d'auteurs ne s'est réellement posé puisque suite à un projet commun et aux accords tacites passés entre les deux protagonistes du projet (INIST et PRODIG) :

- toutes les notices bibliographiques appartenaient aux deux organismes ;
- toutes les photographies, bien qu'appartenant à PRODIG, pouvaient être utilisées dans le contexte d'une manifestation en géographie.

Par contre, il a été précisé que si le serveur d'investigation devait être visible par d'autres utilisateurs, soit dans un autre contexte que celui du Festival International de Géographie, l'INIST devra prévoir un accord écrit avec PRODIG afin d'obtenir une autorisation de diffusion des photographies.

3.2.3 Création d'un modèle de notice

Après avoir compris vers quelle technique d'indexation je devais m'orienter, j'ai créé un cadre formel pour récupérer les indexations de photographies produites par PRODIG. En m'inspirant du modèle précédent, j'ai élaboré la structure suivante pour que toutes les données des notices soient présentes dans l'indexation.

Ces éléments de données sont les suivants :

Étiquette des données

```

name :
tiFR :
tiEN :
tiDE :
icon :
text :
big :
edit :
mcFR :
    li :
mcEN :
    li :
mcDE :
    li :
incc :
$
name :
[suite fichier]
    
```

Fig. 4 Fichier modèle Indexation

Étiquette des données + Éléments de données

```

name : phot01
tiFR : Le tissu urbain ancien et moderne sur les rives du lac Alster à Hambourg
tiEN : The old and modern town-quarter on the Alster lake in Hamburg
tiDE : Alte und moderne Stadtviertel am Alsterufer in Hamburg
icon : phot01.icon.jpg
text : phot01.text.jpg
big : phot01.big.jpg
edit : PRODIG
mcFR :
    li : Eau
    li : Lac
    li : Paysage urbain
    li : Quartier
    li : Rénovation urbaine
    li : Espace bâti
    li : Tourisme
    li : Allemagne
    li : Hamburg
mcEN :
    li : Water
    li : Lake
    li : Urban landscape
    li : Urban district
    li : Urban renewal
    li : Built up area
    li : Tourism
    li : Germany
    li : Hamburg
mcDE :
    li : Wasser
    li : See
    li : Urbane Landschaft
    li : Viertel
    li : Stadtsanierung
    li : Bebauter Raum
    li : Tourismus
    li : Deutschland
    li : Hamburg
incc : 531245
$
name : phot02
[suite fichier]
    
```

Numéro photographie

Titre en français, anglais et allemand

Format des 3 images (petit : icon moyen : text – grand : big)

Zone de l'éditeur

Mots-clés français



Mots-clés anglais

Mots-clés allemands

Délimiteur de notice

Fig. 5 Exemple : Notice d'indexation de la photographie n°1 – phot01

3.2.4 Reformatage des notices « images »

➤ Programme en LEX :

A partir de ce fichier structuré *sous format texte* et afin d'éviter notamment les problèmes de transcription, il a fallu automatiser la structuration des données des photographies en RDF nécessaire à la génération du serveur. La rédaction d'un programme en LEX s'est avérée alors indispensable.

LEX est un générateur d'analyseur lexical qui permet d'associer des actions à des règles de reconnaissance de forme des chaînes de caractères, voire d'expressions régulières. Pour un reformatage simple, cet outil est bien adapté puisqu'il sélectionne des chaînes de caractères, les remplace ou fait un traitement particulier sur celles-ci.

Le programme LEX présenté ci-dessous permet de préparer le fichier pour la conversion en RDF/XML.

*A gauche, les chaînes de caractères à remplacer.
A droite, les chaînes de caractères de remplacement
= 1 règle*

Exemple avec `^"$"` :

A chaque début de ligne matérialisé par `^` puis suivi d'un `$`, l'analyseur LEX remplace cette chaîne de caractère en l'interprétant en un format de type RDF/XML

`<rdf: Description>`

Idem :

- pour la chaîne de caractères name qui se retrouve entre balise `<name>`
- pour le tire en français, le titre en anglais, le titre en allemand....

```
%%
\\015 ;
^"$"
^name[ ]*":"[ ]*
^tiFR[ ]*":"[ ]*
^tiEN[ ]*":"[ ]*
^tiDE[ ]*":"[ ]*
^mcFR[ ]*":"[ ]*
^mcEN[ ]*":"[ ]*
^mcDE[ ]*":"[ ]*
^[ \t]+li[ ]*":"[ ]*
^text[ ]*":"[ ]*
^icon[ ]*":"[ ]*
^big[ ]*":"[ ]*
^edit[ ]*":"[ ]*
^incc[ ]*":"[ ]*
%%
main()
{
  printf("<rdf:Description>\n");
  yylex();
}
```

```
printf("<rdf:Description>\n");
printf("<name>");
printf("<dc:Title xml:lang=\"fre\">");
printf("<dc:Title xml:lang=\"eng\">");
printf("<dc:Title xml:lang=\"deu\">");
printf("<mcFR><li>");
printf("<mcEN><li>");
printf("<mcDE><li>");
printf("<li>");
printf("<inist:inText>");
printf("<inist:icon>");
printf("<inist:big>");
printf("<edit>");
printf("<incc>");
```

Fig. 6 Programme en LEX qui prépare le fichier indexation pour sa conversion en RDF/XML avec le programme en C (ci-dessous)

➤ Programme en langage DILIB

Puis un programme en langage DILIB a permis de récupérer cette nouvelle structuration de données décrites afin de créer des pages HTML pour les visualiser sur le serveur d'investigation.

Inspiré du modèle DOM (Document Object Model), utilisé ici et adapté au contexte de programmation DILIB, l'on peut définir ce modèle comme un arbre avec des relations entre les nœuds.

Ce modèle fournit un ensemble d'objets pour :

- représenter des documents XML ;
- combiner, accéder et manipuler des documents XML.

Il existe des mécanismes de désignation d'un élément dans un document : XPath et XPointer, ainsi que des mécanismes de gestion des liens sur le Web : XLink.

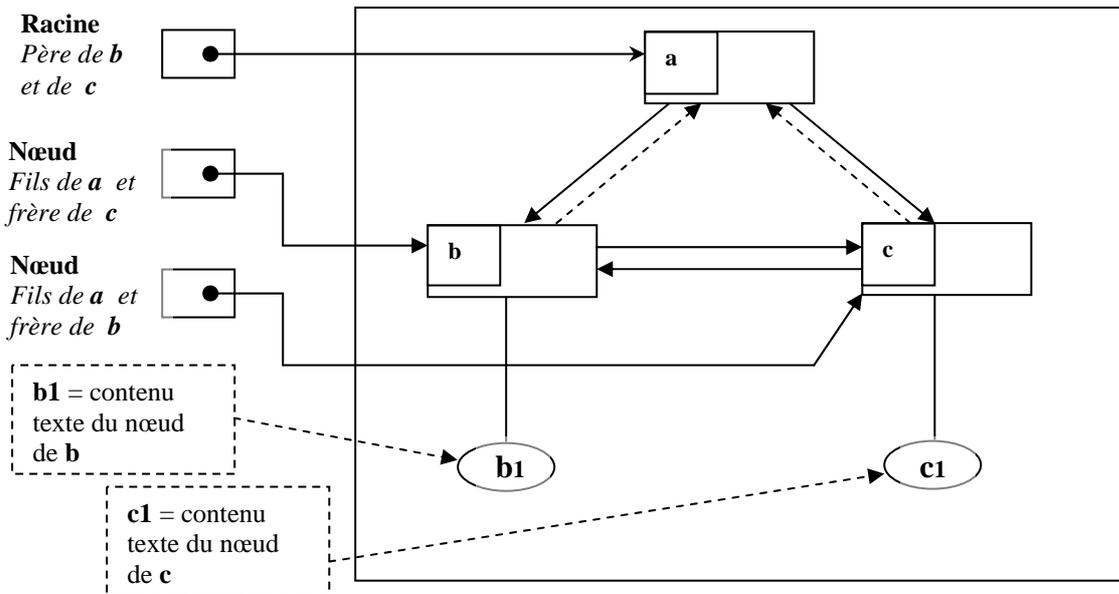


Fig. 7 Construction d'arbre en XML

`<a>b1<c>c1</c><a>`

La programmation dans DILIB permet, entre autre (XmlNode étant un objet DILIB représentant un arbre Xml) :

- d'ajouter un nouvel élément en tête des fils - *Cette méthode est une extension du W3C.*
XmlNode *XmlAddFirstChild(pere, fils) ;
XmlNode *pere, *fils;
- d'ajouter un nouvel élément en queue de la liste des fils
XmlNode *XmlAppenchild(pere, fils) ;
XmlNode *pere, *fils;

Pour créer un arbre :

- on **déclare** les noms de nœuds Xml que l'on va utiliser : Xml *noeudA ; ex : ("*a*")
- on **affecte** une valeur à ces noms noeudA=XmlElementCreate("tag") ;
noeudB=XmlLeafCreate("tag","texte") ; ex : ("*b*", "*b1*")
- on **relie** les nœuds : XmlAppendChild(XmlNodePere,XmlNodeFils) ;

Détails du programme utilisé :

```

#include "XmlNode.h"

main()
{
  XmlNode *doc;
  Buffer *bField;

  bField=BufferCreate(10,10);

  while((doc=XmlInputNextDocumentElement()))
  {
    XmlNode *inputField;
    XmlNode *subject;
    XmlNode *ident;

    XmlAddFirstChild(doc, XmlLeafCreate("dc:Type","image"));
    if ((inputField=XmlGetFirstChildByTagName(doc, "name"))
        {
          XmlSetAttribute(doc, "about", XmlLeafText(inputField));
          XmlFree(inputField);
        }

    subject=XmlElementCreate("dc:Subject");
    if ((inputField=XmlGetFirstChildByTagName(doc, "mcFR"))
        {
          XmlNode *inistFD;
          XmlRemoveChild(inputField);
          inistFD=XmlElementCreate("inist:FD");
          XmlAppendChild(subject,inistFD);
          XmlReplaceName(inputField,"Bag");
          XmlAppendChild(inistFD,inputField);
        }

    if ((inputField=XmlGetFirstChildByTagName(doc, "mcEN"))
        {
          XmlNode *inistED;
          XmlRemoveChild(inputField);
          inistED=XmlElementCreate("inist:ED");
          XmlAppendChild(subject,inistED);
          XmlReplaceName(inputField,"Bag");
          XmlAppendChild(inistED,inputField);
        }

    if ((inputField=XmlGetFirstChildByTagName(doc, "mcDE"))
        {
          XmlNode *inistGD;
          XmlRemoveChild(inputField);
          inistGD=XmlElementCreate("inist:GD");
          XmlAppendChild(subject,inistGD);
          XmlReplaceName(inputField,"Bag");
          XmlAppendChild(inistGD,inputField);
        }

    if ((inputField=XmlGetFirstChildByTagName(doc, "inist:cc"))
        {
          XmlNode *inistcc;
          XmlRemoveChild(inputField);
          inistcc=XmlElementCreate("inist:cc");
          XmlAppendChild(subject,inistcc);
        }
  }
}

```

Bibliothèque DILIB

Déclaration des nœuds que l'on va manipuler pour construire la structure de l'arbre XML/RDF

Création d'un document type métadonnées " image "

Création d'un élément Subject avec comme nœuds :

- des mots-clés
- français
mcFR
(mots-clés français)
- anglais
mcEN
(mots-clés anglais)
- allemands
mcDE
(mots-clés allemands)
- des codes de classement

| | |
|--|---|
| <pre> ident=XmlElementCreate("dc:Identifieur"); if ((inputField=XmlGetFirstChildByTagName(doc, "inist:icon"))) { XmlRemoveChild(inputField); BufferStrcpy(bField, "Images/"); BufferStrcat(bField, XmlLeafText(inputField)); XmlAppendChild(ident, XmlLeafCreate("inist:icon", BufferString(bField))); XmlFree(inputField); } if ((inputField=XmlGetFirstChildByTagName(doc, "inist:inText"))) { XmlRemoveChild(inputField); BufferStrcpy(bField, "Images/"); BufferStrcat(bField, XmlLeafText(inputField)); XmlAppendChild(ident, XmlLeafCreate("inist:inText", BufferString(bField))); XmlFree(inputField); } if ((inputField=XmlGetFirstChildByTagName(doc, "inist:big"))) { XmlRemoveChild(inputField); BufferStrcpy(bField, "Images/"); BufferStrcat(bField, XmlLeafText(inputField)); XmlAppendChild(ident, XmlLeafCreate("inist:big", BufferString(bField))); XmlFree(inputField); } XmlAppendChild(doc,ident); XmlAppendChild(doc,subject); XmlPrint (XmlInputRecord); } } </pre> | <p>Création d'un élément "dc : Identifieur" avec comme nœuds des noms de formats d'images :</p> <ul style="list-style-type: none"> - <i>icon</i> - <i>inText</i> - <i>big</i> <p>Attache les nœuds fils Identifieur et Subject au nœud père (racine) doc. et imprime l'objet « notice indexation image en RDF/XML » (type serveur BIBAN)</p> |
|--|---|

Fig. 8 Programme en C qui convertit en RDF/XML

3.2.5 Traitement des photographies

Parallèlement à ce travail de reformatage, j'ai redimensionné toutes les photographies qui me sont parvenues sous format TIFF (Tagged Image File Format) - choix effectué par l'équipe de PRODIG, soit un total de 88 photographies.

