Alexandra SALAVERT

RAPPORT DE STAGE PROJET REVELEC : REALISATION D'UNE PLATE FORME D'ACCES A DES REVUES ELECTRONIQUES

Mai 2002 à août 2002
Stage effectué à l'INIST-CNRS

Sous la direction de Sylvie Grésillaud

DESS IST-IE promotion 2001-2002

Université Nancy 2- Université Henri Poincaré- INPL

AVANT-PROPOS

Mon stage s'est déroulé au DPS à l'INIST. Je tiens à remercier tous les membres du service pour leur accueil, leur gentillesse et leur soutien mais aussi :

?? Sylvie GRESILLAUD pour son suivi, la confiance qu'elle m'a accordée dans notre travail d'équipe, ses compétences et ses précieux conseils,

?? Thierry CHANIER pour son expérience concernant l'édition, ses remarques lors de nos réunions de travail et son investissement dans le projet,

?? Philippe HOUDRY pour sa disponibilité, ses nombreux conseils et sa relecture,

?? Jacques DUCLOY pour m'avoir permis de réaliser ce stage et m'avoir fait confiance,

?? Sabine BARREAUX pour sa participation dans les questions de vocabulaire en linguistique

*Les * dans le texte signalent un renvoi vers le glossaire en fin de document*

SOMMAIRE

AVANT-PROPOS **3**

INTRODUCTION **6**

PREAMBULE : L'ORGANISME **7**

A. PRESENTATION DE L'INIST 7

B. LE DEPARTEMENT PRODUITS ET SERVICES (DPS) 9

I. PRESENTATION DU PROJET **10**

A. PROJET MCER 10

1. Les Acteurs : ALSIC et l'INIST 10
2. Objectifs 11

B. REVELEC : CAHIER DES CHARGES 12

1. Objectifs généraux 13
2. Les différentes phases du projet REVELEC 14

a. Module de préambule 14

b. Module 1 :" Normalisation des ressources primaires électroniques 14

c. Module 2 " Intégration d'outils de navigation " 15

d. Module 3 " Mise à disposition des ressources sur un service Web " 15

e. Module 4 " Archivage " 15

f. Module 5 " Extension / industrialisation " 15

3. Le planning 16

II. LE SERVEUR REVELEC **17**

A. LA PLATE FORME DILIB 17

1. Présentation 17
2. Fonctionnalités 17

B. MODELE DE NAVIGATION ET STRUCTURE DU SERVEUR 18

1. Identification des ressources 18
2. Modèle de navigation 20

C. DECHARGEMENT ET CONCEPTION DES NOTICES 22

1. Les types de notices 22
2. Normes Dublin Core et RDF 23
3. Création de la structure des notices 24
D. PARAMETRAGE ET CREATION DU SERVEUR 25
1. Création du serveur 25
2. Programmation des cgi-bin 28
E. PLATE FORME WEB 29
III. TRAITEMENT DE LA REVUE EN XML 31
A. LE DOCUMENT STRUCTURE EN XML 32
1. Standard XML et DTDs 32
2. Choix de la DTD 33
3. Conception de la structure et du balisage 35
B. CONVERSION DU HTML EN XML 36
1. Conception des programmes LEX permettant la conversion des articles en XML . 36
2. Parsage, conception d'une feuille de style 37
C. TRAITEMENT DES ARTICLES DANS LE SERVEUR : NAVIGATION ENTRE LES REFERENCES ... 39
1. Identification interne des articles et structuration arborescente 39
2. Extraction des références et des citations 41
3. Schéma de navigation 43
D. REFLEXION SUR UNE CHAINE DE TRAITEMENT 44
CONCLUSION 48
REFERENCES 49
GLOSSAIRE 52
TABLE DES ILLUSTRATIONS 56
ANNEXES 57

INTRODUCTION

L'INstitut de l'Information Scientifique et Technique (INIST*-CNRS) collecte, traite et diffuse l'information scientifique issue de la recherche scientifique, sous forme de copies de documents scientifiques (activité de Fourniture de Documents) ou de notices bibliographiques correspondant à ces documents (activité de production de bases de données). Cependant le monde de l'édition change depuis quelques années : les éditeurs et les agences d'abonnement proposent, en parallèle ou en remplacement du papier, des offres de service permettant l'accès au document original sous forme électronique. C'est ainsi que l'on trouve, aux côtés des abonnements classiques aux revues papier, des services d'abonnement à des portails et kiosques où, en plus des catalogues ou des revues de sommaires, il est possible d'accéder directement au texte intégral des articles eux-mêmes.

L'INIST*-CNRS se doit de prendre part à ces nouveaux enjeux en tenant compte de l'évolution des technologies de l'édition, en adaptant son système d'information et en déployant au sein de son personnel de nouveaux métiers. Tout en continuant son activité de diffuseur d'information il se tourne alors actuellement vers un type de services qui permettra de proposer à ses utilisateurs (en particulier à la communauté des chercheurs) de nouvelles interfaces d'accès aux ressources documentaires numériques et la navigation dans ces ressources. Ces produits seront accompagnés de formation pour la diffusion des compétences.

Issu d'un partenariat entre la revue ALSIC* (www.alsic.org/) et l'INIST, le projet REVELEC* se positionne alors dans cette nouvelle problématique et définira en quelque sorte la place et le rôle de l'INIST par rapport à l'édition électronique. Mon stage de 4 mois à l'INIST a porté sur la conception d'un prototype de plate forme de revues électroniques. Il a en fait servi à faire une étude de faisabilité évaluant les besoins et les moyens à mettre en œuvre pour ce type de services. Ce projet pourra alors permettre à l'INIST de devenir un **"opérateur technique"** dans le cadre de conception de revues électroniques, d'**intégrer l'accès aux documents électroniques** dans le portail **CONNECTSCIENCES et de proposer des outils de navigation** dans ces ressources électroniques.

Le projet REVELEC* s'inscrit au même titre que les mutations technologiques dans la tendance actuelle de l'INIST à évoluer afin d'intégrer les nouvelles technologies dans l'édition et les savoirs-faire associés (évolution des métiers).

PREAMBULE : L'ORGANISME

A. Présentation de L'INIST

L'INstitut de l'Information Scientifique et Technique (INIST) est une unité de service du Centre National de la Recherche Scientifique (CNRS) chargée de collecter, analyser et diffuser les résultats de la recherche. Il est le premier centre européen à mettre à la disposition des centres de recherche des ressources et services permettant un meilleur accès à l'Information Scientifique et Technique (IST*). Fournisseur de copies de documents et producteur de bases de données multidisciplinaires, l'INIST a suivi le développement des nouvelles technologies qui a ouvert de nouvelles perspectives de services sur Internet.

Ses missions sont principalement de :

- **Diffuser l'information** scientifique produite par les différents organismes de recherche publics dans tous les domaines : sciences, technologie, médecine, sciences humaines, sociales et économiques, et en améliorer la collecte et l'analyse. Pour cela, l'INIST dispose d'un fonds documentaire comprenant 26 000 titres de revues (dont 8 500 collections en cours), 60 000 rapports scientifiques, 62 000 comptes rendus de congrès français et internationaux, 110 000 thèses françaises de Sciences et Techniques soutenues dans les universités françaises depuis 1985 et 10 000 ouvrages.
- **Favoriser le transfert de l'information vers le milieu socioéconomique**. De nombreux laboratoires ou entreprises privés ont recours aux services de l'INIST pour la fourniture de documents (700 000 par an au total), la consultation des bases de données et les services de recherche sur Internet.
- **Développer des outils et des services de veille scientifique et technologique**. Qu'il s'agisse d'outils pour le traitement bibliométrique* ou pour l'analyse infométrique des données, de services de veille automatisée ou de prestations particulières sur l'évolution technologique ou scientifique dans un domaine, l'INIST offre en effet une aide capitale pour l'élaboration de stratégies de recherche et développement.

- **Développer l'accès à l'information électronique** qui permet la localisation et l'identification des documents. L'INIST multiplie les accès à l'IST* sur Internet (ConnectSciences*, Articlesciences*, Article@INIST*…) et s'investit dans le développement du document électronique. En prenant en compte l'évolution des technologies de l'édition, l'INIST se doit aujourd'hui de créer de nouveaux produits et services dans le monde de l'édition électronique. C'est alors sur dans type de perspective que s'inscrit le projet REVELEC sur lequel a porté mon stage.

Par rapport à ces grands axes, les différents métiers à l'INIST se répartissent entre plusieurs départements. Mon stage s'est déroulé dans le Département Produits et Services.

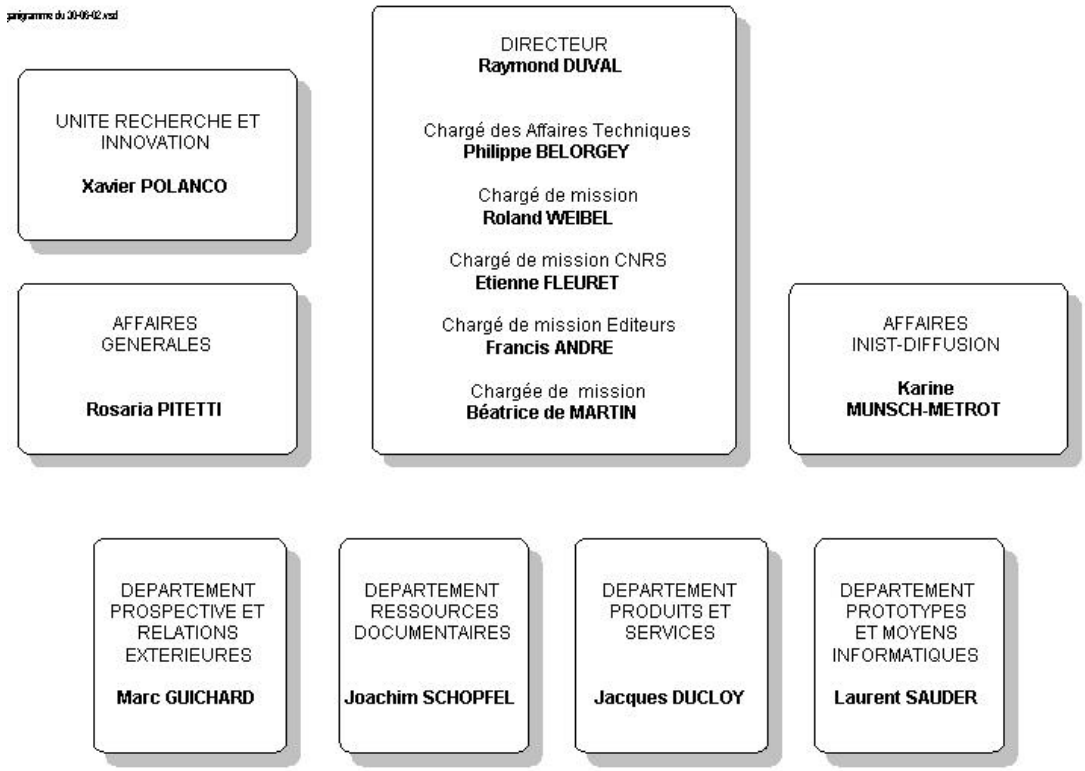


Figure 1 : Organigramme de l'INIST

B. Le Département Produits et Services (DPS)

Le Département Produits et Services, dirigé par J. Ducloy, assure la constitution des bases bibliographiques de l'INIST, la fabrication de produits documentaires en ligne, la FDP (la fourniture de copies de documents) et la mise en place de services de veille et leur exécutions. Il comprend différents services dont le Service Ingénierie et Partenariats (SIP) dans lequel s'est déroulé mon stage. Ce service, sous la responsabilité de P. Pelisson, regroupe les chefs de projet « utilisateurs » et remplit plusieurs missions:

- La fabrication des produits comme les mises à disposition des bases Pascal* (en sciences et techniques) et FRANCIS* (sciences humaines) sur serveurs et sur Cédéroms ou les différents services disponibles sur le site de l'INIST (services@inist*)
- La gestion des projets liés à l'édition dont l'édition du document électronique
- La mise en place et le suivi de partenariats et coopérations
- La mutation technologique (série de formations) qui permet de doter les ingénieurs et techniciens d'un niveau de compétence pointu dans les NTIC

Mon projet de stage rejoint ainsi les problématiques de travail sur le document électronique, projet dont est chargée Sylvie Grésillaud au SIP.

I. PRESENTATION DU PROJET

Le projet REVELEC sur lequel a porté mon stage entre dans le cadre du vaste programme **de Mutualisation des Compétences et des ressources pour l'Édition et la diffusion de Revues numériques** (projet **MCER)** qui a le soutien de la Direction de la Technologie du ministère de la Recherche. Ce projet est commun à la revue ALSIC* et à l'INIST-CNRS.

A. Projet MCER

1. Les Acteurs : ALSIC et l'INIST

La revue électronique ALSIC (Apprentissage des Langues et Système d'Information et de Communication), dirigée par T. Chanier (responsable éditorial), regroupe une communauté de chercheurs et de praticiens (enseignants et formateurs) à la fois auteurs et lecteurs (2 000 lecteurs internationaux) sur le thème de l'apprentissage des langues. A la frontière entre la linguistique et les sciences de l'éducation, la revue, accessible librement en texte intégral, a pour mission de concevoir de nouveaux environnements informatiques pour aider à l'apprentissage des langues, de permettre de mieux comprendre ce processus d'apprentissage et de comprendre les conditions favorables à l'utilisation des TICE* (Technologie de l'Information et de la Communication dans l'Enseignement). Elle propose la publication d'articles de recherche et de comptes rendus de livres, de logiciels, de sites, un glossaire bilingue (français – anglais), des résumés de thèses, une médiathèque (liens vers des sites portails, des activités pédagogiques, des textes de référence, des outils de référence, des distributeurs de matériel) et un annuaire d'associations.