Ce format TIFF m'a permis de travailler sur une qualité d'image excellente. Tout en conservant les originaux sous format TIFF, j'ai pu constituer un corpus d'images sous trois dimensions différentes sous format JPEG (Join Photographic Experts Group) afin d'offrir une meilleure diffusion sur Internet.

Les trois dimensions retenues sont :

| Pour les photographies horizontales : | Pour les photographies verticales : |
|---|---|
| <ul style="list-style-type: none"> - 166 x 107 pixels pour le format icon (vignette) ; - 307 x 199 pixels pour le format inText (pour une image intégrée au texte) - 832 x 538 pixels pour le format big (grande image). | <ul style="list-style-type: none"> - 108 x 166 pixels pour le format icon (vignette) ; - 248 x 384 pixels pour le format inText (pour une image intégrée au texte) ; - 414 x 640 pixels pour le format big (grande image). |

Les trois formats de photographies horizontales et verticales sont dus aux choix effectués en amont pour présenter celles-ci sur un écran 17 pouces sans utiliser l'ascenseur du navigateur (avec une optimisation de 1024 x 768).

3.2.6 Création d'un serveur d'investigation de photographies

Afin de générer un serveur iconographique avec DILIB ([Annexe 2](#)), il a fallu travailler également sur un shell Unix pour lancer toutes les applications iconographiques. Nous avons ainsi repris ce qui existait dans la base BIBAN et avons complété celui-ci de manière à ce qu'elle incorpore les photographies de géographie.

- *Etape 1* : Récupération des paramètres ;
- *Etape 2* : Effacement des anciens répertoires puis création de nouveaux répertoires ;
- *Etape 3* : Création d'un répertoire « plan de classement » ;
- *Etape 4* : Présentation pour le plan de classement ;
- *Etape 5* : Définition du plan de classement ;
- *Etape 6* : Création de l'application géographie
- *Etape 7* : Création de l'application « cartes postales » ;
- *Etape 8* : Création d'un serveur d'investigation base d'images.

Les premiers essais sur ce nouveau serveur faisaient apparaître toutes les images (cartes postales de Nancy et photographies de « l'eau » en Allemagne). Cette première étape d'introduction d'un nouveau corpus d'images a été rendue possible grâce à *l'application pilote* sur laquelle l'expérimentation s'est déroulée les premières semaines de ma mission.

3.3 Serveurs bibliographiques

C'est avec MIRIAD (Module Intranet de Recherche d'Information et d'Analyse Documentaire), une plate-forme bibliométrique qui a pour objet la mise en place d'un système d'aide à la gestion et à la prise de décisions dans la production des bases de données de l'INIST que j'ai récupéré les premiers corpus de notices bibliographiques.

MIRIAD permet de faire :

- une recherche d'information sur Pascal depuis 1984 et sur FRANCIS depuis 1992 ;
- des statistiques standards sur des données bibliographiques, des périodiques et des mots-clés ;
- des téléchargements de résultats (pour PC sous Windows, Macintosh et Unix)

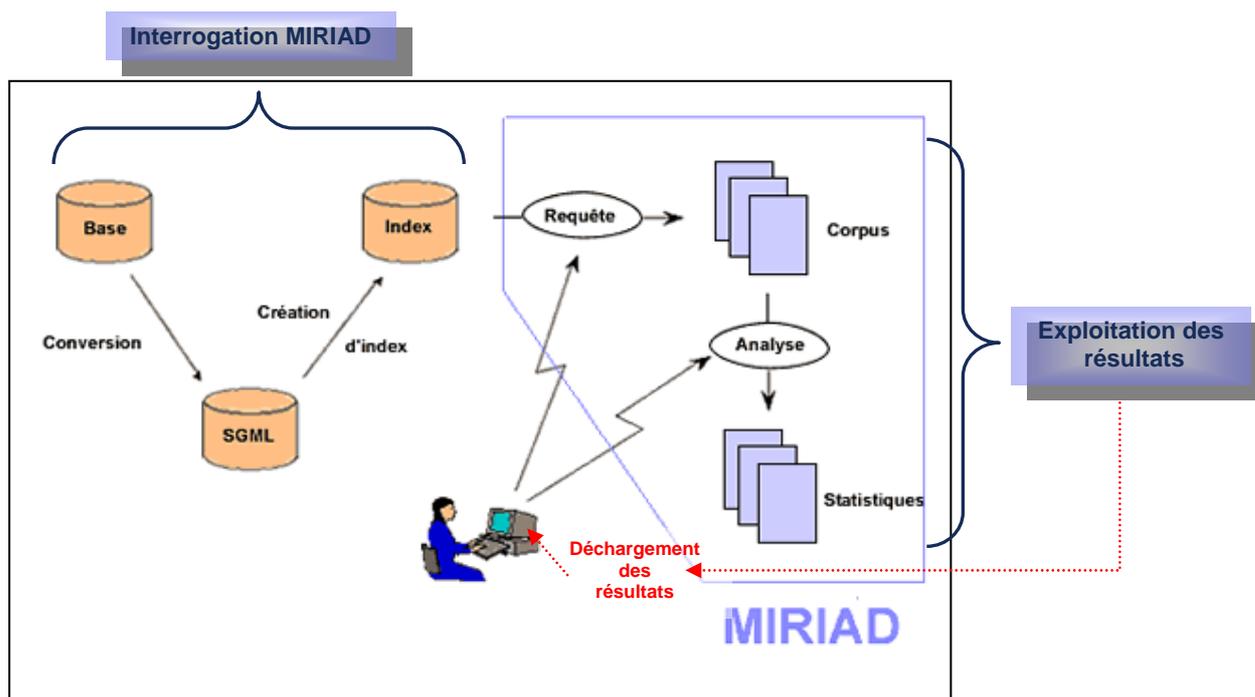


Fig. 9 Plate-forme MIRIAD [Res. 9]

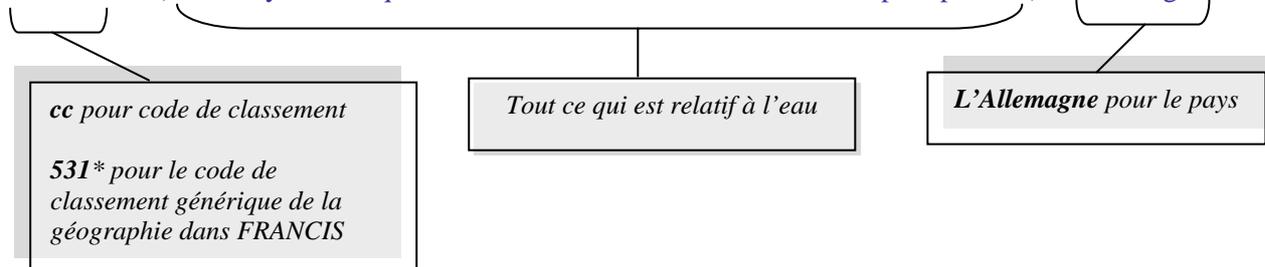
3.3.1 Constitution d'un corpus de notices bibliographiques avec MIRIAD

La réalisation du serveur iconographique et bibliographique a nécessité la réalisation d'un serveur bibliographique sur « l'Eau en Allemagne ». Pour cela, il a été prévu de créer trois serveurs afin de choisir celui qui semblerait être le plus en adéquation avec le serveur iconographique.

J'ai donc élaboré trois recherches multicritères sur la plate-forme bibliométrique MIRIAD en interrogeant la géographie dans la base de données FRANCIS des années 1992 à 2003.

- La 1^{ère} recherche multicritère sur *l'Eau en Allemagne* :

cc= 531* et (eau ou hydr* ou aqu* ou inondation ou crue ou fluviale ou precipitation) et Allemagne



- La 2^{ème} recherche multicritère sur *l'Eau en Allemagne et pays limitrophes* :

cc= 531* et (eau ou hydr* ou aqu* ou inondation ou crue ou fluviale ou precipitation) et (allemagne ou autriche ou tchecoslovaquie ou "ex-tchecoslovaquie" ou "tcheque? republique" ou slovaquie ou suisse ou pologne ou alsace ou lorraine)

- La 3^{ème} recherche multicritère sur *l'Eau en Allemagne et Tchécoslovaquie* :

cc= 531* et (eau ou hydr* ou aqu* ou inondation ou crue ou fluviale ou precipitation) et (allemagne ou tchecoslovaquie ou "ex-tchecoslovaquie" ou "tcheque? republique")

3.3.2 Déchargement des corpus et constitution des serveurs bibliographiques

Le déchargement des corpus des notices bibliographiques extraites de la base de données FRANCIS, puis la constitution des serveurs bibliographiques, m'a permis de produire trois serveurs :

- le premier sur *l'Eau en Allemagne* ;
- le second sur *l'Eau en Allemagne et pays limitrophes* ;
- le troisième sur *l'Eau en Allemagne et Tchécoslovaquie*.

Après plusieurs comparaisons entre le corpus d'images et les trois serveurs réalisés, le service SHS et moi-même avons conservé, avec l'agrément de PRODIG ([Annexe 3](#)), le serveur de notices sur *l'Eau en Allemagne et Tchécoslovaquie* soit un serveur contenant 227 références bibliographiques.

| Descripteurs français | nb docs |
|---|---------|
| Allemagne | 212 |
| Bayern | 34 |
| Précipitation | 32 |
| Bassin-versant | 31 |
| Erosion des sols | 28 |
| Pollution | 27 |
| Hydrologie | 26 |
| Cours d'eau | 26 |
| Pollution de l'eau | 25 |
| Sol | 23 |
| Modèle | 23 |
| Eau | 21 |
| Action anthropique | 21 |
| Fluviale | 18 |
| Protection de la nature | 17 |
| Utilisation du sol | 16 |
| Ecoulement | 16 |
| Baden-Württemberg | 15 |
| Quatenaire | 14 |
| Hessen | 14 |

Dans ce serveur de notices bibliographiques, les descripteurs français sont au nombre de 670 et le tableau, ci-contre, fournit les descripteurs les plus fréquents.

On constate que les descripteurs :

Allemagne (212)

Cours d'eau (26)

Eau (21)

qui correspondent à la requête

Eau Allemagne et Tchécoslovaquie

sont bien présents dans ce résultat statistique.

Les régions (länder) comme « Schleswig-Holstein » et les villes-états comme « Hambourg » et « Berlin » sont également bien représentées.

Fig. 10 Descripteurs français les plus fréquents

4. REALISATION DU SERVEUR D'INVESTIGATION « FIG 2003 »

Suite aux premiers travaux de conception et d'expérimentation, détaillés ci-dessus, le projet de réalisation du serveur regroupant des ressources iconographiques et bibliographiques en Géographie a pu se concrétiser. Toute la démarche suivie lors de cette expérimentation a facilité la mise en œuvre d'un nouveau prototype de serveur réunissant des documents iconographiques et bibliographiques.