Après quatre années d'existence (depuis juin 1998, parution trimestrielle), la revue ALSIC est confrontée, dans le cadre de son effort constant de développement, à de nouveaux défis, dont le passage au format XML*, la constitution d'une chaîne de traitement spécifique, la normalisation de sa description pour le référencement et le catalogage, l'archivage, l'indexation* et, ce qui est lié, l'ouverture de nouveaux services en ligne à son lectorat.

De son côté, l'INIST a traditionnellement eu un rôle plus en aval dans la chaîne de traitement de l'information scientifique et technique. Aujourd'hui, les nouveaux supports et modes de diffusion de l'information (documents multimédias, diffusion en ligne, etc.) lui offrent la possibilité d'ouvrir de nouveaux services à la communauté des chercheurs français, en intervenant plus en amont dans le processus d'élaboration de l'information.

Des organisations comme ALSIC et l'INIST ont donc des intérêts communs : l'une est productrice des données et informations de base, la seconde peut ajouter sur ce substrat de nouvelles informations et de nouveaux services. L'INIST et ALSIC ont donc décidé début 2002 d'établir un partenariat dans le cadre du projet MCER. Il a donc tout d'abord fallu définir, à partir des intérêts communs, des objectifs et une ligne directrice pour la conduite de ce projet.

2. Objectifs

Les premières discussions entre l'INIST et ALSIC ont, en premier lieu, permis de faire émerger un désir commun de partager données et informations, afin que chacun puisse développer ses nouvelles missions, en particulier :

?? **Proposer un accès au texte intégral des revues** librement accessible à des fins d'indexation*, de recherches croisées entre contenus de revues différentes pour les lecteurs, pour l'archivage et le catalogage.

?? **Intégrer des documents multimédias au texte** afin de permettre de nouvelles interactions en ligne, qui viendront profondément modifier l'acte de lecture et écarteront, dans ce cas, le papier comme support alternatif de lecture[1].

?? **Associer des métadonnées* aux documents primaires**, décrites dans le respect des normes internationales (Dublin Core) et permettant ainsi le référencement, le stockage et l'échange de ces documents.

?? **Uniformiser les citations** de documents multimédias pour les références bibliographiques, documents multimédias, sites Internet, logiciels (systématisation de l'effort entrepris par ALSIC).

[1] Par exemple des ressources audio pour alimenter la discussion scientifique sur l'aide à l'apprentissage de l'oral en langue.

?? **Mettre au point une nouvelle chaîne d'édition**, de diffusion, d'indexation / description normée et d'archivage pour revue électronique.

?? **Appliquer cette chaîne au cas particulier de la revue ALSIC** qui, tout en continuant à paraître sur son site propre, serait également diffusée sur le site de l'INIST avec intégration à un serveur de documents électroniques doté de nouveaux services de navigation pour la communauté scientifique.

?? **Étendre ce modèle** (avec les adaptations nécessaires) **à d'autres revues** françaises, revues existant aujourd'hui sur support papier et désireuses de passer en ligne ou revues ayant déjà accompli ce pas.

Le projet comporte alors deux phases :

?? Une phase pilote qui consiste à concevoir un modèle d'édition (métadonnées* associées aux documents primaires décrites dans le respect des normes internationales telles le DublinCore, uniformisation des citations de documents multimédias, référencement des documents…) et à l'appliquer à la revue ALSIC.

?? Une phase de développement d'un système de navigation et d'indexation sur le serveur de documents électroniques et un début d'application de notre modèle à d'autres revues de même thématique (comme la revue américaine LLTJ, *Language and Learning Technology Journal*)[2].

B. REVELEC : cahier des charges

Le projet MCER comporte alors plusieurs phases distinctes qui sont réparties entre les deux organismes. L'INIST-CNRS se propose pour sa part de développer le **prototype d'une plate forme de mise en ligne de revues électroniques avec intégration de ressources documentaires et utilisation d'outils de navigation infométriques appelée REVELEC.** Mon stage a abouti à la réalisation de ce prototype qui s'appuie sur l'exploitation concrète de la revue ALSIC traitée dans le projet MCER. Ce prototype qui consiste à faire une étude de faisabilité pour l'INIST conduira en fait à la rédaction d'un cahier des charges par le chef de

[2] http://llt.msu.edu/

projet chargé du document électronique en vue de **l'industrialisation d'une nouvelle gamme de produits et services**.

1. Objectifs généraux

Le point de départ du projet REVELEC a alors consisté à en définir les différentes phases et à en détailler les composants. Compte tenu des besoins de la revue ALSIC et des possibilités qu'offre l'INIST, il a fallu répartir et définir le travail à effectuer. Les produits pour les revues électroniques que devra proposer l'INIST tournent autour de deux points : une prestation de service pour transformer les ressources (revues en l'occurrence) et la création de produits de diffusion des ressources électroniques. Ces réalisations sont tournées principalement vers le domaine des Sciences Humaines et Sociales (SHS) qui est un des domaine dans lequel les enjeux de ce type de produit sont importants[3]. Une grande partie du stage a été consacrée à la conception d'un serveur spécifique aux revues électroniques, telles qu'ALSIC, qui a du prendre en compte ces différents points :

- **L'intégration de revues électroniques** du domaine des **Sciences Humaines** : ALSIC associée dans un second temps à la revue américaine de même thématique et problématique documentaire *Language Learning and technology Journal*.
- **L'association entre revues électroniques et références bibliographiques** en rapport avec la thématique de la revue déchargées à partir des bases de l'INIST ou trouvées sur Internet.
- L'**accès aux ressources documentaires par différents types de navigation** : navigation "classique" par feuilletage, indexation, navigation par thématique / rubrique (intégration de REVELEC dans le portail ConnectSciences*…), navigation à l'aide d'outils infométriques se basant sur le texte, les citations ou d'autres ressources documentaires.

Ensuite, une autre partie du projet a consisté à **proposer un modèle de traitement pour le passage de la revue en XML***. Il a alors fallu spécifier les types de formats qui entrent en

[3] La plupart des projets de revues électroniques sont en effet dans le domaine des sciences humaines et sociales. Voir par exemple le projet Erudit (http://www.erudit.org/) ou le site de revues.org

jeu, transformer les ressources existantes et aborder les différentes étapes de conception des articles, de l'auteur à la publication. Cette phase a donc porté sur deux points :

- la **normalisation des ressources documentaires** (texte des documents, métadonnées catalogage et identification des documents et des citations, …)
- la définition d'une chaîne de traitement pour la revue ALSIC.

Le projet REVELEC comporte de ce fait plusieurs modules : le premier définit les formats des ressources documentaires ; le second correspond à l'intégration de la revue sur une plate forme web ; le module 3 est consacré à l'apport de valeur ajoutée dans la diffusion de ces ressources documentaires et enfin, le module 4 est une réflexion sur l'archivage des documents concernés par le projet . Afin de mieux cerner les différents éléments du projet et de les répartir entre ALSIC et l'INIST il a alors fallu décomposer les phases, les définir et établir un planning.

2. Les différentes phases du projet REVELEC

a. Module de préambule

Il est nécessaire, avant de se lancer complètement dans le projet, d'établir un inventaire des techniques et travaux existants : inventaire des chaînes de production et d'analyse de ressources électroniques, des sites de revues électroniques.

b. Module 1 :" Normalisation des ressources primaires électroniques

La revue ALSIC.org est "repensée" en XML et autres formats intégrants les différentes ressources présentes, en particulier le multimédia. Le module 1 consiste à :

✍✍**Référencer les ressources nécessaires**

?? Analyser la structure de la revue et définir les principales sous-structures : texte, références bibliographiques, citations, toilithèque (carnet de sites ou webographie)…

?? Recenser les normes correspondant à ces sous-structures

✍✍**Modéliser les documents en XML**

?? « XMLiser » manuellement ou semi-automatiquement un ou plusieurs numéros de la revue ALSIC

?? Récupérer ou appliquer le même modèle XML (ou un autre) à la *revue Language Learning and Technology Journal*

?? Créer de nouvelles notices manuellement pour compléter l'offre dans ce domaine

c. Module 2 " Intégration d'outils de navigation "

✍✍Récupérer les notices bibliographiques dans FRANCIS*

✍✍Récupérer d'autres ressources bibliographiques à identifier

✍✍Développer un serveur d'investigation DILIB*

?? traitement des bibliographies (lien avec corpus PASCAL*/FRANCIS),

?? organisation du référentiel,

?? choix et implémentation du(des) modèle(s) de navigation.

d. Module 3 " Mise à disposition des ressources sur un service Web "

✍✍Intégrer la revue ALSIC et, dans un second temps, la revue *Language Learning and Technology Journal* dans un service web de l'INIST (ConnectSciences, rubrique "revues électroniques")

✍✍Créer le lien avec le portail éditeur

✍✍Créer le lien avec un portail thématique

✍✍Définir la structure organisationnelle de la publication

e. Module 4 " Archivage "

Stocker les ressources utilisées dans ce prototype. Ce module sera géré par l'éditeur détenteurs des ressources "primaires" (ALSIC) mais l'INIST proposera un modèle d'identification et éventuellement d'archivage sur son propre serveur.

f. Module 5 " Extension / industrialisation "

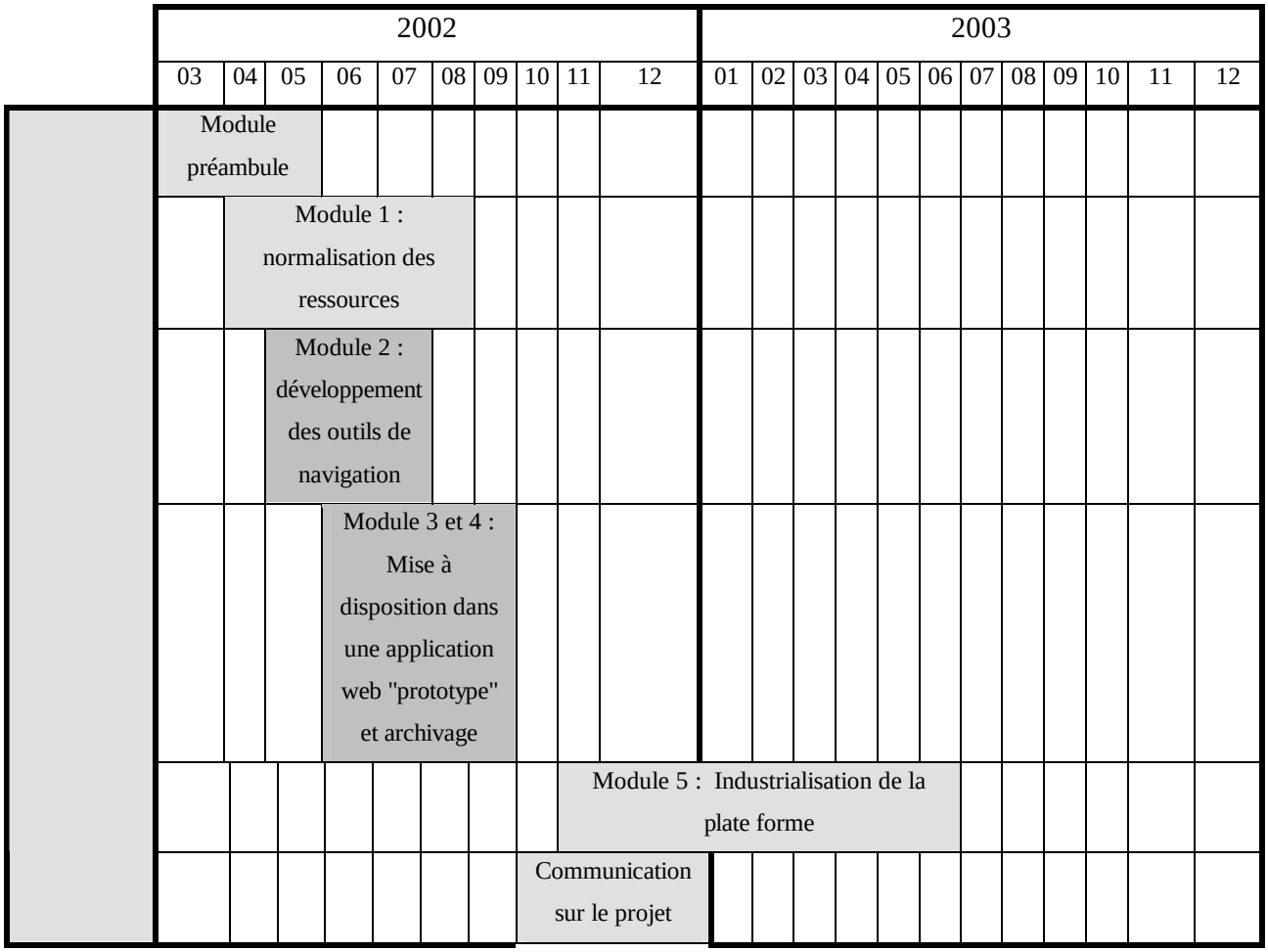
Tout cela aboutira, suite à mon stage à :

✍✍Rédaction du cahier des charges MCER

✍✍Proposition d'une gamme de services et produits sur une plate forme intégrée avec deux fonctions principales : transformer les ressources électroniques, créer un portail de

diffusion des ressources et des services. L'INIST-CNRS deviendra ainsi le prestataire pour l'ensemble des taches de la fabrication à la diffusion des ressources (l'éditeur garde la maîtrise du contenu scientifique de la revue).