4.1 Arborescence du serveur d'investigation « FIG 2003 »

L'arborescence du serveur d'investigation réalisé pour le Festival International de Géographie 2003 se compose de plusieurs répertoires :

- un répertoire nommé **Applications** où j'ai travaillé sur quelques fichiers nécessaires à l'application en géographie dans le répertoire **geo**;
- un répertoire nommé **TestIcon** où tous les fichiers du répertoire **Applications** sont régénérés automatiquement à l'aide de programmes en C et de shells Unix de lancement de programmes. C'est dans ce répertoire qu'il a fallu intervenir à plusieurs reprises pour créer le serveur Multibase regroupant ainsi le serveur d'investigation iconographique et bibliographique.

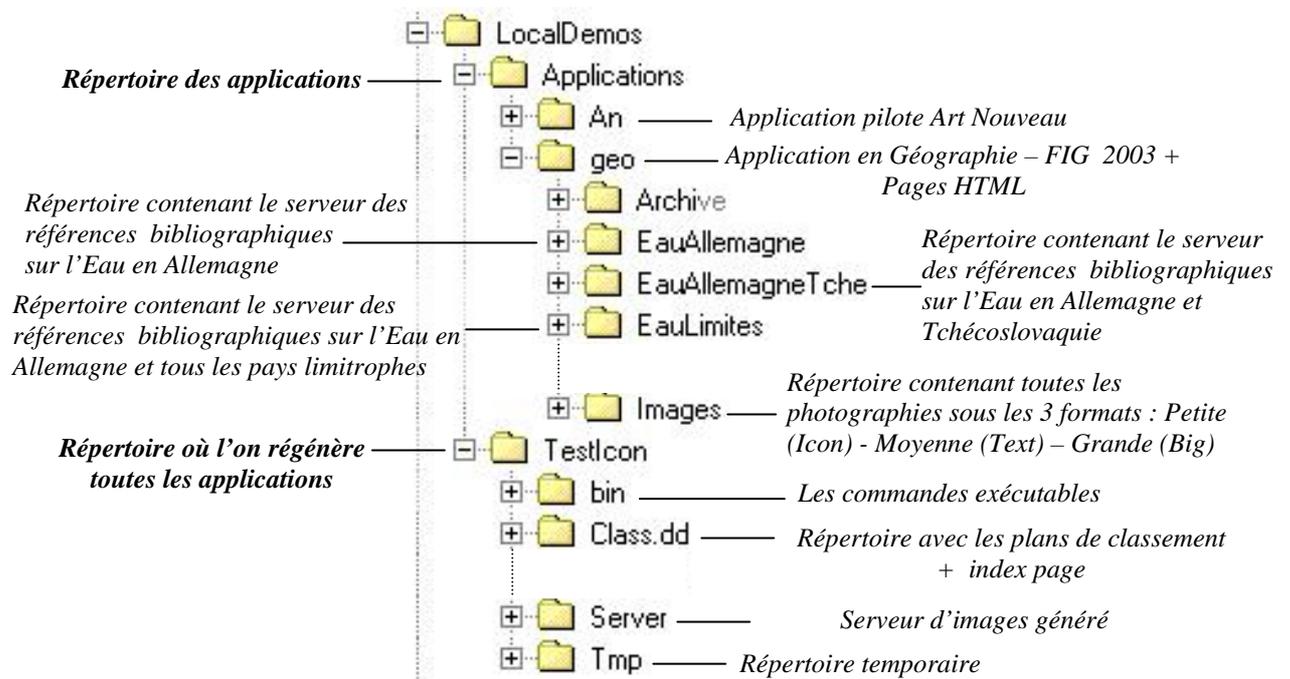


Fig. 11 Arborescence des applications

4.2 Mise en place d'un plan de classement

Lors de la création du fichier plan de classement, j'ai extrait tous les codes de classement disponibles dans le plan de classement de la base de données Géographie de FRANCIS afin de constituer une classification pour les notices d'indexation des photographies.

Ce plan de classement permet également une navigation plus structurée et indique à l'utilisateur un chemin cohérent.

| | |
|-------------------------------|-----------------------------------|
| 531 | Géographie |
| 531 – 153 | Europe |
| 531 – 153 – 239 | Allemagne |
| 531 – 153 – 239 – 243A | Ex-RFA |
| 531 – 153 – 239 – 244 | Etats rhénans – Nord-Ouest |
| 531 – 153 – 239 – 245 | Etats du Nord et de l'Est |
| 531 – 153 – 239 – 246 | Bavière – Etats du Sud |
| 531 – 153 – 239 – 246A | Berlin |
| 531 – 153 – 239 – 246B | Ex-RDA |
| 531 – 287 | Europe de l'Est |
| 531 – 287 – 294 | Ex-Tchécoslovaquie |

Fig. 12 Plan de classement élaboré à partir du plan de classement de la base de données FRANCIS pour l'application geo

A partir de ces codes de classement, l'on peut accéder à l'ensemble des 88 photographies regroupées d'après les codes de classement de BGI retravaillés pour obtenir des codes intermédiaires. Ainsi, l'on a :

- pour l'Allemagne :

Géographie - Europe - Allemagne - Etats du Nord et de l'Est

- sous le code $\overline{531} - \overline{153} - \overline{239} - \overline{245}$, les photographies des villes situées en Europe, en Allemagne et dans les **Etats du Nord et de l'Est** : soit **Brême**, **Hambourg**, **Kiel**, **Lübeck** ;

Géographie - Europe - Allemagne - Berlin

- sous le code $\overline{531} - \overline{153} - \overline{239} - \overline{246A}$, les photographies de la ville de **Berlin** située en Europe et en Allemagne ;

Géographie - Europe - Allemagne - Ex-RDA

- sous le code **531** – **153** – **239** – **246B**, les photographies des villes situées en Europe, en Allemagne, et en Ex-RDA : soit *Dresde* et *Messein*.

- pour la République Tchèque :

Géographie - Europe de l'Est - Ex-Tchécoslovaquie

- sous le code **531** – **287** – **294**, les photographies de la ville de *Prague* située en Europe de l'Est et en Ex-Tchécoslovaquie.

Le rangement des codes de classement s'est effectué dans le fichier *bgi.class.sgm* sous le répertoire « geo » de l'Application en géographie ce qui a facilité l'accès aux images. La création d'un shell a permis de situer toutes ces images sous chaque code de classement.

The image shows an XML snippet with several annotations:

- Record 1:** `<record key="531-1999">`
 - Annotation: "Correspond au plan de classement = 531 de la base de données Francis en *Géographie* de 1999"
 - Annotation: "Nouveau code : 003 pour FRANCIS – 531 pour Géographie" (pointing to `<code>/003/531</code>`)
 - Annotation: "Code de la BGI = 531" (pointing to `<bgi>531</bgi>`)
 - Annotations: "Géographie en français" and "Géography en anglais" (pointing to `<claslib lng="fre"><lib>Géographie</lib></claslib>` and `<claslib lng="eng"><lib>Geography</lib></claslib>`)
- Record 2:** `<record key="531153-1999">`
 - Annotation: "Correspond au plan de classement = 531153 de la base de données Francis en *Géographie* de 1999"
 - Annotation: "Nouveau code : 003 pour FRANCIS – 531 pour Géographie – 153 pour Europe" (pointing to `<code>/003/531/153</code>`)
 - Annotation: "Code de la BGI = 531153" (pointing to `<bgi>531153</bgi>`)
- Record 3:** `<record key="531162-1999">`
 - Annotations: "Europe en anglais" and "Europe en français" (pointing to `<claslib lng="eng"><lib>Europe</lib></claslib>` and `<claslib lng="fre"><lib>Europe</lib></claslib>`)

Commentaires :

Comme il existe un niveau qui correspond à PASCAL :

- 001** - Sciences exactes et technologie
- 002** - Sciences biologiques et médicales

il a été nécessaire de créer artificiellement un niveau général FRANCIS qui a été nommé :

- 003** - Sciences humaines et sociales

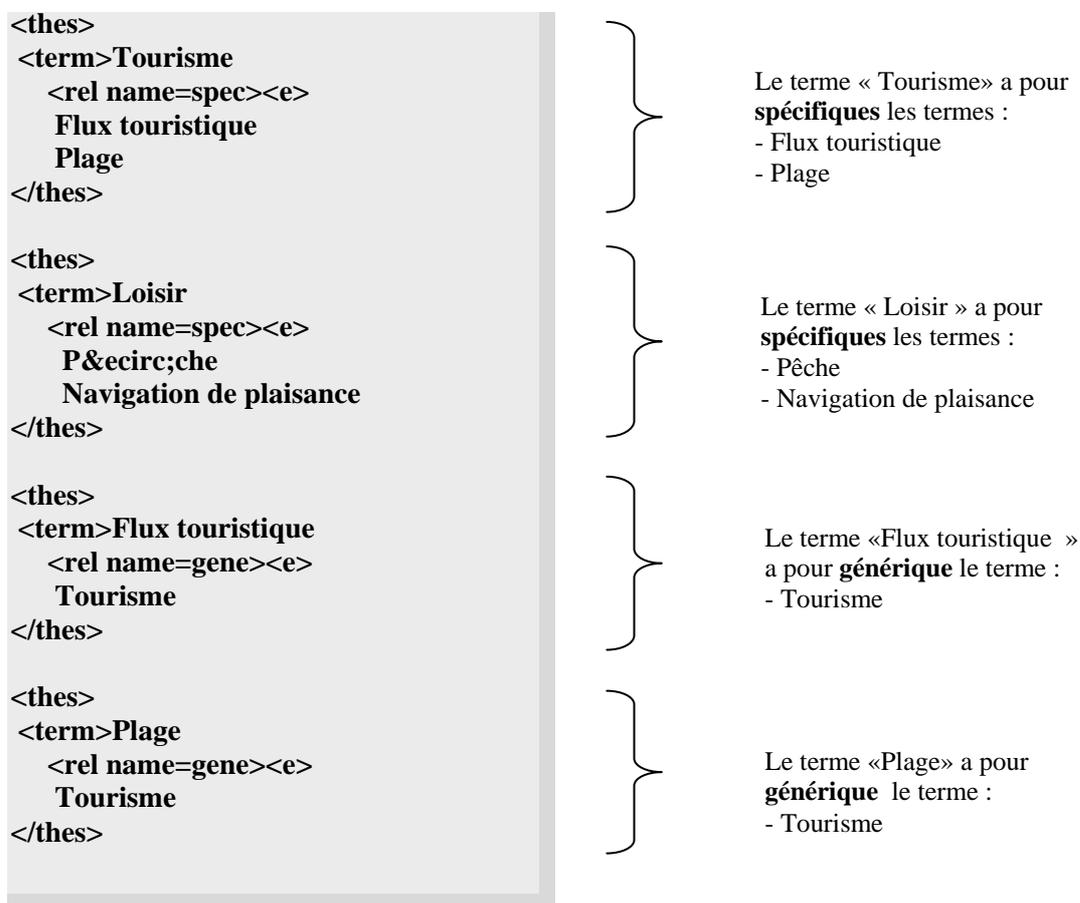
Fig. 13 Extrait du fichier *bgi.class.sgm* utilisé pour insérer le plan de classement dans le serveur

4.3 Création d'un thésaurus

Afin que la recherche d'images soit encore plus précise, la création d'un *thésaurus* s'est avérée indispensable pour établir un arbre de connaissance dans le domaine de la géographie. Ce langage documentaire est apparu d'autant plus nécessaire qu'il facilite l'accès rapide vers d'autres termes par l'intermédiaire de liens hypertextes.

Deux principales étapes pour construire ce thésaurus :

- 1- *La collecte* : en confrontant la liste des descripteurs la base iconographique et la base bibliographique, j'ai collecté les termes communs aux deux bases ;
- 2- *La mise en hiérarchie* : à l'aide des termes retenus, j'ai hiérarchisé ceux-ci en termes spécifiques (TS) ou termes génériques (TG).



Suite fichier

Fig. 14 Extrait du fichier *geo.FD.thes.ed* utilisé pour insérer le thésaurus dans le serveur

Pour construire ce thésaurus, je me suis aidée d'un thésaurus en ligne sur le site du Ministère de l'Équipement, des Transports, du Logement, du Tourisme et de la Mer : [Urbamet](http://thesurb.documentation.equipement.gouv.fr/)¹⁶. Celui-ci a dissipé quelques unes de mes hésitations et a facilité mon travail de hiérarchisation.

¹⁶ <http://thesurb.documentation.equipement.gouv.fr/>

4.4 Création du serveur d'investigation « BIBALEau »

BIBALEau (Base Iconographique et Bibliographique sur l'ALlemagne et l'Eau) est le nom qui a été retenu pour nommer le serveur d'investigation Multibase sur le thème de l'Eau en Allemagne et République Tchèque.

Afin de créer ce nouveau prototype et d'après le modèle du serveur iconographique et bibliographique **BIBAN**, nous avons récupéré les données des deux serveurs (l'un iconographique, l'autre bibliographique) pour obtenir **BIBALEau**.

La génération du serveur contenant les deux corpus préalablement préparés lors des phases de conception, est lancée à partir de shells dans le fichier *testIcon.sh* sous le répertoire TestIcon. Ce dernier définit les chemins d'accès nécessaires au lancement et à la construction des bases et définit les ressources. Il paramètre le chemin de navigation.

4.5 Création d'un dictionnaire

Afin d'améliorer l'ergonomie et la navigation de ce nouveau serveur, un travail de présentation des différentes pages dynamiques s'est avéré nécessaire. Aussi, il a fallu reprendre dans un dictionnaire nommé *bibaleau.FR.dic* certains des intitulés peu compréhensibles pour cette application. Pour cela, la récupération des clés, nécessaires à la reconnaissance des titres, a permis de créer ou de remplacer des titres sur toutes les pages du serveur.

A titre d'exemple, le titre :

Serveur d'Investigation Pascal Monobase a été remplacé par le titre **Serveur BIBALEau**.

Les termes ou expressions utilisés dans ce dictionnaire sont principalement des titres ou groupes de mots dont la fréquence d'utilisation dans le serveur est assez élevée et qui sont supposés être réemployés. Les clés symboliques définissent des éléments d'interface (titre, item de menu) et sont mises en correspondance avec les éléments d'interface spécifiques à l'application. L'ensemble de ces correspondances constitue le dictionnaire. Ce dictionnaire est ensuite utilisé pour créer des pages statiques (accueil) et dynamiques fabriquées avec des CGI (Common Gateway Interface). Ces pages sont personnalisées pour les besoins de l'application.

| <i>Clés</i> | <i>Correspondances</i> |
|---------------------|------------------------------------|
| ... | |
| MULTI/longName | Base Commune |
| homePage/select | Moteur de recherche |
| Cgi/listDoc/n/long | Nombre de documents traités |
| Cgi/listDoc/n/short | Nb de documents |
| ... | |

Fig. 15 Extrait du dictionnaire français : fichier « bibaleau.FR.dic »

5. FONCTIONNEMENT DU SERVEUR « BIBALEau »

Le serveur BIBALEau a été conçu de manière à accéder à trois serveurs : l'un iconographique, l'autre bibliographique, puis le serveur mixte. Mais avant d'interroger les serveurs d'investigation, l'on peut accéder directement aux photographies par l'intermédiaire d'une carte d'Allemagne cliquable.

5.1 Présentation des photographies

A l'aide d'un fond de carte de l'Allemagne retravaillé avec le logiciel de retouches d'images Paint et le logiciel MapExpert Pro 2.0 pour obtenir des zones cliquables sur chaque ville ou région hachurée, l'on peut visualiser les photographies des villes citées dans le plan de classement. Cette page a été conçue pour accéder directement aux images sans passer par le serveur d'investigation iconographique.



Fig. 16 Carte d'Allemagne, d'une partie de la République Tchèque et des cours d'eau avec des pages HTML

La conception des pages HTML sur le thème de l'Eau dans les différentes villes s'est fait d'un commun accord avec l'équipe du projet en ce qui concerne la charte graphique et la mise en page. Nous avons également souhaité que ces photographies soient présentées à la manière d'un album-photo d'où le choix de ne garder que quatre photographies par page. De plus, en cliquant sur chaque photographie, l'on obtient une version agrandie (*big*) de la photographie.

5.2 Fonctionnement des serveurs d'investigation

A partir de la page d'accueil ([Annexe 4](#)), le lien sur BIBALEau nous permet d'accéder à la carte de l'Allemagne et aux trois serveurs d'investigation. Ces serveurs nous donnent la possibilité d'explorer soit :

- un corpus iconographique ;
- un corpus bibliographique ;
- un corpus commun appelé multibase.

• Pour les trois serveurs, *l'accès par navigation* indique :

- les auteurs « auteurs » ;
- les descripteurs français « FD » ;
- les descripteurs anglais « ED » (*excepté la base image*) ;
- les mots du titre en français « FTI » ;
- les mots du titre en anglais « ETI ».

• *L'accès par navigation*, à partir d'un index (auteurs, descripteurs, mots du titre) facilite la recherche des photographies et/ou des notices bibliographiques (en fonction du serveur choisi).

• *L'accès par moteur de recherche* permet une recherche sur chaîne de caractères (pour les spécialistes permet une recherche par les index).

L'exploration est rapide et les clusters révèlent des associations internes.

Détail des fonctions proposées :

- les descripteurs les plus fréquents ;
- la liste des clusters – nombre de co-occurrences ;
- la liste des clusters – distances des cosinus ;
- la liste des associations les plus fréquentes ;
- l'exploration par sélection d'un descripteur.

La même navigation peut se poursuivre depuis des clusters obtenus selon les méthodes précitées, ou la liste des associations les plus fréquentes.

5.2.1 Le serveur iconographique



Ce serveur donne accès à un résultat statistique réalisé grâce à l'analyse bibliométrique effectuée par DILIB. Le corpus traité comprend **88** photographies.

Liens vers les différents Index :

- Descripteurs français
- Auteurs
- Mots du titre en anglais
- Mots du titre en français

A titre d'exemple, voici les premiers liens disponibles lorsqu'on navigue à partir de la fonction des descripteurs les plus fréquents.

| Descripteurs Français | Nb de documents |
|---------------------------------------|-----------------|
| Eau | 88 |
| Allemagne | 79 |
| Cours d'eau | 46 |
| Paysage urbain | 37 |
| Elbe | 22 |
| Hamburg | 20 |
| Hambourg | 20 |
| Hambourg (ville-état) | 20 |
| Berlin | 18 |
| Berlin (ville-état) | 18 |
| Spree | 16 |
| Port | 16 |
| Saxe | 15 |
| Sachsen | 15 |
| Schleswig-Holstein | 13 |
| Architecture | 13 |
| Canal | 12 |
| Ville | 11 |
| Espace bâti | 11 |
| Trafic | 10 |
| Pour continuer... | |

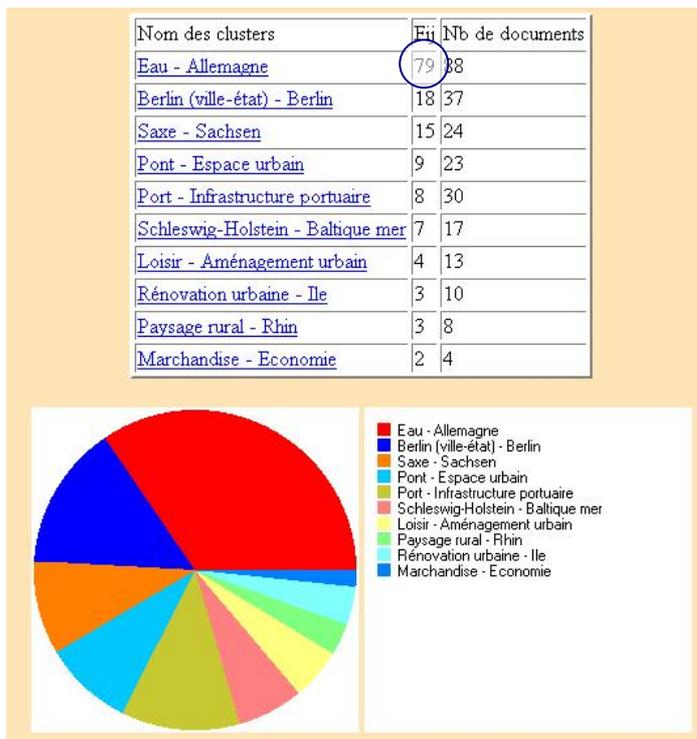
| Descripteurs Français | Nb de documents |
|--|-----------------|
| Rénovation urbaine | 10 |
| Quartier | 10 |
| Pont | 10 |
| Espace urbain | 10 |
| Tchèque, république | 9 |
| Praha | 9 |
| Prague | 9 |
| Photographie aérienne | 9 |
| Meissen | 9 |
| Infrastructure portuaire | 9 |
| Loisir | 8 |
| Centre-ville | 8 |
| Transport fluvial | 7 |
| Tourisme | 7 |
| Paysage rural | 7 |
| Baltique mer | 7 |
| Vltava | 6 |
| Lübeck | 6 |
| Dresden | 6 |
| Dresde | 6 |
| Végétation | 5 |
| Pour continuer... | |

– 1^{er} résultat –

Liste des 20 premiers descripteurs français les plus fréquents.

Sur 88 photographies :

- 79 traitent de l'Allemagne ;
- 9 traitent de la République Tchèque.



– 2^{ème} résultat –

Mesure utilisée : nombre de co-occurrences

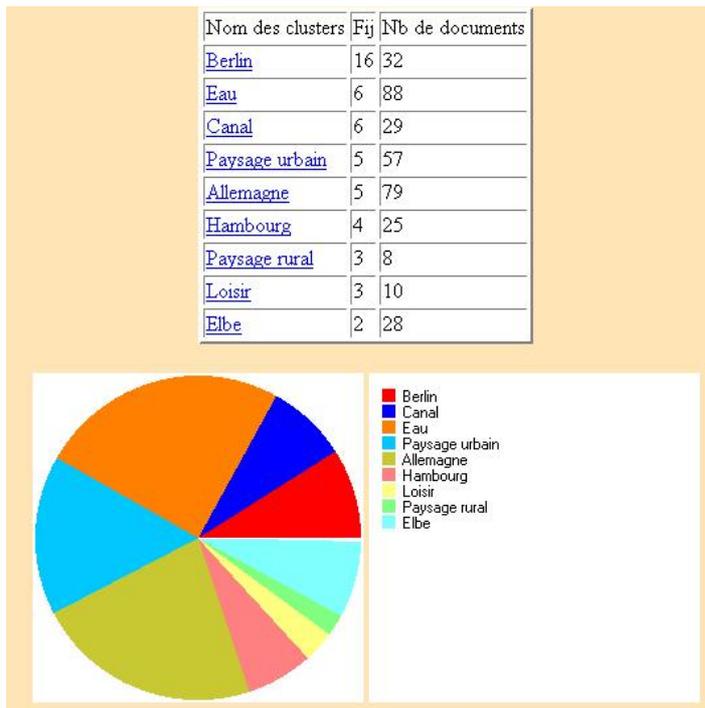
La **co-occurrence** est la présence simultanée de deux ou plusieurs éléments ou classes d'éléments dans le même énoncé.

Ici, le nombre de co-occurrences, appliqué au traitement et à l'analyse des photographies, correspond au nombre de fois où deux mots sont utilisés ensemble dans un document.

Par exemple :

- les 2 mots *Eau* et *Allemagne* sont utilisés ensemble 79 fois.

– 3^{ème} résultat –



Liste des clusters - Mesure utilisée : distance des cosinus

La force des associations entre les termes est rapportée au nombre de notices concernées par la clusterisation.

Par rapport à la méthode des associations simples, cette méthode permet :

- de relativiser l'importance des associations, dont les plus fréquentes ne sont pas forcément les plus intéressantes ;
- de mettre en évidence des "clusters" intéressants mais moins volumineux.