3. Le planning



En gris plus foncé apparaissent les modules sur lesquels il était prévu que je travaille en priorité. La demande de stagiaire a en effet été formulée pour le traitement des modules 2 et 3, c'est-à-dire la création du serveur et son intégration dans une application. Les questions concernant l'archivage ne devaient pas faire partie de mon sujet de stage. Cependant, le stockage et l'identification des articles étant nécessaire à la création du prototype, j'ai alors entamé une partie de ce module et proposé des solutions quant à l'archivage. Aussi, il n'était pas exclu, selon le temps disponible, que je m'occupe du module 1, c'est-à-dire de la normalisation des ressources primaires.

II. LE SERVEUR REVELEC

Une grande partie de mon stage a alors consisté à réaliser le serveur d'investigation REVELEC permettant de naviguer à partir des revues électroniques vers d'autres types de ressources. Ce prototype utilise la plate forme DILIB*.

A. La plate forme DILIB

1. Présentation

DILIB est une plate-forme pour l'Ingénierie du Document et l'Information Scientifique et Technique (IST), développée par l'INIST et le LORIA*, permettant de générer un serveur d'investigation avec lequel il est possible de naviguer dans un corpus de notices bibliographiques (voir annexe 1 sur les étapes de création d'un serveur). Cette plate forme permet d'exploiter des documents de formats initialement différents (DILIB peut intégrer des corpus au format XML, SGML* et RDF*) et de générer des applications multi-bases.

DILIB permet non seulement d'explorer l'information technique, documentaire et multimédia, mais aussi de la traiter grâce à un ensemble de composants qui construisent des Systèmes de Recherche d'Information. Elle permet alors l'exploration de corpus documentaires par la navigation hypertexte et leur exploitation par une analyse infométrique. Ses principaux domaines d'application sont l'investigation ou la recherche documentaire, la construction de systèmes de recherche d'information ou la mise en place d'outils pour les bibliothèques électroniques. Il ne s'agit encore que d'un prototype dont certaines fonctionnalités sont en cours d'amélioration.

2. Fonctionnalités

Qualifié de « boîte à outils » SGML*/XML disposant de commandes pour la manipulation de données SGML ou de fonctions de conversions, DILIB permet de construire des serveurs d'investigation. Générés à partir d'un corpus normalisé, il est alors possible de (voir annexe1):

- Naviguer et retrouver des notices à partir de plusieurs index* (descripteurs, mots du titre, auteurs)
- Explorer des sous thématiques (cluster) mises en évidence grâce à des méthodes de calculs infométriques liées à la méthode des mots associés (co-occurrence)
- Trouver pour chaque cluster la liste des termes qui le constituent, la liste des associations à ces termes, des documents pertinents, des auteurs, périodiques ou pays d'affiliation associés et un histogramme d'évolution par année
- Envoyer une requête sur Internet à partir d'une analyse des mots du titre.

La première étape de conception de ce type de serveur a alors consisté à proposer un modèle qui réponde aux objectifs fixés dans le cahier des charges.

B. Modèle de navigation et structure du serveur

Etant donné qu'il n'existait pas de serveur DILIB intégrant des articles de revues électroniques, nous avons en premier lieu réfléchi à sa structure afin de proposer un modèle de navigation adapté. Nous avons pris en considération les besoins de T. Chanier et la valeur ajoutée qu'apporte l'INIST par ce service. La premier étape a été de déterminer les ressources à utiliser pour établir un modèle général de navigation.

1. Identification des ressources

Comme nous l'avons précisé dans le cahier des charges et la définition des besoins, le serveur qu'il faut concevoir doit intégrer les articles de la revue ALSIC mais aussi d'autres ressources qu'il s'agira d'identifier. Le problème a donc été de savoir de quels types de sources et sur quels critères ces ressources seront choisies. Suite à différentes discussions, nous avons décidé de prendre des notices d'articles de la base de données* FRANCIS et de chercher des articles intéressants en ligne.

Sur quels critères seront-ils alors retenus ? Compte tenu des possibilités de navigation dans le serveur, nous avons tout d'abord sélectionné les ressources qui correspondaient au thème d'ALSIC, à savoir l'apprentissage des langues et les NTIC. Nous avons alors déchargé des notices de FRANCIS qui portaient sur ce sujet et quelques articles

en ligne trouvés le plus souvent sur des sites personnels d'universitaires ayant eux même écrit des articles pour ALSIC. Aussi, dans l'idée de développer une navigation entre les articles à partir des citations, nous avons alors identifiées celles-ci dans ALSIC pour rechercher dans FRANCIS ou sur le web la notice ou l'article cité. Pour chaque citation, nous avons alors fait des requêtes sur Miriad* (module intranet de recherche dans les bases de l'INIST), et sur des moteurs de recherche sur Internet[4]. Nous avons ainsi pu rassembler plusieurs types de documents :

- Des notices FRANCIS correspondant thématiquement à ALSIC
- Des notices FRANCIS correspondant aux citations relevées dans les articles d'ALSIC
- Des articles en texte intégral trouvés sur Internet de même thématique qu'ALSIC
- Des articles en texte intégral trouvés sur Internet et correspondant aux citations relevées dans les articles d'ALSIC.

Nous avons donc identifié deux types de ressources : des notices bibliographiques et des articles. La question a alors été de savoir comment les traiter dans le serveur. Il a tout d'abord fallu déterminer le type de serveur, mono ou multi-base et la manière de traiter les ressources sélectionnées. Pour pouvoir présenter un serveur maquette à T. Chanier, nous avons décidé de ne procéder qu'à un signalement des articles (type notice en RDF*) dans le serveur avec un accès possible à l'article en ligne et de laisser pour le moment de côté l'idée d'intégrer l'article en entier, celui-ci devant être pour cela au format XML[5]. Aussi, afin de naviguer entre ces deux types de ressources, les signalements d'articles et les notices FRANCIS, nous avons opté pour un serveur multi-base. Il y aura donc d'un côté des notices FRANCIS déchargées en SGML avec Miriad* et d'un autre des notices signalant les articles de la revue et les articles trouvés sur Internet.

Une fois les ressources identifiées, sélectionnées et déchargées, nous avons alors pu proposer un modèle de navigation entre ces deux bases.

---

[4] Les requêtes ont été faites avec *google* et l'interrogation portait sur l'auteur ou le titre de l'article recherché ou sur le thème de l'apprentissage des langues. Aussi, nous avons exploré les revues en ligne proches d'ALSIC.

[5] En effet, comme nous le verrons par la suite, il n'est pas intéressant pour notre projet d'effectuer un traitement basé sur du texte intégral en HTML* puisque la navigation reposera sur un balisage des citations ce qui n'est pas possible avec un balisage en HTML*.

2. Modèle de navigation

Dès le départ la possibilité d'intégrer l'article, afin de pouvoir naviguer à partir des citations d'ALSIC, ne pouvait se faire que si celui-ci était au format XML. Il a donc fallu proposer deux modèles de navigation : un premier modèle avec un serveur simple contenant uniquement des signalements de notices (et un accès au texte intégral) en attendant que les articles soient convertis en XML et un deuxième modèle permettant la navigation entre le texte intégral des articles en XML et d'autres ressources.

?? **1ère phase**

Compte-tenu de la particularité de ce serveur, qui intégrera alors deux types de ressources de formats et de composition différents, nous avons proposé deux nouvelles fonctions : l'accès au texte intégral à partir de la liste des documents pertinents et la navigation entre citation et document cité (2ème phase).

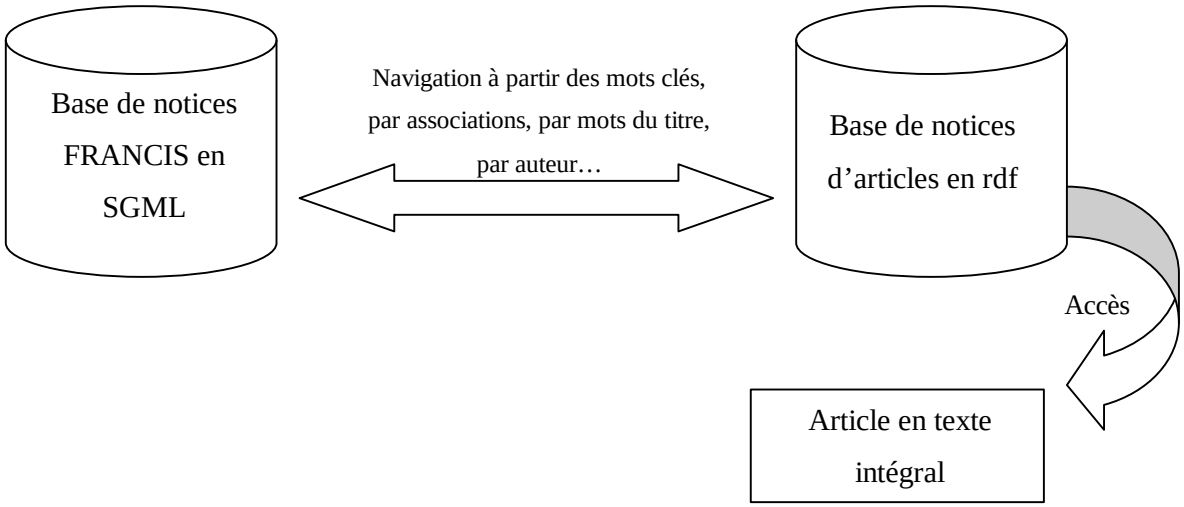


Figure 2: Le modèle de navigation classique

?? **2ème phase**

Une fois la revue et les articles en XML, le serveur devra intégrer les articles en texte intégral et permettre ainsi une navigation à partir des citations.

Pour l'instant, le premier serveur ne contient que des signalements et des notices. Cependant, les ressources on été choisies en envisageant la deuxième phase et le travail de sélection de notices a alors porté sur les citations afin de proposer par la suite un modèle de navigation à partir des citations. S'agissant d'une plate forme de navigation sur des revues électroniques, nous pensons en effet que ce type de fonctionnalité peut apporter une grande valeur ajoutée.

De manière générale, notre travail sur les citations a permis de faire ressortir plusieurs types de liens possibles entre les citations et d'autres documents:

- Lien entre les citations et des notices FRANCIS déchargées en SGML avec Miriad ou à partir de l'interface Bibliosciences* (à reformater en sgml)
- Lien entre les citations et des documents électroniques (articles) sur des sites internet.
- Lien entre des citations dans des articles et des analyses de livres dans la même revue ou liens entre des articles des deux revues ALSIC et LLTJ

Le modèle de navigation que nous proposerons alors pour ce type de serveur (à l'issue du stage) devra permettre de naviguer entre ces différentes ressources à partir de ces éléments de citation.

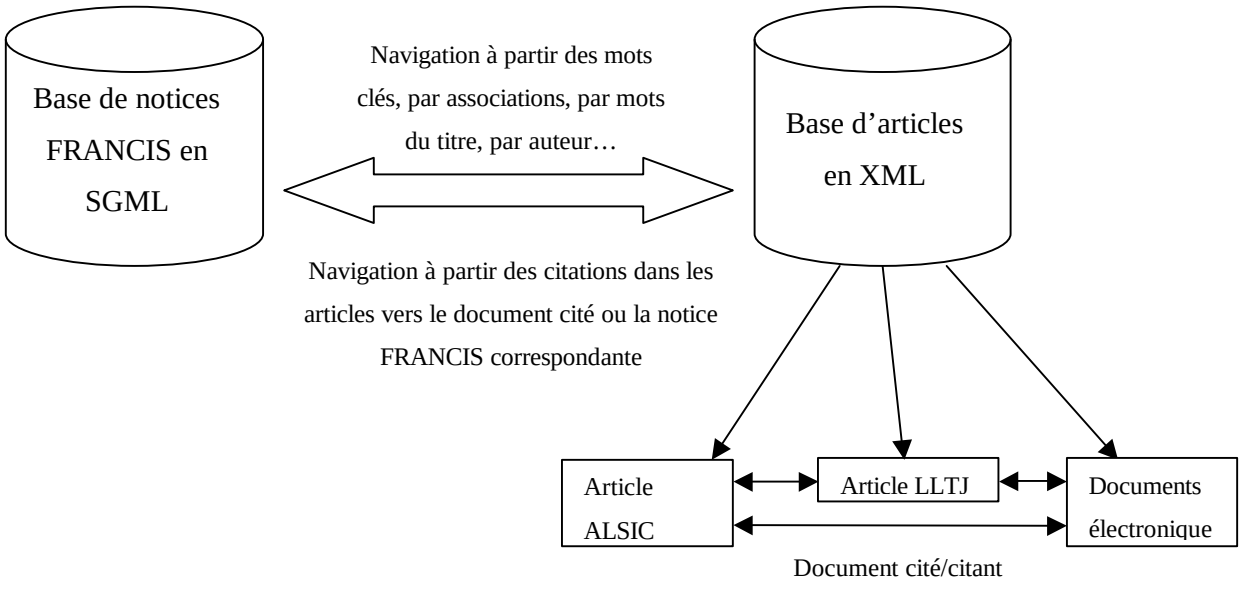


Figure 3: Modèle de navigation avec texte intégral en XML

Une fois le modèle de navigation ébauché, nous avons alors pu passer à la conception du serveur. La première phase a consisté dans la création des notices permettant de signaler les articles dans le serveur.

C. Déchargement et conception des notices

1. Les types de notices

Une fois le modèle de navigation décidé, la question a alors été de définir le type de notices qui constitueront les deux bases. Nous avons opté pour l'intégration de notices des bases de l'INIST dans un premier temps. L'interrogation de la base de données* FRANCIS a porté sur le titre quand il s'agissait d'une citation faite dans l'article ou sur les mots clés dans l'autre cas. Les notices FRANCIS ont ainsi été déchargées à partir de Miriad au format SGML et à partir de l'interrogation de Bibliosciences* qui contient des notices au format de diffusion « FRANCIS ancien » ce qui a nécessité un reformatage vers le format SGML Miriad.