Cette mesure permet ainsi de faire remonter les signaux faibles, mais son inconvénient est qu'elle permet aussi de faire remonter du "bruit".

– 4^{ème} résultat –

Liste des associations les plus fréquentes

Liste des associations de mots associés entre eux par ordre décroissant :

- Eau et Allemagne sont présents **79** fois ;
- Eau et Cours d'eau sont présents **46** fois...

| Association | Fij |
|---|-----|
| Eau - Allemagne | 79 |
| Eau - Cours d'eau | 46 |
| Cours d'eau - Allemagne | 39 |
| Paysage urbain - Eau | 37 |
| Paysage urbain - Allemagne | 34 |
| Elbe - Eau | 22 |
| Paysage urbain - Cours d'eau | 21 |
| Elbe - Allemagne | 21 |
| Hamburg - Hambourg (ville-état) | 20 |
| Hamburg - Hambourg | 20 |
| Hamburg - Eau | 20 |
| Hamburg - Allemagne | 20 |
| Hambourg (ville-état) - Hambourg | 20 |
| Hambourg (ville-état) - Eau | 20 |
| Hambourg (ville-état) - Allemagne | 20 |
| Hambourg - Eau | 20 |
| Hambourg - Allemagne | 20 |
| Eau - Berlin (ville-état) | 18 |
| Eau - Berlin | 18 |
| Berlin (ville-état) - Berlin | 18 |

5.2.2 Le serveur de références bibliographiques



Le corpus traité comprend **227** références bibliographiques.

Liens vers les différents Index :

- Auteurs
- Descripteurs français
- Descripteurs anglais
- Mots du titre en français
- Mots du titre en anglais

La présentation des résultats ne diffère pas de la précédente, mais cependant elle apporte un complément d'information avec l'intégration des bases d'arrière-plan de la BGI des années 1999 – 2000 – 2001 – 2002 – 2003. Ces bases apportent ainsi une nouvelle fonctionnalité au serveur bibliographique en proposant pour chaque terme retenu une représentation du terme sur plusieurs années (ont été retenus les cinq dernières années).

| Nom de la base | Nb de documents |
|----------------|----------------------------|
| BGI 2003 | 61 / 1939 |
| BGI 2002 | 193 / 3731 |
| BGI 2001 | 182 / 3773 |
| BGI 2000 | 152 / 3883 |
| BGI 1999 | 180 / 3942 |
| Toutes | 768 |

Ce corpus d'antériorité permet :

- de voir si le terme est stable d'une année à l'autre ;
- de s'assurer s'il est toujours utilisé.

Sur l'exemple ci-contre, le terme exploré est *Allemagne*. Celui-ci a subi une baisse d'usage en 2000 ; il est de nouveau fortement présent en 2001 et 2002. Pour 2003, il faut signaler qu'il s'agit d'une demi-année (soit sur 6 mois d'indexation).

| Terme | Allemagne |
|-----------------------|---|
| spec | Baden-Württemberg |
| | Basse-Saxe |
| | Bavière |
| | Bayern |
| | Berlin (ville-état) |
| | Brême (ville-état) |
| | Danube |
| | Elbe |
| | Hambourg (ville-état) |
| | Main |
| | Moselle |
| | Rhénanie du Nord-Westphalie |
| | Rhénanie-Palatinat |
| | Rhin |
| | Saxe |
| | Schleswig-Holstein |
| | Spree |
| Trave | |
| Weser | |
| gene | Europe |
| englishTerm | Germany |

Tout comme pour le serveur iconographique, ce serveur permet de naviguer dans le **thésaurus** construit en amont et offre ainsi une meilleure connaissance de l'environnement du document.

Ainsi le terme *Allemagne* a bien :

- pour *générique* Europe [**gene**] ;
- pour *spécifique* toutes les régions de l'Allemagne présentes dans les deux corpus (iconographique et bibliographique) ainsi que tous les cours d'eau qui la traversent [**spec**].

Légende (non représentée dans le serveur) :

- = Cours d'eau de l'Allemagne
- = Région et ville-état de l'Allemagne

5.2.3 Le serveur commun - multibase

SERVEUR BIBALEau
Base Commune

Nombre de documents traités: 315

| Index | Code index |
|--------------------------|---------------------|
| Auteurs | AU |
| Descripteurs anglais | ED |
| Descripteurs français | FD |
| Mots du titre (anglais) | ETI |
| Mots du titre (français) | FTI |

Le corpus traité comprend **315** documents, dont 88 photographies et 227 références bibliographiques.

Liens vers les différents Index :

- Auteurs
- Descripteurs anglais
- Descripteurs français
- Mots du titre en anglais
- Mots du titre en français

Ce serveur est le résultat du mixage des deux précédents serveurs. Il a l'avantage d'offrir les résultats statistiques regroupant deux types de documents et apporte aux utilisateurs une nouvelle navigation en regroupant différents types de ressources.

Autre fonctionnalité offerte : possibilité de retrouver le terme *Allemagne* dans les autres bases :

- le nombre de documents présents sur *l'Allemagne* dans la Base Image : 79 ;
- le nombre de documents présents sur *l'Allemagne* sur la Base de références bibliographiques : 212 ;
- le nombre de documents présents sur *l'Allemagne* dans la Base commune : 291.

| Nom de la base | Nb de documents |
|---|-----------------|
| Base Image | 79 |
| Base de références bibliographiques | 212 |
| Base Commune | 291 |

Navigation vers chacune des bases

6. BILAN DE LA MISSION

6.1 Difficultés rencontrées

D'un point de vue personnel, la première des difficultés a été l'apprentissage des systèmes informatiques et des outils nécessaires à la réalisation du projet. La mise en place d'un programme en LEX et en langage C m'a contraint à m'exercer et à travailler par essais successifs.

En ce qui concerne cette étude spécialisée, réalisée sur un corpus d'images provenant du fonds iconographique de PRODIG et indexées d'après le vocabulaire utilisé dans la base de données Géographie de FRANCIS, les difficultés sont intervenues lors de la récupération du « fichier indexation ».

Sur ce fichier :

d'un point de vue technique :

- les consignes d'écriture n'ont pas été toujours respectées ce qui a provoqué des « coquilles » dans ce fichier (par exemple : majuscules ou minuscules oubliées) ;
- j'ai également constaté des problèmes de lecture du fichier avec le logiciel de traitement de texte utilisé chez PRODIG (Word). Lors de sa lecture en mode texte, sous Xemacs, les espaces n'étaient plus respectés.

d'un point de vue documentaire :

- les descripteurs utilisés dans l'indexation étaient parfois différents de ceux utilisés dans le vocabulaire contrôlé de la base de données Géographie de FRANCIS ;
- des oublis de mots-clés, tel que les régions dans lesquelles les villes étaient situées ;
- des doublons dans l'envoi des photographies.

Ces difficultés ont quelque peu perturbé la poursuite du projet. Ainsi et, à titre d'exemple, dans le champ mot-clé, l'ajout d'une lettre m'a bloqué quelques jours avant de comprendre la raison pour laquelle je ne pouvais visualiser toutes les photographies enregistrées sur le serveur.

Aussi et afin d'avoir une vue d'ensemble sur toutes les erreurs qui avaient pu se glisser dans ce fichier, j'ai récupéré tous les mots-clés les plus fréquents dans le serveur de photographies ([Annexe 5](#)).

C'est ainsi que je me suis aperçue que pour la République Tchèque, l'indexeur avait retenu quatre descripteurs :

- Tchèque (République)
- Tchèque République
- Tchèque (république)
- République (Tchèque).

Alors qu'il ne fallait retenir que celui de : Tchèque, république (terme retenu dans le vocabulaire contrôlé de la base de géographie de FRANCIS).

Il faut toutefois préciser que toutes ces erreurs se sont produites uniquement sur le corpus d'images lors d'une indexation manuelle, car lorsque les notices bibliographiques sont indexées avec le logiciel PSILOG, il ne peut y avoir de confusion parce que le logiciel ne retient que le descripteur contenu dans le vocabulaire de la base de données Géographie de FRANCIS. L'on peut se rendre compte ici des limites d'une indexation non-automatisée et non-informatisée.

6.2 Les résultats obtenus

D'une part, ma mission aura permis d'expérimenter un nouveau mode d'exploitation de corpus documentaire autre que bibliographique en instaurant une méthodologie de travail réutilisable lors de la mise en œuvre d'un serveur de ressources iconographiques et bibliographiques.

D'autre part, avec ce prototype présenté au FIG 2003, l'INIST aura valorisé la recherche en géographie et sera en mesure d'élargir sa gamme de produits en proposant cette nouvelle application. Enfin, ce prototype aura été la première pierre de la future "Base d'images" qui, pour l'instant, englobent les cartes postales de Jacques DUCLOY et les photographies en géographie de PRODIG à condition cependant de lever tous les problèmes de droits.

6.3 Les limites du projet

Il ne faut cependant pas oublier que ce nouveau produit reste encore en l'état, un prototype, qui ne demande qu'à être amélioré. Aussi, il faudra sûrement encore quelques mois avant que cette expérience-pilote aboutisse en un produit INIST.

6.4 Les améliorations possibles

Pour que cette étude spécialisée sur un serveur d'investigation des ressources iconographiques et bibliographiques soit rapidement exploitable et réutilisable, il faudra inclure quelques améliorations indispensables à la mise en œuvre d'un produit INIST.

✎ *Pour le fichier texte* qui contient toutes les notices, il faudra prévoir un *formulaire* avec un contrôle syntaxique sur le vocabulaire contrôlé de la base pour rentrer les données afin de récupérer un fichier « propre » sans erreurs ou fautes de frappe.

✎ *Pour les photographies*, il faudra prévoir un traitement d'images automatisé à partir d'un script. Avec celui-ci, on pourra générer automatiquement les trois formats d'images retenus ce qui permettra de traiter un grand nombre d'images et de ce fait, sera un gain de temps pour la réalisation du produit.

✎ *Pour les pages HTML*, il faudra également rendre possible la génération automatique des pages HTML.

6.5 Les perspectives à court terme

La réalisation de ce projet a donné la possibilité au DPS de l'INIST et, en particulier au service SHS, d'être à même de proposer un prototype qui traite de notices iconographiques et bibliographiques grâce à la plate-forme logicielle DILIB.

C'est ainsi que j'ai eu l'occasion de rencontrer (*lors de 8^{ème} Festival du Film de chercheur qui s'est tenu à Nancy du 27 mai au 6 juin 2003*) Laurence POURCHEZ, *Ethnologue* au CNRS à qui j'ai pu présenter ce prototype. Très intéressée, lors de ma démonstration, par les premiers résultats statistiques du serveur en cours de réalisation, cette scientifique a souhaité que le service SHS lui produise un outil qui lui permette d'avoir une approche plus synthétique des centaines de photographies qu'elle possède sur le thème de « *la Religion à la Réunion* ».

Nous lui avons expliqué la procédure qu'elle devrait adopter pour que ce travail soit réalisé par le service. Aussi, avant de traiter un tel corpus, soit environ 1000 photographies, nous lui avons conseillé de nous confier une vingtaine de photographies en lui demandant de nous fournir une indexation pour chacune des celles-ci. En effet, il nous semblait indispensable que l'indexation de ces documents iconographiques soit réalisée par la personne qui avait effectué les prises de vue afin d'obtenir une indexation scientifique. Nous lui avons également fourni le modèle de notices d'indexation présenté ci-dessus dans la phase de conception du serveur et transmis le vocabulaire d'indexation de la base de données d'Ethnologie de FRANCIS.

Dernièrement et juste avant de conclure ce rapport, j'ai reçu les premières indexations et photographies correspondant au corpus demandé et avant mon départ, je vais essayer de mettre en œuvre le mode opératoire élaboré lors de la conception de l'application en géographie afin de vérifier si toutes les phases de réalisation sont applicables pour ce nouveau corpus de ressources iconographiques et bibliographiques en Ethnologie.

CONCLUSION

L'application en géographie réalisée à l'occasion de mon stage à l'INIST pour le Festival International de Géographie de Saint-Dié-des-Vosges est le résultat d'un travail effectué en étroite collaboration avec le partenaire PRODIG, mais aussi avec tout l'encadrement technique du service SHS et du DPS. Ces deux composantes ont facilité mon adaptation aux techniques requises pour concevoir et réaliser un serveur d'investigation mixte (images et références bibliographiques). La création des premiers serveurs bibliographiques lors de la période de test m'a permis de me rendre compte des différentes possibilités du logiciel DILIB et m'a ainsi aidé à son appropriation.

La phase d'expérimentation de ce prototype a donné lieu à la mise en œuvre d'une méthodologie pour automatiser et réaliser des serveurs d'investigation à base de documents électroniques issus de fonds documentaires divers, mais a également révélé les différentes difficultés d'exécution de l'application lors de la mise en place du serveur commun. A longue échéance, l'aboutissement de ce projet devrait permettre d'intégrer cette réalisation dans une base d'image INIST selon un plan de classement précis et constituera une nouvelle ressource de documentation tout en valorisant le travail des chercheurs.

D'un point de vue personnel, ce stage a contribué à approfondir mes connaissances en informatique documentaire (RDF-XML) et m'a familiarisée avec les techniques de programmation. Le travail en partenariat avec PRODIG et les différentes équipes m'a donné l'opportunité de tester mes capacités à gérer le travail avec plusieurs intervenants tout en respectant les objectifs de la mission. Travaillant pour une activité de Recherche et de Développement, j'ai également pu évaluer et répondre aux attentes de l'équipe en matière de créativité et de rigueur dans le travail.

Suite à cette expérimentation, une nouvelle approche d'analyse des photographies va se concrétiser ces prochaines semaines par l'utilisation d'un modèle informatique qui s'inspire des réseaux de neurones biologiques : *les réseaux de neurones*. L'idée principale est d'obtenir une représentation des photographies en fonction des couleurs, des formes détectées, des noms attribués... Cette étude devrait déboucher sur une présentation quelque peu différente des méthodes de restitution des photographies, mais semble une suite logique de la poursuite de ce projet.

TABLE DES SIGLES & ABREVIATIONS

| | |
|-----------------|---|
| ADBS | Association des professionnels de l'information et de la documentation |
| ABES | Agence Bibliographique de l'Enseignement Supérieur |
| BIBALEau | Base Iconographique et Bibliographique sur l'Allemagne et l'Eau |
| BIBAN | Base Iconographique et Bibliographique Art Nouveau |
| BGI | Bibliographie Géographique Internationale |
| CGI | Common Gateway Interface |
| CNRS | Centre National de la Recherche Scientifique |
| DILIB | Document & Information LIBrary |
| DOGE | Documentation Automatisée en Gestion des Entreprises |
| DOM | Document Object Model |
| INIST | INstitut d'Information Scientifique et Technique |
| INPI | Institut National de la Propriété Industrielle |
| IST | Information Scientifique et Technique |
| FIG | Festival International de Géographie |
| HFD | Hierarchical File organisation for Documentation |
| HTML | HyperText Markup Language |
| JPEG | Join Photographic Experts Group |
| LORIA | Laboratoire Lorrain de Recherche en Informatique et ses Applications |
| LORIASI | LORIA dans le Société de l'Information |
| MARC | Machine Readable Cataloguing |
| MIRIAD | Module Intranet de Recherche d'Information et d'Analyse Documentaire |
| PRODIG | Pôle de Recherche pour l'Organisation et la Diffusion de l'Information Géographique |
| RDF | Resource Description Framework |

| | |
|-------------|--------------------------------------|
| SGML | Standard Generalized Markup Language |
| TIFF | Tagged Image File Format |
| UMR | Unité Mixte de Recherche |
| UPS | Unité Propre de Service |
| W3C | World Wide Web Consortium |
| XML | eXtensible Markup Language |

GLOSSAIRE

- Bibliométrie** Application des méthodes statistiques et mathématiques aux livres et autres médias de communication. Elle consiste en l'application d'outils de statistique descriptive pour analyser notamment la distribution et la fréquence des données bibliographiques (champs date, langue, affiliation...) et/ou textuelles (mots-clés). Elle comprend deux volets principaux : les lois bibliométriques et l'analyse des citations.
- CGI** Common Gateway Interface – Petit programme informatique, écrit en langage script, qui permet de réaliser des pages dynamiques et d'interagir avec le serveur Web. Il s'agit en fait d'une convention entre le serveur Web et un programme qu'il appelle pour traiter une requête utilisateur.
- Cluster** Ensemble d'unités présentant une corrélation plus forte entre elles qu'avec les unités des groupes voisins. Agrégat d'éléments résultant d'une méthode de classification automatique.
- Cooccurrence** Présence simultanée de deux ou plusieurs éléments ou classes d'éléments dans le même énoncé.
- Descripteur** Mot ou groupe de mots retenu dans un thésaurus (ou vocabulaire contrôlé) et choisi parmi un ensemble de termes équivalents pour représenter sans ambiguïté une notion apparaissant dans un document ou dans une demande de recherche documentaire.
- Distance des cosinus** Mode de calcul mesurant l'intensité de l'association réalisée entre deux mots i et j par l'ensemble des documents du fichier.
- DILIB** Document & Information LIbrary - Plate-forme documentaire développée par l'INIST et le LORIA servant à l'ingénierie des documents et en particulier à générer des serveurs d'investigation. Ce prototype fonctionne sur des données XML et des données hétérogènes.
- Dublin Core** Elaborée depuis 1995, le standard sur lequel la communauté internationale se met d'accord concernant la production de métadonnées exploitées par les moteurs de recherche. Le Dublin Core définit ainsi un ensemble de 15 éléments (title, author, subject, description, format...) constituant une base (core) de métadonnées générales, indépendantes du domaine d'application.
- Fichiers d'associations** Fichier permettant la mise en relation de termes co-occurents dans un même document.
- Fichier Inverse** Fichier d'indexation permettant la gestion des relations entre les termes d'un index et les notices qui les contiennent.

| | |
|-----------------------------------|---|
| HFD | Hierarchical File organisation for Documentation. Cette structure est utilisée par DILIB pour stocker les fichiers inverses et les fichiers d'associations. Cette structure permet de traiter en standard jusqu'à un million de références bibliographiques réparties en 100 répertoires de 100 fichiers contenant chacun 100 références. |
| HTML | HyperText Markup Language – Format ou langage permettant l'échange de données sur le web et qui définit la présentation graphique des pages par un système de balisage des documents. |
| Index | Document secondaire présentant une liste ordonnée de termes choisis figurant dans un document avec une indication permettant de les y localiser. |
| Indexation | Opération qui consiste à décrire et caractériser un document à l'aide d'une analyse des concepts contenus dans ce document. |
| Infométrie | Terme adopté en 1987 par la FID (Fédération Internationale de la Documentation) pour désigner l'ensemble des activités métriques relatives à l'information, couvrant aussi bien la bibliométrie que la scientométrie. |
| LEX | LEX est un générateur d'analyseur lexical qui exploite un code source contenant des instructions en langage C complété par un ensemble de règles de reconnaissance lexicale pour produire un programme C. Permet de traiter des chaînes de caractères. |
| MIRIAD | Module Intranet de Recherche d'Information et d'Analyse Documentaire – Permet la consultation de notices déchargeables dans plusieurs formats dont SGML. |
| Métadonnées | Les métadonnées sont des « données structurées à propos des données », parfois un substitut permettant de décrire le contenu d'un document pour le retrouver à partir de recherches sur Internet. |
| Mot-clé | Mot ou groupe de mots choisi : soit dans le titre ou le texte d'un document, soit dans une demande de recherche documentaire, pour caractériser le contenu. |
| Pixel | Plus petite partie homogène constitutive d'une image enregistrée par un support informatique. Le nombre de pixels par pouce (ppi ou dpi) est l'unité de mesure de la résolution d'une image numérique. Un écran standard affiche dans la résolution 800 x 600 pixels ou 1024 x 768 pixels. |
| Mots-associés (Méthode) | Cette méthode repose sur la comptabilisation des co-occurrences de mots indexant les documents. En comptabilisant toutes les co-occurrences et en représentant les liens entre mots qu'elles tracent, il est possible d'obtenir un réseau de mots-associés, l'existence des liens et leur force dépendant du nombre de co-occurrences entre mots dans les différents documents. A l'aide d'une classification hiérarchique ascendante, cette méthode découpe le réseau en clusters (ou agrégats) et permet d'identifier les thèmes ou problèmes de recherche et les liens qui les unissent (un thème correspondant à un agrégat). |

| | |
|-------------------------------|--|
| RDF | Resource Description Framework (Structure de description des ressources) permet de prendre en compte des métadonnées d'origines différentes, en utilisant plusieurs domaines nominaux. |
| Serveur (on line) | Organisme exploitant un système informatique permettant à un demandeur de consulter et d'utiliser directement une ou plusieurs banques de données fournies par différents producteurs. Il assure la commercialisation de ces produits et développe un logiciel commun pour leur interrogation. |
| Serveur (sur Internet) | Ordinateur offrant des données consultables à distance par d'autres ordinateurs. |
| SGML | Standard Generalized Markup Language – C'est un langage structuré normalisé international pour la documentation technique (ISO 9979 : 1986) qui permet de décrire la structure logique d'un document. |
| Shell | Langage de script utilisé sous Unix. Ce langage permet d'écrire des petits programmes appelés script shells. |
| Thésaurus | Langage documentaire organisé employant des termes normalisés reliés entre eux par des relations sémantiques hiérarchisées. |
| UNIX | Système d'exploitation créé par Bell, autrefois réservé à de grosses plates-formes mais aujourd'hui accessible sur un ordinateur. |
| W3C | World Wide Web Consortium – Les industriels s'y accordent sur les standards du Web. |
| XML | eXtensible Markup Language - Format d'échange et de structuration des documents. C'est une version allégée de SGML (Standard Generalized Markup Language) qui permet de définir des entités abstraites. C'est un langage de balisage qui permet une description structurelle, physique et logique d'un document. |

BIBLIOGRAPHIE

Articles de périodiques

- [Art. 1] ADBS. Dossier : Le droit de l'image. *Actualités du droit de l'information*, 2002, n° 26, p. 1-2
- [Art. 2] ADBS. Dossier : Le domaine public. *Actualités du droit de l'information*, 2002, n° 27, p. 1-2-4
- [Art. 3] ADBS. Foire aux questions. L'œuvre photographique et les droits des photographes. *Actualités du droit de l'information*, 2001, n° 10, p. 4

Dictionnaires

- [Dic. 1] Le Petit Robert de la langue française. Paris : Le Robert, 2001

Livres

- [Liv. 1] CHARBONNEAU, N., ROBERT, M. *La gestion des archives photographiques*. Sainte-Foy (Québec) : Presse de l'Université du Québec, 2001. 306 p.
- [Liv. 2] INRIA. *Bibliothèques numériques : cours INRIA, 9-13 octobre 2000, La Bresse*. Paris : ADBS, 2000. 246 p.
- [Liv. 3] JAKOBIAK, F. *L'information Scientifique et Technique*. Paris : PUF, 1995. 126 p.
- [Liv. 4] JOLION, J.M. *L'indexation*. Paris : Hermès Science publications, 2001. 182 p.
- [Liv. 5] KATTING, Cécile. *Gestion et diffusion d'un fonds d'images*. Paris : Nathan-ADBS, 2002. 127 p.

Rapports de stage

- [Rap.1] HEBTING, C. Rapport de stage de DESS IST-IE : Enrichissements du serveur BIBAN – Base Iconographique et Bibliographique Art Nouveau – conçu à partir de la plate-forme logicielle DILIB. Nancy 1, Nancy 2, INPL : LORIA, 1999. 60 p.
- [Rap.2] SALAVERT, A. Rapport de stage de DESS IST-IE : Rapport de Stage Projet Revelec : Réalisation d'une Plate-Forme d'Accès à des Revues électroniques. Nancy 1, Nancy 2, INPL : INIST, 2002. 73 p.
- [Rap.3] VIDAL, S. Rapport de stage de DESS IST-IE : Nouvelles Fonctionnalités du Serveur d'Investigation Transcriptome : Ajout d'un sous-ensemble du Méta-Thésaurus UMLS – Ajout de base d'arrière-plan. Nancy 1, Nancy 2, INPL : INIST, 2002. 60 p.