D'autre part, pour constituer la base ALSIC et articles, nous avons conçu des notices en RDF et Dublin Core* qui peuvent non seulement servir à signaler les articles dans le serveur mais aussi être récupérées comme métadonnées lorsque les articles seront en XML. Ce type de notices en RDF permet de décrire et signaler les articles ALSIC et les documents électroniques avec des termes correspondants au vocabulaire FRANCIS de linguistique[6]. Ainsi, il est possible, à partir de ces notices, d'intégrer des liens vers le texte intégral ce qui n'est pas le cas avec de simples notices FRANCIS, puisqu'il n'y a pas la possibilité d'ajouter de nouveaux champs.

Afin que nous puissions, à partir des notices décrivant les articles et documents électroniques, naviguer vers des notices FRANCIS (par les descripteurs, les citations ou les auteurs) et pointer vers les documents en texte intégral, nous avons alors créé des notices RDF avec des Meta-données en Dublin Core* pour chacun de ces articles et documents électroniques.

[6] Cela afin d'avoir une homogénéité terminologique entre les bases du serveur.

2. Normes Dublin Core et RDF[7]

Le Dublin Core (DC) est une norme permettant de décrire les caractéristiques essentielles de documents électroniques. La croissance d'Internet a en effet accru le nombre de ressources mais elles restent souvent difficiles à localiser. D'autre part, la disparité des modèles et standards pour la description des ressources dans les différentes communautés (bibliothèques, musées, systèmes d'information...) rend difficile la recherche de ressources à travers plusieurs disciplines. Il était donc nécessaire de définir un formalisme simple et universel permettant de décrire des ressources quelles qu'elles soient dans un format compatible avec les standards existants. Elaboré depuis 1995, le DC semble être le standard sur lequel la communauté internationale voudrait se mettre d'accord. Son rôle est la création de métadonnées qui devraient être exploitées par les moteurs de recherche afin d'améliorer la précision des recherches effectuées sur le web. Il permet de mettre l'accent sur la sémantique en proposant une syntaxe assez simple. Le DC définit ainsi un ensemble de 15 éléments (title, author, subject, description, format…) constituant une base ("core") de métadonnées générales, indépendantes du domaine d'application. Il est conçu autour de 5 principes fondamentaux qui sont que tous les éléments sont optionnels, répétables, et qu'il est extensible, multidisciplinaire et international. Si nécessaire, chaque élément peut être précisé de deux manières: à l'aide de "qualifieurs" ou "attributs" fixant le contexte dans lequel la valeur d'un élément doit être interprétée et à l'aide de sous-éléments qui spécifient les facettes d'un élément.

Le *Resource Description Framework* (RDF) (Structure de description des ressources) est un modèle ou 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 les 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. Aussi, RDF a besoin de la facilité des espaces de nom* XML pour associer précisément chaque propriété avec le schéma qui définit la propriété.

---

[7] Voir les recommandations sur ces normes : http://dublincore.org/documents/dces/ et http://www.w3.org/TR/REC-rdf-syntax/

3. Création de la structure des notices

Comme nous l'avons dit, les recommandations du DC définissent 15 éléments de base, les éléments qualifiés et le vocabulaire type. A partir de ces recommandations, des besoins de navigation du serveur et des informations que nous avions, il était donc nécessaire de réfléchir au préalable à une structure pour les notices à créer. Afin d'avoir des notices utilisables par le serveur, nous avons donc conçu leur structure en RDF et pensé au contenu des balises Dublin Core. Pour cela, nous nous sommes appuyés sur les recommandations concernant notamment les éléments de base, les éléments qualifiés et le vocabulaire type.

Ont été ainsi créées deux bases de formats différents, une base de notices FRANCIS en SGML (en collaboration avec les ingénieurs du service scientifique SHS du département pour le vocabulaire) et une base de notices en RDF pointant vers des articles. La navigation dans le serveur étant basée sur les mots du titre (français et anglais), les descripteurs (français et anglais) et les auteurs, nous avons alors adapté les éléments du Dublin Core en integrant des attributs qui nous permettent par exemple de naviguer entre les langues.

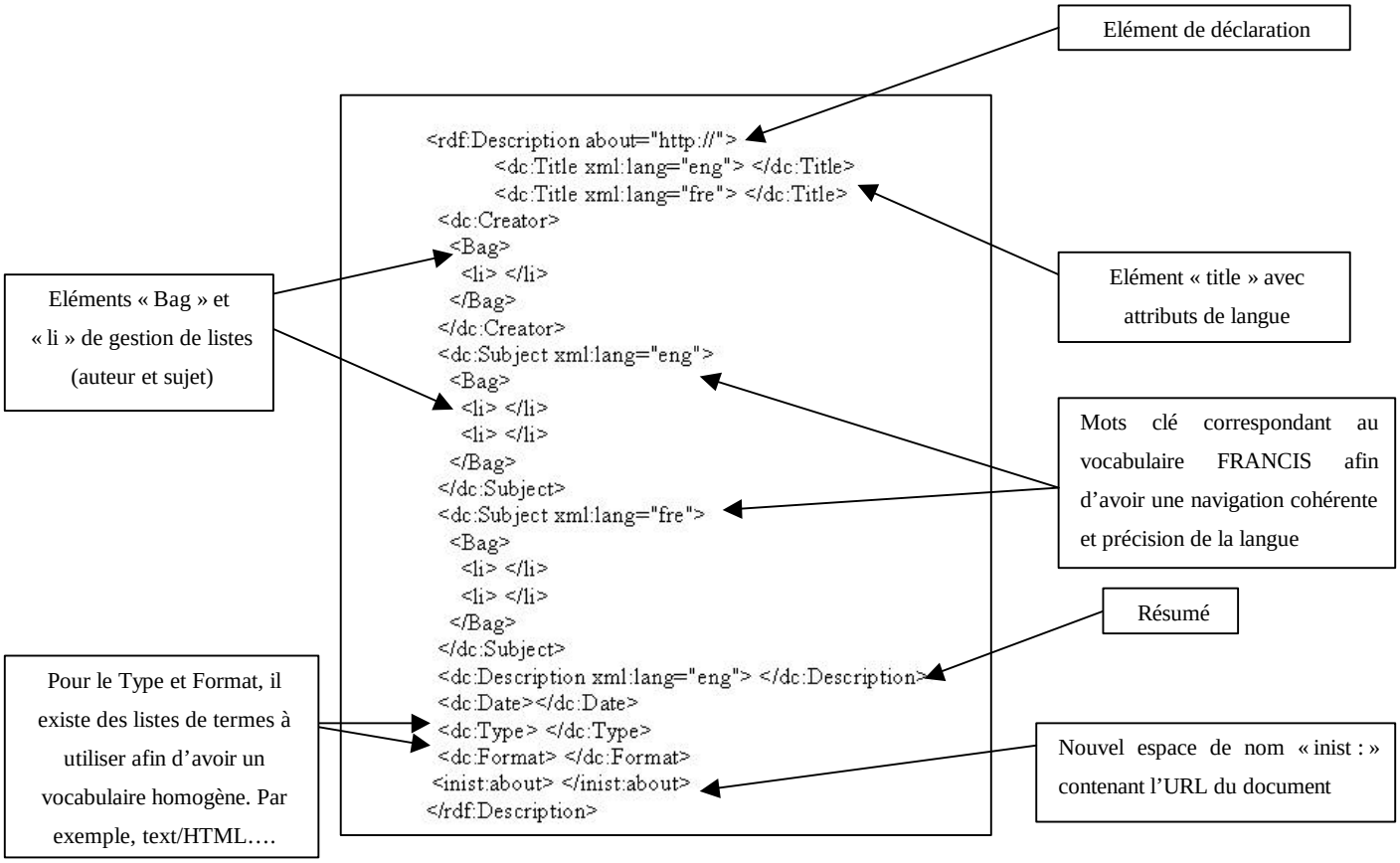


Figure 4: La structure des notices en RDF et Dublin Core

La structure en RDF nous permet de gérer des listes. Les balises « Subject » (mots-clés) et « Creator » (auteur) devaient en effet intégrer des éléments répétitifs. Ce problème a pu être résolu avec l'utilisation de l'espace de nom* RDF « Bag » qui est multiséparateur et et de la balise « li » qui identifie un élément de liste. Aussi, dans l'optique de pouvoir pointer vers les documents, nous avons ajouté une nouvelle balise RDF about dans l'espace de nom « inist » contenant l'adresse vers le texte intégral de l'article signalé.

Nous avons ainsi créé une vingtaine de notices en RDF qui signalent les articles d'ALSIC, les articles de LLTJ* et des documents électroniques trouvés sur Internet. Ces notices permettront de naviguer entre des ressources FRANCIS et d'autres types de documents n'ayant pas la même nature ni le même format.

Nous obtenons ainsi :

- Un corpus composé de notices en RDF signalant les articles d'ALSIC et documents électroniques
- un corpus de notices FRANCIS déchargées à partir de Miriad et concaténées en XML.

Une fois ces ressources identifiées et rassemblées, nous avons pu passer à la création du serveur en lui-même.

D. Paramétrage et création du serveur

1. Création du serveur

La création du serveur contenant ces deux corpus préalablement préparés doit passer par un paramétrage qui se définit dans le fichier desc.ed (voir annexe 1) . Il permet de définir les index* souhaités, les chemins d'accès nécessaires à la construction des bases, de définir les ressources et de paramétrer le schéma général de navigation (les croisements entre les bases puisqu'il s'agit d'un serveur multi base). Le serveur présenté a donc été constitué à partir de deux bases déclarées dans le paramétrage tel que le montre la figure 5.

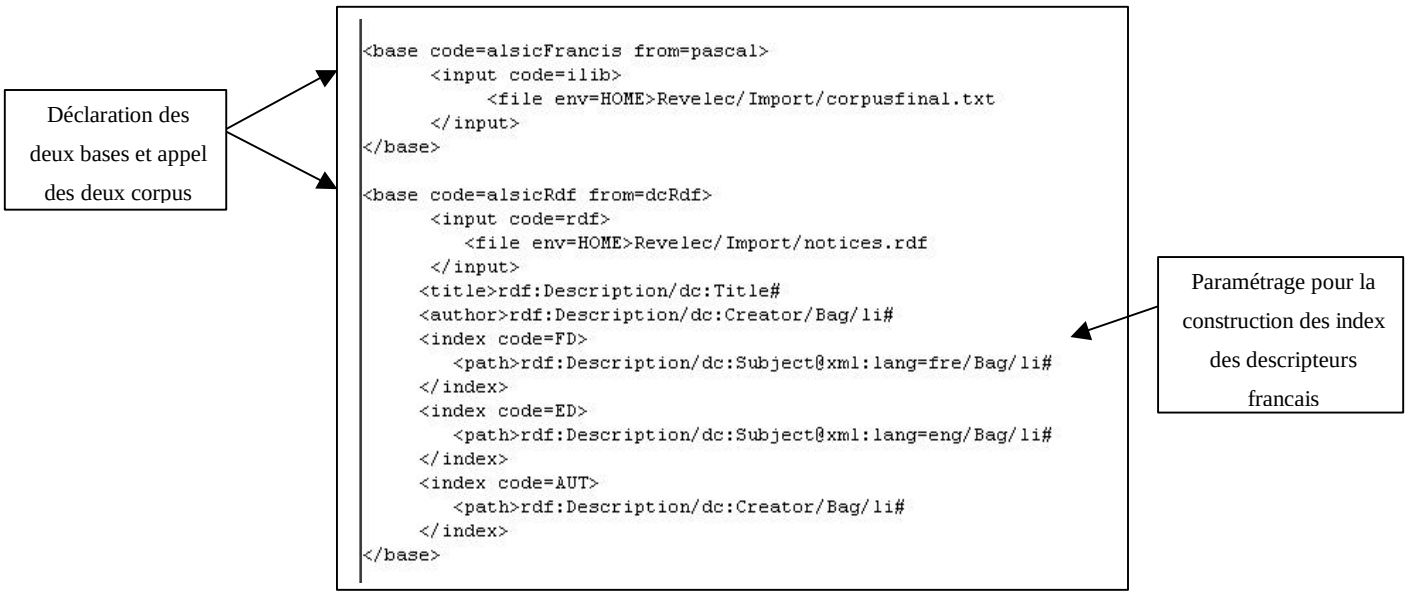


Figure 5 : le fichier desc.ed ou document paramètre

Une fois le paramétrage des bases effectué et les chemins nécessaires à la construction des index déclarés nous avons alors généré le serveur. Il permet de naviguer à partir de chaque base ou en accès multi-base, par des entrées d'index sur les descripteurs français, anglais, les mots du titre et les auteurs.