Ressources Internet

- [Res. 1] DUFFING, G. « L'indexation des images » *LORiatech* [En ligne] http://loriatech.loria.fr/LIBRE/2002_1024/EXPOSES/G-Duffing.ppt (pages consultées le 2 juin 2003)
- [Res. 2] DUFFING, G. « Approche thématique Thématico-Visuelle pour l'organisation et l'interrogation interactive d'une collection d'images hétérogènes » LORIA, équipe L&D Strasbourg [En ligne] <http://cdsweb.u-strasb.fr/meeting2/strsgb02.pdf> (pages consultées le 16 août 2003)
- [Res. 3] DUBOST, K. « RDF et les métadonnées » W3C [En ligne] <http://www.w3.org/2000/Talks/1019-rdf-poly/Overview.html> (pages consultées le 16 juillet 2003)
- [Res. 4] DUCLOY, J. « Architecture générale d'un serveur » *INIST* <http://dilib.inist.fr/dilib/v04Public/Documentation/Tutorial/2/1/archi.fre.html> (pages consultées le 20 juillet 2003)
- [Res. 5] DUCLOY, J. « Introduction générale et découverte de Unix, Xml, WWW et Dilib » *INIST* [En ligne] <http://dilib.inist.fr/dilib/v04Public/Documentation/Tutorial/1/1/index.fre.html> (pages consultées le 16 juillet 2003)
- [Res. 6] Grand dictionnaire terminologique. <http://granddictionnaire.com> [En ligne] (pages consultées le 16 juillet 2003)
- [Res. 7] Dictionnaire de l'Informatique et de l'Internet. Définition de serveur [En ligne] <http://www.dicofr.com/cgi-bin/n.pl/dicofr/definition/20010101004517> (pages consultées le 16 juillet 2003)
- [Res. 8] INIST – CNRS. « INIST-CNRS : Aide serveur d'investigation » [En ligne] <http://www.inist.fr/DEMOS/veilleset.htm> (pages consultées le 27 juillet 2003)
- [Res. 9] INIST – CNRS. « Miriad URI » <http://www.inist.fr/uri/miriad.htm> [En ligne] (pages consultées le 28 juillet 2003)

TABLE DES ANNEXES :

ANNEXE 1 : Bibliographie sélective sur le classement et l'indexation des images

ANNEXE 2 : Etapes de création du Serveur DILIB

ANNEXE 3 : Echange de courrier avec le partenaire PRODIG

ANNEXE 4 : Page d'accueil « BIBALEau »

ANNEXE 5 : Problèmes rencontrés sur les mots-clés les plus fréquents

ANNEXE 1 : Bibliographie sélective sur le classement et l'indexation des images

Mots-clés utilisés sur le SUDOC (Système Universitaire de Documentation) :

classement - indexation – image – photographie

3 monographies sélectionnées :

1 – ADBS. Le thésaurus de l'image : étude des langages documentaires pour l'audiovisuel. Paris : ADBS, 1994, 94 p.

2 – CHARBONNEAU, N., ROBERT, M. La gestion des archives photographiques. Sainte-Foy (Québec) : Presse de l'Université du Québec, 2001, 326 p.

Résumé :

Les auteurs traitent de manière complète la gestion des archives photographiques sous des angles à la fois théoriques et pratiques. Ils présentent l'histoire et les techniques de la photographie, l'acquisition, l'estimation de la valeur financière, la classification, le tri, la description, l'indexation, la préservation, la diffusion, ainsi que des études d'utilisateurs.

3 – KATTING, C. Gestion et diffusion d'un fonds d'images. Paris : Nathan université - ADBS, 2002, 128 p.

Résumé :

Cet ouvrage méthodologique décrit les différentes étapes de la mise en place d'une banque d'images (images fixes reproduites sur support numérique ou électronique), en s'attachant aux aspects techniques, juridiques et documentaires. L'auteur présente des collections et banques d'images, le traitement matériel et intellectuel de l'image, son exploitation et sa diffusion.

Mots-clés utilisés sur ARTILCE@INIST :

((images) <AND> (indexation) <AND> (français))

● Résultat de la recherche avec K2 : Base monographie

2 thèses sélectionnées :

1 – BOUET, M., BRIAND, H. Traitement de l'information multimédia : modélisation, indexation, traitement de la forme et recherche d'images dans un SGBD à objets. Nantes Université : Thèse doct, 2000, 218 p.

2 – PICARD, S., MOHR, R. Etude de descripteurs non paramétriques pour l'indexation d'images par le contenu. Application au cas des images aériennes. Grenoble : Thèse INPG doct., 1999, 125 p.

● Résultat de la recherche avec K2 : Base article

6 articles de périodiques sélectionnés - Années 2002 à 1999

1 – CORNUEJOLS, M. L'organisation des réseaux sémantiques verbaux et imagés: De l'analyse psycholinguistique au traitement automatique du langage. *Bulletin de linguistique appliquée et générale*. 2002, no 27, p. 39 – 56

Résumé :

Les études psycholinguistiques sur l'organisation des réseaux sémantiques en mémoire humaine peuvent servir de base pour élaborer des représentations sémantiques dans les systèmes de traitement automatique du langage (indexation, traduction, constitution de dictionnaires,...) et les systèmes de recherche d'informations (Web sémantique, e-learning,...). L'étude psycholinguistique développée dans cet article montre l'existence de réseaux sémantiques différenciés pour l'image et pour le mot. La connaissance des réseaux d'associations, iconiques d'une part et verbaux d'autre part, leur mode d'organisation et leur interconnexion est proposée comme base de réflexion pour la constitution de représentations sémantiques de contenus plus adaptées à l'utilisateur final. La représentation de contenu informationnel sous forme de graphes conceptuels et la prise en compte des associations sémantiques verbales et imagées sont envisagées pour constituer des systèmes de traitement ou de recherche d'information.

2 – DOMBRE, J., RICHARD, N., FERNANDEZ-MALOIGNE, C. Nouvelle utilisation de l'information spatiale pour l'indexation d'images basée contenu. *Annales des télécommunications*. 2002 vol. 57, no 9-10, p. 943 – 957

Résumé :

En indexation d'images, la plupart des descripteurs développés sont basés sur des statistiques globales des images. Aujourd'hui, des méthodes essayant d'aller plus loin dans la représentation apparaissent. Sans aller jusqu'à une description sémantique, elles représentent les images plus finement en utilisant des statistiques localisées ainsi que les interactions entre celles-ci. Dans la plupart des cas, des structures de graphes sont utilisées. Cet article présente ces diverses approches ainsi que celle que nous avons développée en construisant un graphe pyramidal. Les premiers résultats obtenus pour des requêtes globales et partielles avec cette structure sont aussi présentés.

3 – HUDON, M., TURNER, DEVIN, Y. Description et indexation des collections d'images en mouvement : résultats d'une enquête. *Documentation et bibliothèques*. 2001, vol. 47, no 1, p. 5 – 12

Résumé :

Dans le cadre d'un projet de recherche récemment complété, nous nous sommes intéressés au lexique et à la structure des langages documentaires utilisés pour représenter le contenu d'images en mouvement décrivant des objets et des situations de la vie courante. Onze organismes, gérant quatorze collections d'images ont répondu à un questionnaire et ont été visités. Les données recueillies ont permis de constater que les collections sont imposantes et qu'au moins la moitié d'entre elles sont indexées au niveau du plan à l'aide d'outils langagiers plus ou moins contrôlés, souvent un thésaurus. Mais malgré la similarité des collections, les divers lexiques montrent peu de recouvrements. Les collections d'images en mouvement sont encore décrites et indexées selon des principes et des techniques établis localement, peu normalisés et rarement compatibles.

4 – MOULIS, A.M. L'analyse documentaire des images animées. *Documentaliste*. 1999, vol. 36, no 3, p. 171 – 178

Résumé :

Le traitement des documents audiovisuels impose des méthodes adaptées à leurs spécificités et exige des compétences particulières : l'analyse de ces documents ne peut être menée de la même manière

que celle des documents textuels, ni avec les mêmes outils. Cet article fait le point sur l'analyse documentaire et l'indexation des images animées, leurs particularités, leurs difficultés et leur richesse. Il évoque in fine les perspectives ouvertes par l'indexation automatique de ces images et par le projet de norme MPEG7 qui établit un ensemble standard de descripteurs multimédias pour favoriser l'accès direct aux images animées.

5 – TUNER, J.M., ROULIER, J.F. La description d'images fixes et en mouvement par deux groupes linguistiques, anglophone et francophone, au Québec. *Documentation et bibliothèques*. 1999, vol. 45, no 1, p. 17 – 22

Résumé :

Le présent article rapporte les résultats d'un projet de recherche entrepris dans le cadre d'une série d'études portant sur l'indexation d'images fixes ou en mouvement. L'objectif était de déterminer le niveau d'équivalence dans le choix d'expressions verbales pour représenter le contenu d'une image entre les termes d'indexation choisis par des francophones et ceux choisis par des anglophones. En comparant les résultats obtenus, nous constaterons un réel degré de similitude dans la façon de décrire les images chez les deux groupes linguistiques. De plus, ces résultats viennent appuyer l'hypothèse voulant que l'indexation d'images en français pourrait être obtenue à partir d'une indexation existante en anglais, et vice versa.

6 – WAUTELEY, M., LESPINASSE, K. James Turner à la recherche du stock-shot perdu. *Archimag*. 2001, no 149, p. 33 – 34

Résumé :

Tout documentaliste qui s'y est frotté sait que l'indexation des images animées est une autre paire de manches que celle des documents textuels. Différents chercheurs se penchent sur cette question ; parmi eux le Canadien James Turner dont voici un panorama des travaux.

Mots-clés utilisés sur MIRIAD :

(mc="image" ou mc="photographie*") et (mc="indexation*")*

3 résultats sélectionnés - Années 1999 à 2002

1– CORNUEJOLS, M., LIMAME D., SKOURATOV, I., THOMAS, I. L'organisation des réseaux sémantiques verbaux et imagés : de l'analyse psycholinguistique au traitement automatique du langage. *Bulletin de linguistique appliquée et générale*. 2002; NO. 27 ; 39-56 p.