Figure 6: Les bases et index du serveur REVELEC. Exemple du multi

Code base	*auteurs*	*titre*
alsicFrancis	BERGER (Jean-Francois), RIEBEN (Pierre)	Environnements interactifs d'apprentissage sur Internet. Stratégies de conception et expérimentations
alsicFrancis	GIARDINA (M.), LAURIER (M.), MEUNIER (Claire), JACQUINOT (Geneviève)	Modélisation de l'apprenant et interactivité
alsicFrancis	BRIEN (R.), BOURDEAU (J.), ROCHELEAU (J.), MEUNIER (Claire), JACQUINOT (Geneviève)	L'interactivité dans l'apprentissage : la perspective des sciences cognitives
alsicFrancis	WEISSBERG (J.-L.), MEUNIER (Claire), JACQUINOT (Geneviève)	Retour sur interactivité
alsicFrancis	CHARLIER (P.), MEUNIER (Claire), JACQUINOT (Geneviève)	Interactivité et interaction dans une modélisation de l'apprentissage
alsicFrancis	PAQUETTE (G.), MEUNIER (Claire), JACQUINOT (Geneviève)	L'ingénierie des interactions dans les systèmes d'apprentissage
alsicFrancis	DIAZ (P.), AEDO (I.), TORRA (N.), MIRANDA (P.), MARTIN (M.), BRNA (Paul), DICHEVA (Darina)	Meeting the needs of teachers and students within the CESAR training system
alsicRdf	Lydia PLOWMAN	The 'Primitive Mode Of Representation' and The Evolution Of Interactive Multimedia
alsicRdf	Didier PAQUELIN	Analyse d'applications multimédias pour un usage pédagogique. A la recherche de l'intentionnalité partagée
alsicRdf	Marie-Noelle LAMY, Robin GOODFELLOW	"Conversations réflexives" dans la classe de langue virtuelle par conférence asynchrone

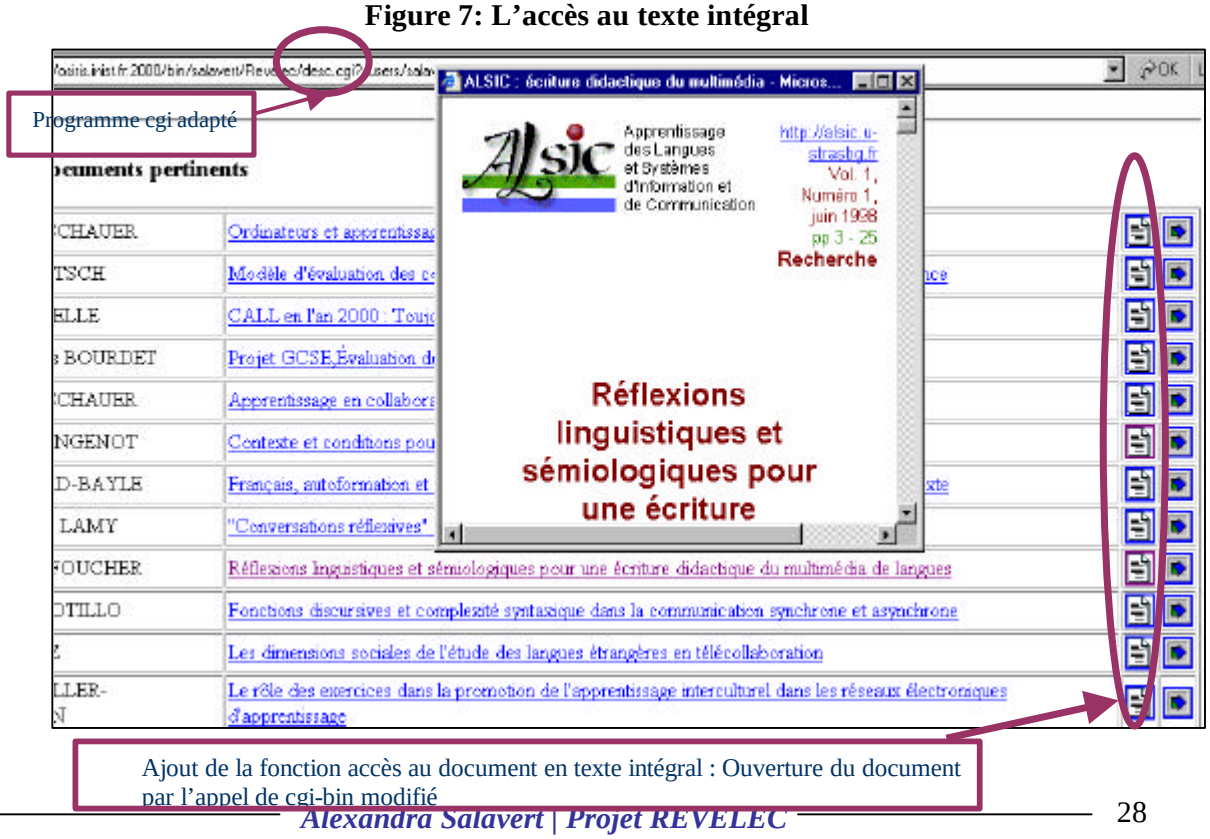
Une fois le serveur créé, nous avons alors pu envisager les différentes évolutions à réaliser. En effet, la navigation simple dans des notices n'est pas suffisante pour un serveur de revues électroniques, **un accès au texte intégral de l'article en ligne s'avère nécessaire**.

2. Programmation des cgi-bin

Ce problème de l'accès au texte intégral pourra être résolu lorsque les articles auront été transformés en XML et intégrés au serveur en texte intégral, puisque l'on pourra alors naviguer dans l'article en lui-même.

En attendant, il était intéressant d'intégrer un lien vers l'article afin de pouvoir l'afficher à partir de sa notice dans le serveur d'investigation. Les pages du serveur étant générées dynamiquement, le développement de cette fonction n'a pu se faire qu'avec la programmation de cgi-bin*. Ces programmes ont consisté à récupérer le contenu de l'espace de nom « inist » créé dans les notices RDF, à savoir l'URL du document, pour chaque requete lancée. L'icône texte ajoutée permet ainsi de générer une URL dynamique ce qui permet, à partir de la liste des documents pertinents dans le serveur en fin de navigation, de pouvoir accéder directement à l'article sur Internet. Nous nous sommes plongé dans l'enchaînement des programmes DILIB permettant la création des cgi* et les avons créé ou paramétré pour la base ALSIC puis pour le MULTI. Le langage de commande utilisé pour le paramétrage et la programmation est un mélange de langage C et de commandes DILIB. Cette modification de programmes et le paramétrage des cgi donne une nouvelle dimension aux serveurs d'investigation, l'accès au texte intégral.

Figure 7: L'accès au texte intégral



Ce type de serveur a ainsi été développé en fonction des documents particuliers que sont les revues électroniques avec un accès indispensable au texte intégral. Nous reviendrons sur le serveur lorsque les articles auront été XMLisés puisqu'il nous faudra alors les intégrer en eux mêmes. Nous pourrons alors suivre le plan de navigation précédemment défini et établir les liens entre les articles et les notices à partir des citations sur lesquelles nous avons déjà travaillé. Pour l'instant, nous intégrerons le serveur tel quel sur une plate forme et une page web.

E. Plate forme web

L'INIST doit proposer l'accès aux revues électroniques à la communauté des chercheurs. Loin de constituer un remplacement des revues « papier », ces revues leur sont en effet complémentaires et les utilisateurs du portail de l'INIST (ConnectSciences[8]) sont en attente de ce genre de services. Ainsi, il s'agira d'intégrer le serveur d'investigation et la revue elle-même sur le portail. En suivant la charte graphique, nous avons alors créé une page web (avec GOLIVE 5.0) proposant ces services à partir du portail de l'INIST dans une nouvelle rubrique « Revues électroniques ». Nous y proposerons deux types d'accès. Un accès à forte valeur ajoutée permettant de naviguer entre plusieurs types de ressources, des ressources INIST (type base de données FRANCIS) et des ressources articles, mais aussi un accès plus classique à la revue en texte intégral. Outre ALSIC, nous proposons aussi un lien vers la revue américaine LLTJ* qui est thématiquement très proche et que nous avons intégré dans le serveur. Nous avons ensuite réalisé une petite rubrique d'aide accessible à partir de la page des revues en sciences du langage qui explique à l'utilisateur l'intérêt d'un serveur de revues électroniques et surtout son fonctionnement. Il a alors fallu réfléchir aux points à mettre en avant concernant l'utilisation du serveur REVELEC. Compte tenu du public auquel s'adresse ce type de service, à savoir une communauté de chercheurs spécialistes du domaine, les clusters et associations de termes peuvent être intéressants puisqu'ils permettent d'élargir et de trouver d'autres sources. Il a donc fallu bien expliquer la marche à suivre et les différentes opérations proposées par un serveur d'investigation.

[8] http://connectsciences.inist.fr/

Figure 8 : Page Web pour Connect Sciences

La création et l'intégration (dans le portail de l'INIST) du premier serveur REVELEC qui contient alors les signalements des notices étant terminées, il nous faut alors réfléchir à la normalisation des ressources et à leur passage en XML. Toute cette première partie du stage aura permis de définir les ressources et leur organisation pour un traitement dans un serveur de revues électroniques. Aussi, nous avons pu soulever certaines problématiques propres à l'évolution des nouvelles technologies sur Internet comme l'utilisation du Dublin Core et sa structuration en RDF. En effet, l'utilisation de ce type de normes et de formats, qui ont un rôle d'indexation et qui nous ont permis de créer des notices, s'est avéré très intéressante puisqu'ils pourront être réutilisés pour gérer les métadonnées des articles sur Internet.

La prochaine étape nous permettra de proposer à T. Chanier un modèle de document structuré pour les articles et pour sa revue. Aussi, nous pourrons créer par la suite un deuxième serveur contenant le texte intégral dans lequel il sera possible de naviguer à partir des citations. Aussi, à travers certaines problématiques posées par les outils de navigation dans des revues électroniques, nous nous interrogerons sur la chaîne de traitement d'une revue électronique au format XML en dégageant des questions indispensables pour le traitement des articles, telles leur identification locale, leur archivage…

III. TRAITEMENT DE LA REVUE EN XML

La chaîne de traitement documentaire aboutissant à la publication de la revue utilise des formats normalisés pour structurer et présenter l'information. Pour l'instant, les recommandations aux auteurs[9]. leur proposent de faire un document Word selon certains critères pour générer des articles HTML T. Chanier souhaite que sa revue passe au format XML. Ce format nous permettra en effet d'avoir des documents structurés et de mettre en place une navigation basée sur les citations. Nous avons alors envisagé les différentes problématiques que ce changement de format impose et proposé des solutions et perspectives pour la conversion. La méthodologie adoptée a été la suivante :

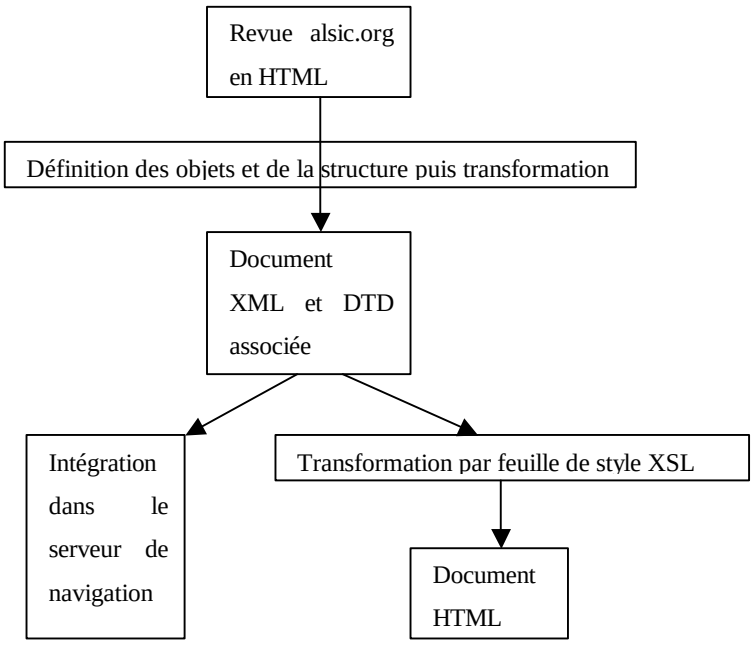


Figure 9 : Passage de la revue existante en XML

Aussi, le but étant par la suite d'avoir une revue générée dès le départ en XML, toute la chaîne de traitement, d'identification et de publication doit être repensée.

[9] http://alsic.u-strasbg.fr/Infrevue/auteurs.htm

A. Le document structuré en XML

1. Standard XML et DTDs

XML (*eXtensible Markup Language*, ou Langage Etendu de Balisage) est un langage recommandé en 1998 par le *W3 Consortium** et utilisé sur le web. Comme HTML (*Hypertext Markup Language)*, XML est un langage de balisage (*markup*), c'est-à-dire un langage qui structure l'information au moyen de *balises*[10]. En tant que format structuré, il permet donc de rendre lisible la structure et le contenu d'un document. Chaque élément se voit attribuer des balises qui le délimitent et l'identifient. Contrairement à HTML, qui présente un jeu limité de balises orientées présentation (titre, paragraphe, image, lien hypertexte, etc.), un langage structuré comme XML (ou *métalangage*) va permettre d'inventer à volonté de nouvelles balises pour isoler toutes les informations élémentaires ou agrégats d'informations élémentaires que peut contenir une page Web. (Source www.chez.com/XML).

Aussi, le principal avantage d'un passage de HTML à XML (dans le cas de la revue ALSIC également) est de pouvoir séparer la structure logique du document de son affichage et ainsi de créer des documents structurés indépendants de leur support. D'un point de vue éditorial, l'intérêt d'utiliser un format normalisé et international est double : l'INIST peut intégrer au serveur de documents électroniques des outils de traitement automatique de qualité et la revue peut décider d'échanger une partie de son fonds avec des centres universitaires de traitement spécialisés à l'étranger.

Descendant du SGML, XML représente avant tout un format d'échange. Il décrit n'importe quel type de données et permet de modéliser un document grâce à des balises qui en décrivent la structure et le contenu. L'intérêt du langage XML est alors pour nous de baliser le contenu et/ou la structure des articles d'ALSIC afin de faire des traitements et requêtes à l'intérieur de ce document. Nous obtiendrons alors un article final propre, balisé en fonction de sa structure et non de sa présentation et nous pourrons de ce fait automatiser certaines tâches telle la constitution d'index.

[10] Alain Michard, dans *XML langages et applications*, définit le XML comme « un langage de description et d'échange de documents structurés » qui « permet de décrire la structure logique de documents, à l'aide d'un système de balises permettant de marquer les éléments qui composent la structure et les relations entre ces éléments. » page 1.

Un document XML peut avoir sa structure ou sa grammaire définie dans une DTD* (*Definition Type Document* ou *Définition de Type de Document*). Les documents XML réalisés à partir d'une DTD et conformes à celle-ci sont dits « valides », c'est-à-dire qu'ils obéissent à une structure type définie. La structure arborescente du document XML (imbrications des balises, caractère obligatoire ou facultatif des balises et de leur ordre de succession…) basée sur le modèle DOM (racine, nœud parent, nœud fils, nœud frère) est en effet déclarée formellement dans le corps du document XML ou dans un fichier à part. C'est cette déclaration que l'on appelle DTD. Elle s'effectue selon un formalisme particulier normé lui-aussi dans la spécification XML. En XML cette déclaration est facultative, ce qui donne une grande souplesse aux développeurs. La DTD définit donc le nom des éléments, leur contenu, l'ordre et la hiérarchie de ces éléments, les attributs et le nom des entités utilisées.

2. Choix de la DTD

On peut construire sa propre DTD ou se baser sur une DTD existante. Nous avons choisi cette deuxième option dans l'optique de contribuer à une certaine homogénéité des revues électroniques en XML sur Internet. Nous avons alors comparé et étudié les différentes DTD par rapport à notre type de besoin.

Premièrement, il a fallu identifier la structure et les objets des articles HTML. Le choix d'une DTD se base en effet sur les possibilités de balisage et de structuration qu'elle offre par rapport aux types d'objets à modéliser. Nous avons donc tout d'abord procédé à l'identification des balises HTML utilisées et des différentes rubriques des revues ALSIC et LLTJ afin d'en lister les objets.

Pour choisir la DTD la mieux adaptée, une confrontation de plusieurs DTDs par rapport aux besoins identifiés est indispensable. Il existe en effet plusieurs DTDs normalisées ou recommandées et que l'on utilise pour structurer des articles. **La norme ISO 12083** (*Electronic Manuscript Preparation and Markup*) est certainement la plus répandue puisqu'elle définit un ensemble de 4 DTD qui sont issues des travaux de l'*Association of American Publishers* (AAP). On peut les utiliser pour représenter des livres, des articles de périodiques, des périodiques au complet ou des équations et formules mathématiques. Nous

nous sommes alors penchés sur cette DTD ISO 12083[11] article et l'avons comparé à la **DTD TEI**[12] (*Text Encoding Initiative* développée en 1987 pour les domaines de la linguistique et des sciences humaines) ainsi qu'à la **DTD Docbook**[13] qui a été créée pour décrire les livres électroniques. Après avoir analysé les différents éléments de ces DTDs et la finesse de structuration du document qu'ils permettent, nous avons alors tiré plusieurs conclusions.

Nos critères de sélection ont porté sur :

- La richesse de description des éléments par rapport au document original et aux objets à structurer.
- La qualité de signalement des citations dans le texte puisque l'objectif sera de faire des traitements sur ces types d'objets dans le serveur.
- La souplesse : le fait de pouvoir rajouter et restructurer des éléments

La DocBook est ressortie comme étant la plus adaptée pour notre besoin. En effet, T. Chanier envisage d'intégrer des éléments multimédia à ALSIC. La TEI ne le permettant pas a donc été éliminée. Par rapport à la 12083, **la Docbook est beaucoup plus riche** car :

- Elle comporte un grand nombre d'éléments et semble particulièrement adaptée aux articles, malgré son objectif principal et son nom.
- Elle prévoit l'intégration et la manipulation du multimédia.
- Elle permet d'avoir une structure avec des pointeurs sur les objets, les URL, les citations.
- La finesse de structuration des citations est très intéressante.
- Elle permet de fabriquer des documents au format SGML ou XML puis de les publier dans une multitude de formats : HTML, RTF*, PDF*.

Une fois la Docbook sélectionnée, nous avons alors conçu un modèle de document structuré par rapport au format initial des articles en HTML (voir annexe 2). Cette structure

[11] Concernant la DTD 12083 voir http://www.xmlxperts.com/articledtd.htm
[12] Concernant la DTD TEI voir http://www.tei-c.org/P4X/DTD/
[13] Concernant la DTD Docbook voir www.docbook.org et possibilité de télécharger le livre de référence sur http://www.docbook.org/tdg/en/html/docbook.html

nous permet en effet de normaliser et reformater les documents initiaux et de donner une base pour la réflexion autour de la chaîne de traitement.

3. Conception de la structure et du balisage

Une fois les différents objets recensés et la DTD choisie, nous avons pu concevoir la structure XML des articles (voir annexe 2). L'important a alors été de choisir des éléments et balises qui pourraient structurer le contenu des articles selon la DocBook et concevoir une structure qui soit assez proche du texte actuel en HTML afin que le reformatage ne soit pas trop complexe.

Comment appliquer une DTD ? Il faut tout d'abord analyser les différents éléments de la DTD et choisir ceux qui pourraient être utiles. Ensuite, même si ça n'est pas obligatoire chaque élément doit être prioritairement employé tel que la DTD le préconise, c'est-à-dire en respectant la structure de l'arbre mais aussi les caractéristiques optionnelles et/ou répétitives de ces éléments ainsi que de leurs attributs.

Quels sont les objets des articles HTML ALSIC et comment la DTD y répond-elle ? Nous nous sommes en effet basés sur la structure des documents existants. En analysant les différents objets en HTML (en-tête, figure, tableau, image, citation, titre, paragraphe, exemple, liste….), nous avons pu sélectionner les éléments de la DocBook à utiliser et proposer un modèle type de document XML basé sur cette DTD qui définit les liens de parenté entre les éléments. Même s'il n'a pas toujours été possible de suivre ce que prévoyait la DocBook dans la manière d'imbriquer nos balises, l'intérêt de cette DTD était justement de pouvoir la modifier. Ces changements ont alors été déclarés dans la DTD finale.

En nous basant sur le texte HTML de départ et sur nos objectifs en matière de traitement du document XML, nous avons alors établi un **modèle de document**. Pour nos objectifs concernant la navigation dans l'article, nous avons balisé plusieurs types d'objets et particulièrement les références bibliographiques et les citations. En effet, la Docbook permet de repérer dans un texte une entrée dans la bibliographie renvoyant à une citation identifiée et de la structurer. Ce marquage nous permettra par la suite d'établir des tables de références à partir de l'auteur, l'année et le titre.

Une fois la structure du document XML arrêtée, comment passer du document HTML vers XML ? Il est tout d'abord nécessaire de choisir les outils pour ce type de traitement.

B. Conversion du HTML en XML

1. Conception des programmes LEX permettant la conversion des articles en XML (voir annexe 2)

LEX* est un générateur d'analyseur lexical qui permet d'associer des actions à des règles de reconnaissance de formes lexicales. Il permet de traiter des chaînes de caractères dans des flots de données. 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.

Nous avons commencé par faire un premier essai d'automatisation en LEX pour la conversion des articles HTML vers XML. La question a donc été de savoir comment passer d'un document HTML à un document structuré ? Les difficultés à prendre en considération pour une transformation automatique concernent :

- La longueur des articles, 40 pages environ.
- La multiplicité des objets (l'en-tête, la table des matières, le texte en lui même, les titres…..).
- La généralisation de l'automatisation à plusieurs articles HTML différents qui ne sont pas toujours conçus avec la même structure et qui n'ont pas tous les mêmes objets.
- La nécessité de nettoyer le document HTML de départ (certaines balises n'étant pas fermées par exemple).

Pour résoudre ces difficultés, nous avons développé 5 programmes LEX qui s'enchaînent grâce à un shell (voir annexe 2). Certaines balises HTML ont été transformées en balises XML, d'autres ont été supprimées. Nous avons enchaîné un grand nombre de règles afin de traiter les différents objets pour les différents documents HTML.

Cette phase d'automatisation en LEX nous a permis de nous rendre compte des erreurs commises notamment quant à la façon d'aborder le problème. Nous ne sommes pas parvenu à faire une transformation automatique complète et à produire un article XML valide. Par exemple, nous avons été confrontés à des problèmes de structure puisque certaines balises ouvertes n'étaient pas forcément fermées. LEX permet en effet de faire des traitements ligne par ligne mais ne s'avère pas sûr pour produire un document structuré valide. Les outils de traitements de chaînes de caractères ne permettent pas de structurer un document de manière logique mais d'effectuer des remplacements de balises. Nous avons alors réfléchi aux types d'outils à mettre en place. Une première structuration à l'aide d'outils plus puissants se serait avérée nécessaire. Les outils et commandes DILIB auraient été intéressants puisqu'ils permettent de structurer un document logiquement c'est-à-dire en encadrant des parties de texte repérées entre balises. Il aurait alors fallu baliser les différentes sections du document avec ce type d'outils et créer une arborescence, pour ensuite appliquer des chaînes de traitement en LEX. Une fois la structure générale balisée, le texte pouvant alors être repris en LEX dans le traitement des chaînes de caractères à l'intérieur des sections repérées.

Une fois le document XML généré, semi-automatiquement pour l'instant, nous avons réfléchi à deux types de problèmes : l'affichage et le « parsing » du document XML, et les différents traitements que nous pourrons effectuer dans les articles afin de pouvoir naviguer à partir des citations dans le document vers d'autres ressources identifiées dans la bibliographie.

2. Parsage, conception d'une feuille de style

XML permet de définir un format d'échange selon les besoins de l'utilisateur et offre des mécanismes pour vérifier la validité du document manipulé. Cette opération est possible à l'aide d'un outil appelé analyseur (en anglais *parser*, parfois francisé en *parseur*) qui permet d'une part d'analyser les données d'un document XML (on parle de *parsing* du document) ainsi que d'en vérifier éventuellement la validité.

XMLSpy? *** permet entre autres de parser et de valider un document XML**. Le logiciel extrait la structure du document et la compare à l'arborescence de la DTD associée. Plusieurs outils on été testés mais XMLSpy? est le plus complet, puisqu'il traite également la technologie XSL et XSL-FO*, et donc le mieux adapté. Il permet de vérifier la structure d'un

document avec sa DTD et de générer automatiquement une DTD particulière basée cependant sur la Docbook originale à partir de ce document validé.

Nous avons également testé des outils et commandes DILIB appropriés qui vérifient la structure d'un document balisé. Ceux-ci ont été améliorés au cours des tests. A partir de ce document XML valide, nous pouvons alors faire différents traitements et sélections d'éléments. Aussi, pour passer du XML à un format affichable, le XHTML nous utiliserons une feuille de style en XSL*.

Les feuilles de style XSL sont des documents XML composés de différents scripts ou règles de transformation appliqués aux éléments d'un document XML. Il s'agit d'un langage permettant d'associer des informations stylistiques à un document XML sous un navigateur quelconque. Intégrés dans le document XML, la feuille de style (voir annexe 2) a permis de générer un document HTML très proche des articles originaux.

Même si cette structure et la feuille de style associée ne seront pas utilisées telles quelles par la suite, cette phase nous a confronté à un type de technologie qu'il est nécessaire de maîtriser pour la manipulation d'objets XML et nous a amené ainsi à soulever certaines problématiques liées aux outils et langage à utiliser pour générer une feuille de style XSL*.

Toute cette phase de conversion et de mise en application de ces programmes sera révisée lorsque la DTD et donc la structure des documents XML seront validés par T. Chanier et l'INIST.

Une fois le document XML créé, testé et affiché (voir annexe 2), nous pouvons alors proposer un modèle plus affiné de ce que sera la maquette « finale » du serveur. En effet, dans la première phase du serveur nous en étions restés à une navigation simple entre plusieurs types de ressources et vers le texte intégral. Les articles étant eux-mêmes structurés, nous avons pu proposer et définir un modèle de navigation à partir d'éléments balisés tels que les citations ou références.

C. Traitement des articles dans le serveur : navigation avec les références

Les articles au format XML peuvent alors être utilisés dans notre serveur de navigation. En effet, l'idée était au départ du projet de pouvoir, à partir du texte intégral de l'article, naviguer entre citation et article cité ou entre article et article citant. Nous avons dans un premier temps créé un serveur simple afin d'identifier les ressources, de poser les problèmes et de cerner les besoins liés à ce type de serveur. Aussi, nous pouvons maintenant passer à l'étape suivante qui apportera réellement une valeur ajoutée à une revue électronique comme ALSIC, **la navigation entre le texte intégral, les citations et références et des notices correspondant à ces références**. La conception de la maquette de serveur devra prendre en compte plusieurs étapes. La première consistera tout d'abord à structurer et à stocker les articles afin de pouvoir travailler en local. Cela posera alors le problème important des identifiants.

1. Identification interne des articles et structuration arborescente

Le premier problème à résoudre a consisté à pouvoir **identifier** les articles dans un environnement local de manière à les localiser efficacement et rapidement lors de la navigation. Nous avons alors besoin d'une donnée d'identification pour chaque article et chaque notice. Nous avons donc importé tous les articles de la revue et les avons identifié dans une arborescence locale (*repository)*.