[Voir résumé dans références articles de périodiques ci-dessus n° 1]

2 – SALWAY, A., FREHEN, C., MELBY, A.K. Terminology and knowledge engineering/6/2002-08-28/Nancy FRA ; Allemagne A. Words for pictures: analysing a corpus of art texts Le Chesnay : INRIA; DA. 2002, p. 113-118

Résumé :

The use of collateral text for indexing images is a promising approach for closing the 'semantic gap' between image databases and their users, as demonstrated by systems which process HTML tags and newspaper photograph captions to extract metadata for images. This approach can be extended in

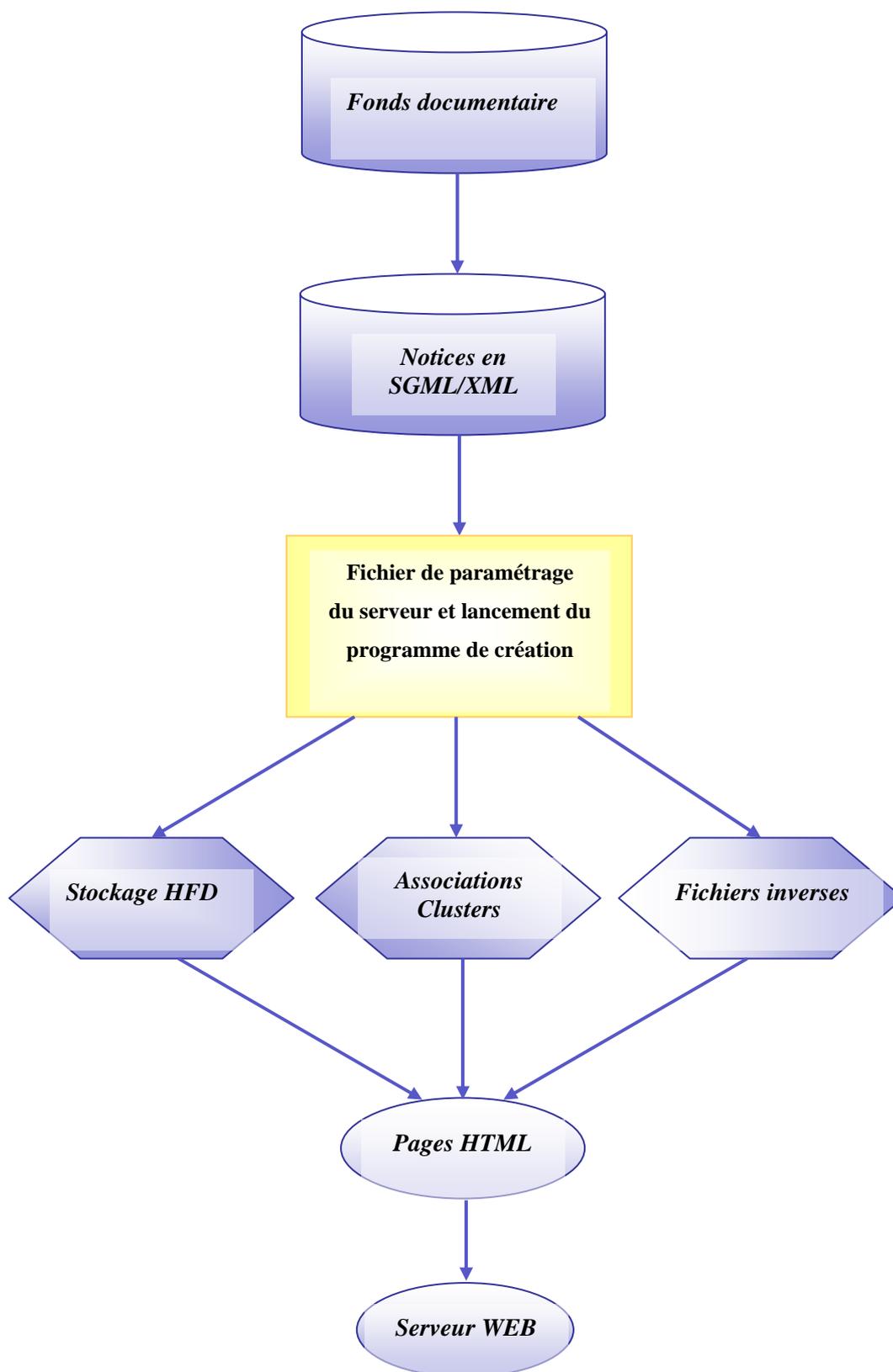
specialist domains like fine art in which the 'semantics' of paintings are articulated in the extensive discourses of subject experts. Thus we may talk of a corpus of collateral texts, gathered from the Web, to provide a conceptual framework for the organization of visual information through the extraction of terminology and knowledge-bases. The premise is that experts systematically articulate the subject matter of paintings and other domain knowledge using a special language. This paper analyses a corpus of art texts, specifically painting captions and artist biographies, to examine evidence for a special language of art and to evaluate the use of such a corpus for organising and accessing visual information in an on-line art gallery.

3 – JÖRGENSEN, C., GOODRUM, A., O'CONNOR, B., TURNER, J. Access to pictorial material: A review of current research and future prospects. *Computers and the humanities* ; DA. 1999; VOL. 33; NO. 4 ; p. 293-318

Résumé :

Rapid expansion in the digitization of image and image collections has vastly increased the numbers of images available to scholars and researchers through electronic means. This research review will familiarize the reader with current research applicable to the development of image retrieval systems and provides additional material for exploring the topic further, both in print and online. The discussion will cover several broad areas, among them classification and indexing systems used for describing image collections and research initiatives into image access focusing on image attributes, users, queries, tasks, and cognitive aspects of searching. Prospects for the future of image access, including an outline of future research initiatives, are discussed. Further research in each of these areas will provide basic data which will inform and enrich image access system design and will hopefully provide a richer, more flexible, and satisfactory environment for searching for and discovering images. Harnessing the true power of the digital image environment will only be possible when image retrieval systems are coherently designed from principles derived from the fullest range of applicable disciplines, rather than from isolated or fragmented perspectives.

ANNEXE 2 : Etapes de création du serveur DILIB



ANNEXE 3 : Echange de courrier avec le partenaire PRODIG

Date: Tue, 8 Jul 2003 10:57:08 +0200
De: [Claire List <list@univ-paris1.fr>](mailto: Claire List <list@univ-paris1.fr>)
A: [brigitte.parant@etudiant.univ-nancy2.fr](mailto: brigitte.parant@etudiant.univ-nancy2.fr)
Sujet: Tr: Serveurs eau et géographie

 2 [unnamed](#) text/html 2.58 KB 

----- Original Message -----

From: [Claire List](mailto: Claire List)
To: [Gaelle Hallair](mailto: Gaelle Hallair) ; [Rageot@inist.fr](mailto: Rageot@inist.fr)
Sent: Tuesday, July 08, 2003 10:54 AM
Subject: Serveurs eau et géographie

Bonjour,

Nous avons exploré avec beaucoup d'intérêt les 3 serveurs et sommes en admiration devant le travail réalisé.

Il nous semble que ce sera plus intéressant de coupler le serveur images au serveur bibliographique le plus large (Allemagne et pays limitrophes) afin de générer plus de liens entre les 2, notamment avec les photos concernant Prague.

Dans une prochaine étape, ce serait aussi intéressant d'enrichir le serveur images avec des possibilités d'interrogation à partir des descripteurs et les mots du titre en anglais et en allemand. Gaelle Hallair continue à travailler sur l'indexation et la traduction des titres des derniers lots d'images et vous les fera parvenir dès que possible.

Très cordiales salutations

Claire LIST
Responsable BGI

UMR 8586 PRODIG
191 rue Saint Jacques
75005 Paris
Tel : 01 42 34 56 30
Fax : 01 43 29 63 83
e-mail : [list@univ-paris1.fr](mailto: list@univ-paris1.fr)

ANNEXE 4 : Page d'Accueil « BIBALEau »

A l'occasion du FIG-2003



14^{ème} Festival International de Géographie de Saint-Dié-des-Vosges (2-5 10 2003)

sur le thème de :

l'Eau et la Géographie en Allemagne

Source de vie, source de conflits, trait d'union entre les hommes



vous proposent

BIBALEau

- Une **B**ase **I**conographique et **B**ibliographique sur l'**AL**lemagne et l'**Eau** -

Cette base est un prototype. Attention cette version est optimisée par Netscape 7.0 et Internet Explorer

ANNEXE 5 : Problèmes rencontrés sur les mots-clés les plus fréquents

| Descripteurs Français | nb docs |
|---------------------------------------|---------|
| Eau | 88 |
| Allemagne | 79 |
| Cours d'eau | 45 |
| Paysage urbain | 34 |
| Elbe | 22 |
| Hambourg | 20 |
| Berlin | 18 |
| Spree | 16 |
| Port | 16 |
| Architecture | 13 |
| Ville | 11 |
| Espace bâti | 11 |
| Canal | 11 |
| Trafic | 10 |
| Quartier | 10 |
| Pont | 10 |
| Espace urbain | 10 |
| Rénovation urbaine | 9 |
| Photographie aérienne | 9 |
| Meissen | 9 |

| Descripteurs Français | nb docs |
|--|---------|
| Infrastructure portuaire | 9 |
| Centre-ville | 8 |
| Tourisme | 7 |
| Sachsen | 7 |
| Sachsen (Saxe) | 7 |
| Praha (Prague) | 7 |
| Paysage rural | 7 |
| Loisir | 7 |
| Vltava | 6 |
| Schleswig-Holstein | 6 |
| Lübeck | 6 |
| Baltique mer | 6 |
| Végétation | 5 |
| Trave | 5 |
| Habitat urbain | 5 |
| Aménagement urbain | 5 |
| Transport fluvial | 4 |
| Tchèque (République) | 4 |
| Rhin | 4 |
| Port maritime | 4 |
| Mer du Nord | 4 |

| Descripteurs Français | nb docs |
|--------------------------------------|---------|
| Mer du Nord | 4 |
| Marchandise | 4 |
| Hiver | 4 |
| Flux touristique | 4 |
| Dresden (Dresde) | 4 |
| Crue | 4 |
| Bremen (Brême) | 4 |
| Rivage | 3 |
| Port commercial | 3 |
| Parc urbain | 3 |
| Lit majeur | 3 |
| La Hanse | 3 |
| Ile | 3 |
| Dégradation | 3 |
| Conteneur | 3 |
| Tchèque République | 2 |
| Tchèque (république) | 2 |
| Prague | 2 |
| Port fluvial | 2 |
| Littoral | 2 |
| Lac | 2 |

| Descripteurs Français | nb docs |
|---|---------|
| Lac | 2 |
| Histoire urbaine | 2 |
| Espace public | 2 |
| Economie | 2 |
| Construction urbaine | 2 |
| Complexe industrialo-portuaire | 2 |
| usine | 1 |
| cours d'eau | 1 |
| Würtzburg | 1 |
| Vignoble | 1 |
| Vigne | 1 |
| Transport | 1 |
| Transport ferroviaire | 1 |
| Trafic maritime | 1 |
| Tourbe | 1 |
| Schleswih-Holstein | 1 |
| Sédimentologie | 1 |
| Rheingrafenfels | 1 |
| Rhénanie-Palatina (Rheinland-Pfalz) | 1 |
| République (Tchèque) | 1 |
| Réhabilitation | 1 |

| Descripteurs Français | nb docs |
|---|---------|
| Réhabilitation | 1 |
| Port de plaisance | 1 |
| Port de commerce | 1 |
| Port Fluvial | 1 |
| Plage | 1 |
| Pêche | 1 |
| Nordrhein-Westfalen (Rhénanie du Nord-Westphalie) | 1 |
| Moselle | 1 |
| Montagne | 1 |
| Mer | 1 |
| Main | 1 |
| Logement | 1 |
| Kiel | 1 |
| Inondation | 1 |
| Industrie | 1 |
| Hydrocarbure | 1 |
| Humidité | 1 |
| Hamburg (Hambourg) | 1 |
| Fluviatile | 1 |
| Finance | 1 |
| Falaise | 1 |

| Descripteurs Français | nb docs |
|--------------------------------------|---------|
| Falaise | 1 |
| Espace vert | 1 |
| Espace rural | 1 |
| Equipement portuaire | 1 |
| Embouchure | 1 |
| Duisbourg (Duisburg) | 1 |
| Dresden | 1 |
| Culturel | 1 |
| Cana | 1 |
| Bayern (Bavière) | 1 |
| Baltique (mer) | 1 |
| Aménagement fluvial | 1 |
| Agriculture | 1 |
| Administration | 1 |

Résumé :

Le serveur d'investigation iconographique et bibliographique réalisé pour le Festival International de Géographie 2003 de Saint-Dié-des-Vosges est le résultat d'un projet commun entre l'INIST (INstitut d'Information Scientifique et Technique) et l'UMR (Unité Mixte de Recherche) 8586 PRODIG (Pôle de Recherche pour l'Organisation et la Diffusion de l'Information Géographique).

Ce prototype intégrant des photographies et des notices bibliographiques sur le thème de l'eau en Allemagne et en République Tchèque a permis de mettre en œuvre une méthodologie d'exploitation de ressources iconographiques grâce à la plate-forme DILIB.

Les différentes étapes de cette étude spécialisée, détaillées dans ce rapport, devraient faciliter la poursuite de cette expérimentation et pourrait aboutir à l'offre d'un nouveau produit INIST.

Mots-clés :

Allemagne, BIBALEau, Bibliométrie, Festival International de Géographie, FIG 2003, Eau, INIST, Plate-forme DILIB, PRODIG, Saint-Dié-des-Vosges, République Tchèque, Serveur d'investigation, Serveur d'investigation iconographique et bibliographique.