Ce problème de l'identifiant a d'ailleurs été soulevé à plusieurs reprises avec T. Chanier. En effet, un identifiant est également nécessaire pour signaler les articles sur Internet. Même si un article (ou ses métadonnées) peut-être retrouvé à partir du nom de l'auteur ou du titre, il est bon de le doter également d'un identifiant unique (au sens URN*). Pour C. Lupovici, « les identifiants sont une courte séquence numérique ou alphanumérique attribuée à une unité d'information et servant de manière univoque et permanente à désigner cette unité ». Ils sont importants « pour effectuer le lien entre le créateur de l'information et son utilisateur »[14]. Ils évitent par exemple les problèmes d'erreur 404 puisque ce concept

[14] C. Lupovici, « Le Digital Object Identifier », *BBF*, T. 43 n°3, 1998, pp. 49-54.

d'identification « recouvre à la fois l'identification et la localisation » en fournissant en effet une identification unique et pérenne du document. Aussi, l'identifiant sera constitué de deux parties : l'en-tête identifiant l'organisme qui gère officiellement cet identifiant, et assure ainsi la fiabilité des requêtes sur Internet, et la partie locale sur laquelle nous réfléchissons. Nous avons donc recherché les différents types d'identification existants sur Internet. Nous avons écarté pour l'instant le principe des OpenURL mais pourrons y revenir à la suite du stage

Le principe du DOI* qui est composé de deux parties séparées par un slash, nous propose un système d'identification des articles avec un préfixe type www.inist.fr/ et un suffixe identifiant l'objet en local. L'identifiant d'un article ALSIC serait ainsi constitué d'un id *namespace_identifier* (espace de nom identifiant) comme www.inist.fr, si l'INIST assure le rôle de *repository*, et d'un *local-identifier* (identifiant local).

Comment structurer cet identifiant local ? Nous penserons le problème en terme de stockage. Pour chaque article d'ALSIC, nous avons choisi de l'insérer dans une arborescence de répertoires et de fichiers qui permette de l'archiver et de le traiter. Nous avons alors décidé de construire une arborescence qui reprenne le schéma général de la revue, à savoir volume, numéro, place de l'article dans le numéro. Prenons par exemple un article d'ALSIC : volume 3, numéro 2, en deuxième position. La cote qui permettra de le localiser sera ALSIC :3 :2 :2. Aussi, cette même structuration sera utilisée pour stocker les notices et articles de la bibliographie autres que les articles ALSIC, à partir d'un répertoire biblio.dd avec un classement par année et par ordre alphabétique du nom de l'auteur. Nous avons ainsi un répertoire par article qui contiendra les différents fichiers relatifs à celui-ci dans une arborescence qui localisera et identifiera de manière unique chacun de ces fichiers (voir figure 8).

Pour chaque répertoire numéro d'ALSIC, un sommaire automatique est généré. Il est important que l'identifiant dont est doté un article ne serve pas uniquement à le localiser mais soit aussi utilisé pour la fabrication d'index et de sommaire. Dans le sous-répertoire contenant un article, nous avons alors plusieurs types de fichiers dont la notice en RDF, l'article en XML et en HTML, les images, c'est-à-dire tous les fichiers qui participent à l'affichage ou au traitement de l'article. Nous obtenons donc une arborescence de répertoires et de fichiers dans laquelle chaque article sera logiquement localisé et archivé.

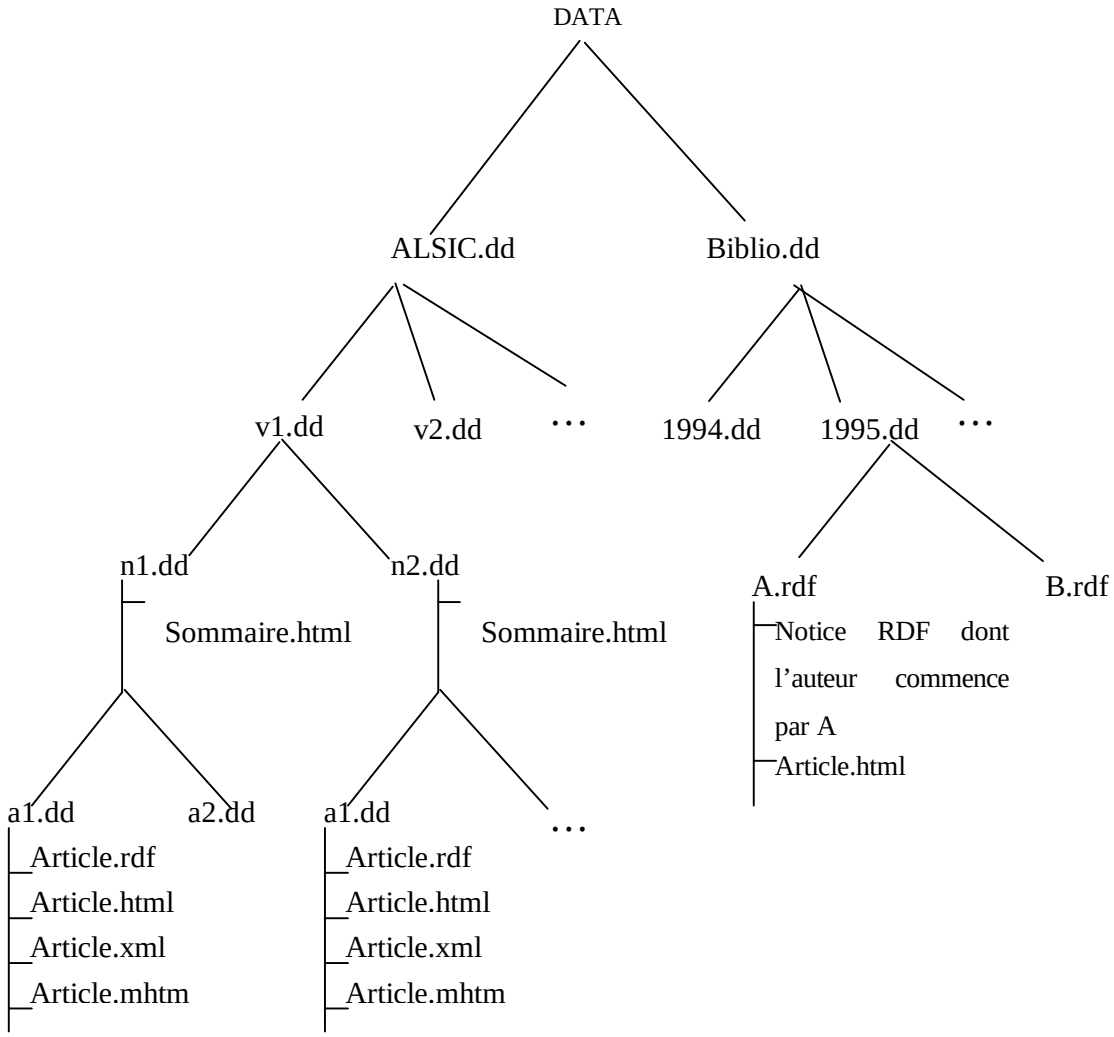


Figure 10 : Arborescence permettant de localiser et archiver les articles

Une fois l'arborescence créée et chaque article et ses fichiers localisés et identifiés, nous pouvons envisager un traitement sur les références et les citations dans le corps du texte XML.

2. Extraction des références et des citations

Nous avons deux types de marquage pour les citations dans le texte : la citation en elle-même, encadrée par des balises, et la référence dans la bibliographie à laquelle correspond la citation. Nous avons réfléchi à la manière de procéder pour que les citations

renvoyant à d'autres articles dans le serveur soient utilisées. Il serait intéressant de pouvoir naviguer entre toutes les références et ainsi rapprocher des articles cités des articles citants et de mettre en rapport tous ceux citant les mêmes sources. **Mais comment signaler les références n'ayant pas de notice dans le serveur** ? La solution qui permet de naviguer dans les citations et à partir des références a alors été de créer un programme qui extrait la bibliographie de chaque article XML afin de créer pour chaque référence une notice exploitable dans le serveur. Ensuite, nous avons crée des index sur ces références. Chaque référence citée est ainsi présente dans le serveur sous forme d'une notice RDF utilisable. Aussi, suite à diverses discussions autour de l'intérêt de conserver un serveur multi-base[15], nous avons décidé de faire un serveur mono-base et de ne faire que des notices RDF pour l'ensemble des types de documents à utiliser. Nous avons alors procédé à l'encapsulage des notices FRANCIS dans une structure RDF. Nous pouvons désormais faire des requêtes sur les références et naviguer du texte intégral de l'article vers les références citées (notice ou article).

Divers problèmes se sont posés comme celui des doublons. En effet, les références étant extraites de chaque article, plusieurs notices ont été fabriquées pour une même référence lorsque celle-ci était citée dans plusieurs articles. Nous devons alors **réfléchir au traitement de ces doublons dans le serveur**. L'autre problème sur lequel nous devons également nous pencher concerne les ressources FRANCIS. Il serait en effet intéressant de pouvoir accéder à l'article en lui-même si celui-ci est disponible sous forme numérique à l'INIST.

Au terme du stage, le serveur de navigation integre donc les revues électroniques aux références bibliographiques classiques. Afin de mieux comprendre son intérêt et son fonctionnement, nous avons reproduit ci-dessous le schéma général de navigation.

[15] L'intérêt d'un serveur Multi-base est de faire des croisements entre plusieurs bases de formats différent. Or, nous avons la possibilité d'uniformiser les formats de nos deux bases pour n'avoir qu'un mono-base qui n'enleverait rein à notre serveur puisque la navigation ne se base pas sur une mise en correspondance de deux corpus différents mais porte sur les références.

3. Schéma de navigation

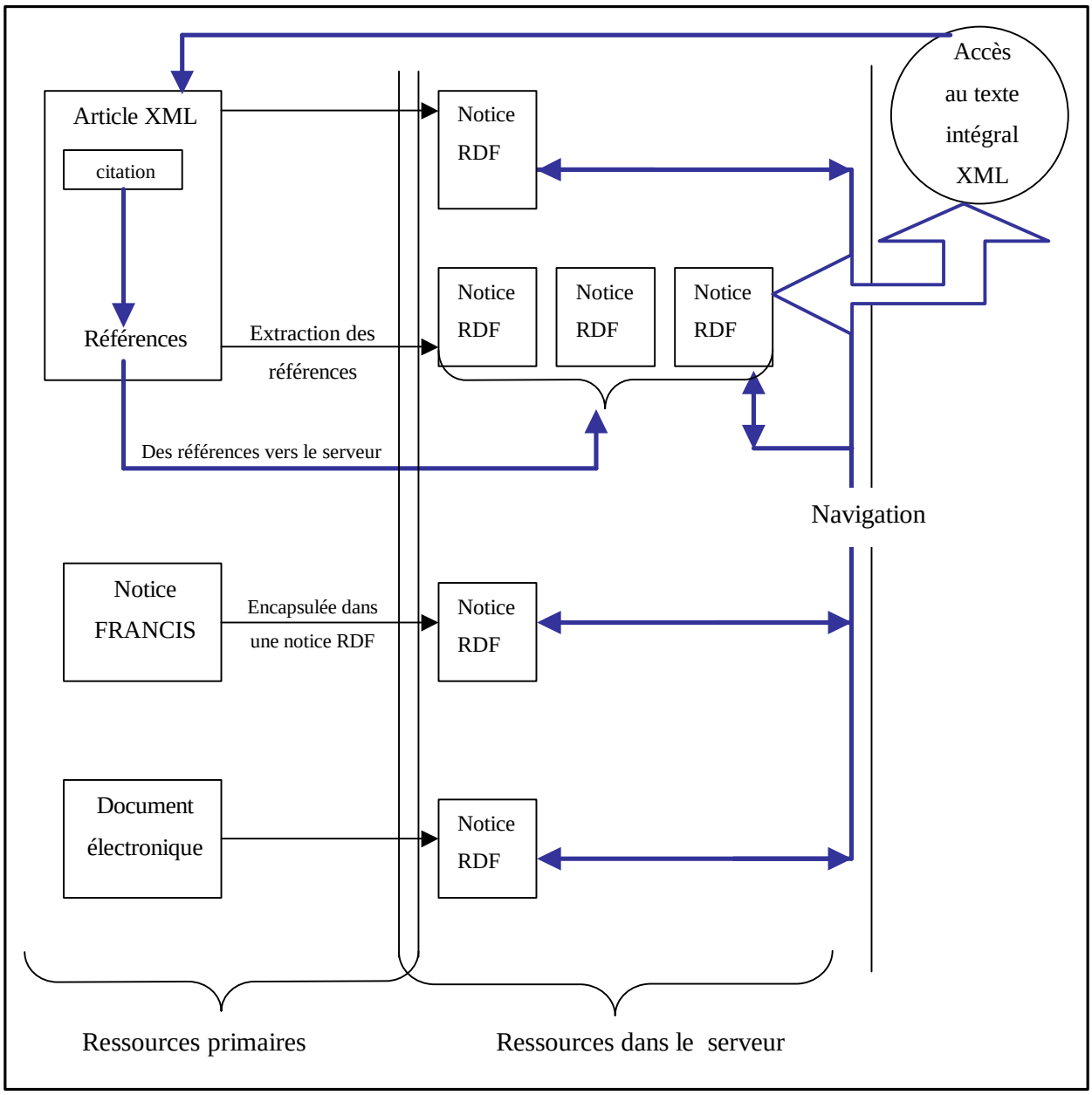


Figure 11 : Schéma de construction et de navigation du serveur

Ce serveur mono-base, dans lequel toutes les notices sont au même format RDF et traitées comme un seul corpus, repose sur une navigation entre les citations et les références. Il est ainsi possible d'accéder, à partir du serveur, à la revue (archivée en local), aux notices de références et vice-versa des notices vers la revue. La navigation forme alors une boucle.

Le serveur, même à titre de prototype, apporte une réelle valeur ajoutée à une revue électronique telle qu'ALSIC puisque nous avons :

- Une plate forme de revues électroniques en local dans laquelle chaque article est identifié, localisé, et archivé en local.
- Des articles en XML normalisés.
- Une navigation entre les différentes ressources et le texte intégral des articles qui apporte aux chercheurs une ouverture sur d'autres réflexions ou thématiques.

Nous finirons notre étude par une réflexion plus générale concernant la constitution de la nouvelle chaîne de traitement à mettre en place pour la publication d'une revue en XML.

D. Réflexion sur une chaîne de traitement

Toute cette réflexion autour de la normalisation des ressources nous a alors amené à nous pencher sur la chaîne de traitement que nécessite la création d'une revue en XML et sur le rôle qu'aura l'INIST dans cette chaîne. Si les documents HTML sont normalisés en XML, **comment concevoir les prochains articles par rapport aux consignes aux auteurs et comment se déroule le processus d'édition**? En fait, jusqu'à présent, les auteurs devaient fournir des documents word en suivant des consignes de présentation[16]. Ces documents étaient ensuite traités manuellement et transformés en HTML. Or, la revue devant désormais être au format XML, **comment traiter automatiquement les articles** ? Il s'agira alors de convenir, dans le cadre du partenariat entre ALSIC et l'INIST, de la manière dont vont être gérés et conçus les documents (archives et futurs) dans une chaîne en XML. Nous pourrons

[16] http://alsic.u-strasbg.fr/Infrevue/auteurs.htm

donc faire un bilan sur les différents formats et traitements envisagés et situer notre travail dans la chaîne éditoriale.

Du point de vue de l'éditeur, la revue en ligne doit être constituée selon un schéma précis : à partir de l'article tel que l'écrit l'auteur, en utilisant les styles word, il faut effectuer une conversion automatique en XML, selon la DTD retenue dans le cadre du projet, puis à l'aide d'une feuille de style XSL ou XSL-FO (pour la génération du document en PDF) l'afficher au format HTML ou PDF*. Les réflexions de l'INIST portent donc sur cette nouvelle chaîne éditoriale tout en respectant le schéma d'édition classique. La chaîne passe alors par plusieurs étapes :

- Création de l'article par l'auteur, en principe distant du comité de rédaction de la revue, avec un outil habituel (traitement de texte) en suivant les recommandations du comité de lecture. L'auteur observe alors les règles générales de typographie et de présentation indiquées dans les consignes aux auteurs. Le futur schéma éditorial implique une adaptation de ces consignes en conséquence. L'auteur fournit également les informations de base correspondant à la fiche de description de l'article (métadonnées).
- Correction de l'article et mise en forme au niveau du comité de rédaction (selon une feuille de style dépendante de la structure finale de l'article XML) et création des métadonnées.
- Transformation du document Word stylé en document XML structuré directement ou via XHTML.
- Diffusion de l'article : plusieurs processus de transformation sont ensuite à envisager et détermineront le type de feuille de style à utiliser. L'article correspondra alors à :
    - Soit un fichier XHTML généré directement à partir de word
    - Soit un fichier XHTML créé par transformation du documents XML + XSL
    - Soit un fichier PDF (format permet aux lecteurs d'imprimer aisément les articles avec une mise en page proche de celle de la

Toile, habitude courante aujourd'hui et qu'ALSIC pratique) par l'application de XSL-FO sur le document XML.

Toute cette phase de production a eu lieu au sein de la revue ALSIC conjointement aux réflexions de l'INIST. Ce type de processus décrit ci-après (voir figure 12) est dorénavant recommandé pour la rédaction des thèses universitaires, pour certains types de rapports et d'autres revues.

Enfin, les métadonnées sont intégrées sur le site et permettent à la revue de garder un bon référencement par les moteurs de recherche et autres robots qui vont, de plus en plus, tenir compte de ces formats. Dans le même esprit, la revue peut échanger ces métadonnées avec d'autres revues. L'INIST est donc dépositaire de ces métadonnées (*repository*). Il met également en place les services gérant les identifiants dont nous avons parlé sur le site www.inist.fr. Ses serveurs sont donc prêts à satisfaire les requêtes des formats correspondants provenant de services extérieurs. De nombreux services vont ainsi se développer qui ne manqueront pas de constituer des répertoires de données sur des articles.

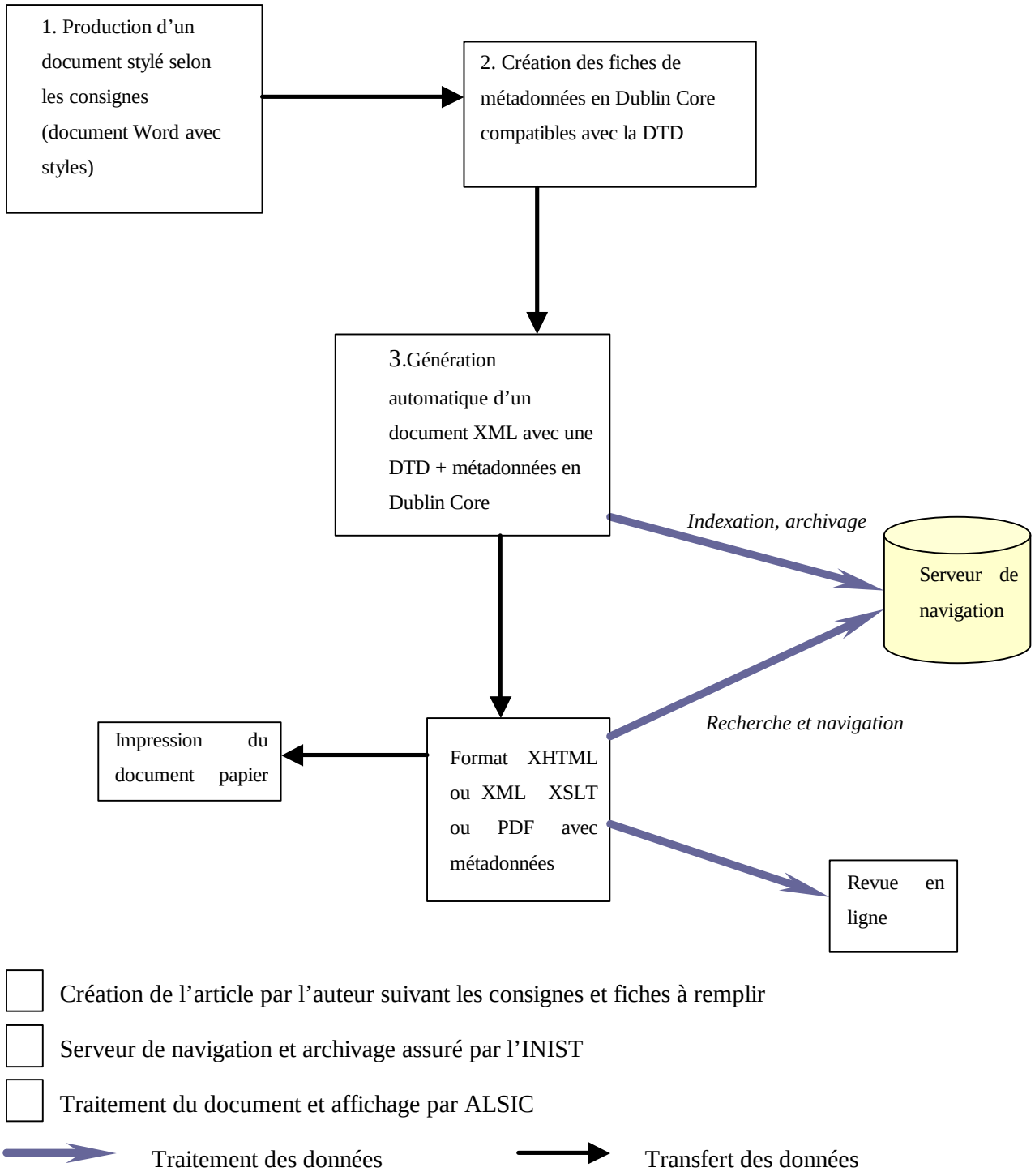


Figure 12 : La chaîne de traitement d'ALSIC

Nous proposons cette chaîne de traitement qui permet de replacer le rôle et la place de l'INIST par rapport à ALSIC et de situer le niveau d'intervention des différents partenaires.

CONCLUSION

Le projet REVELEC est un projet phare pour l'INIST puisqu'il l'engage vers un nouveau type de service de plus en plus important dans le monde de l'IST, **l'édition électronique dans le domaine SHS**. L'étude de faisabilité que nous avons menée a permis de dégager les différentes phases et enjeux de ce projet. Outre les réalisations en tant que telles, nous avons en effet pu mesurer l'ampleur et **la transversalité du dossier et les différents outils techniques et moyens humains qu'il nécessite**. Le prototype réalisé dégage **les différentes problématiques d'une plate-forme d'édition basée les technologies XML**. Aussi, nous élaborerons dès le mois de septembre le cahier des charges définitif du projet REVELEC avant de procéder à son industrialisation.

D'un point de vue personnel, ce stage m'a permis de m'investir dans un projet en équipe faisant intervenir différents acteurs internes et externes avec des compétences variées dont certaines sont en cours d'acquisition, d'expertise ou de déploiement par mes collègues de l'INIST. Toutes les phases de réalisation que j'ai menées et présentées dans ce rapport m'ont permis de m'enrichir de techniques et compétences essentielles pour un professionnel de l'Information Scientifique et Technique (le monde de l'édition, la maîtrise des langages et outils informatiques pointus qui nous dote d'une autonomie indispensable, la manipulation des métadonnées type Dublin Core, le XML et XSL qui sont les technologies et formats de demain sur Internet…).

Mon stage est prolongé jusqu'au mois d'octobre (sous forme de contrat). Ce prolongement me permettra de poursuivre et de généraliser le processus de XMLisation de la revue, de préparer l'industrialisation du prototype présenté et d'assister Sylvie Grésillaud dans ses activités de déploiement des compétences acquises auprès du personnel des départements de l'institut.

REFERENCES

Bibliographie

BEAUDRY, G. & BOISMENU, G, « Conception d'un portail de production, de diffusion et de gestion de publications électroniques, étude de faisabilité », http://www.erudit.org/erudit/etude/accueil.html

CHARTRON, G., (1999). « Revues électroniques: développement de l'offre et questions actuelles »: http://www.ccr.jussieu.fr/urfist/presse/offre/

DE BRITTO, M., (2001). « Présentation de la bibliothèque Scielo et de ses DTD », *Expertise de ressources pour l'édition de revues electroniques*, Chartron, G. & Salaün, J.M. (dirs). ENSSIB. En ligne : http://revues.enssib.fr/titre/8etudca/2scielo/7exemple.htm

LEGENTIL-GALAND, M. (2001). "Edition de revues scientifiques", In *Expertises de ressources pour l'édition de revues numériques*, Chartron, G. & Salaün, J.M. (dirs).

LUPOVICI, C. (1998). « Le Digital Object Identifier », *BBF*, T. 43 n°3, pp. 49-54.

MICHARD A. (2001). *XML, langage et applications*, Editions Eyrolles, 2001.

ROUSSEAU E. (2001). *Etude d'une chaîne de publication web de texte intégral au format XML*, Rapport de stage DESSID, ENSSIB.

WALSH, N., MUELLNER, L. (2001)., *Docbook, la référence*, traduction de l'Américain par S. Blondel et P. Zanettacci, éditions O'Reilly, Paris.

BEAUDRY, G. & BOISMENU, G. (2001). "Expertise technique et organisationnelle". In *Expertises de ressources pour l'édition de revues numériques*, Chartron, G. & Salaün, J.M. (dirs). ENSSIB. En ligne (http://revues.enssib.fr/Index/indextecnic.htm)

Sites et documents de référence

Recommandations du W3C, 2 May 2001 sur XML http://www.w3.org/XML/Schema

Recommandations du W3C sur Dublin Core et RDF
http://www.w3.org/TR/xmlschema-0/

DCrecom (2002). *Page de référencement des documents au stade de recommandation*. Dublin Core: http://dublincore.org/documents/

DClib (2002). *DC-Library Application Profile (DC-Lib)*. Proposition du 2002-04-16. Dublin Core. http://dublincore.org/documents/2002/04/16/library-application-profile/

DCschema (2002) *Recommendations for XML Schema for Qualified Dublin Core.* Proposition du 2002-07-14. Dublin Core.
http://www.ukoln.ac.uk/metadata/dcmi/xmlschema/

Webographie **(tous les sites ont été consultés en août 2002)**

?? **Les partenaires**

ALSIC (2002). Site de la *revue Apprentissage des Langues et Systèmes d'Information et de Communication*. http://alsic.org

INIST (2002). Site de *l'INstitut de l'Information Scientifique et Technique, CNRS*. http://www.inist.fr

LLTJ (2002). Site de la revue *Language Learning and Technology Journal*. http://llt.msu.edu

?? **Sur les revues électroniques**

Les enjeux de l'édition électronique :
http://www.bu.univ-ubs.fr/Cours_UBS/ELEC/Ed_electro2002.html

Revue SOLARIS, N°6, (1999-2000). « Normes et documents numériques: quels changements? », Sous la direction de Ghislaine Chartron et Jean-Max Noyer, http://www.info.unicaen.fr/bnum/jelec/Solaris/d06/

Etude du labo RECODOC (1999). « Les nouvelles pratiques de production et d'usage des revues scientifiques dans leur passage du papier à l'électronique », http://www.recodoc.univ-lyon1.fr/publications/CISI99/CISI99.htm

M. Sévigny, (1997). « Un modèle de traitement pour l'édition électronique de revues savantes », http://www.culture.fr/culture/mrt/numerisation/gde/edition_pum.html

?? **Sur XML**

Portail sur XML : http://www.chez.com/xml/

Concertation sur l'information bibliographique enrichie Groupe XML (1999), http://www.abf.asso.fr/enrichi/xml/991214.htm

S. FLEURY, Sur toutes les formats des documents électroniques et leurs composantes (DTD et XML),
http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/cours/parcours/slides/slsgml/

Exemple d'un passage du html au xml : http://www.chez.com/xml/htmlt/index.htm

Le site XML MUTU : http://www.mutu-xml.org/xml-base/shared/KEY-DTD.html

Site sur les langages du web : http://www.laltruiste.com/

DCXML (2002). *Guidelines for implementing Dublin Core in XML*.
http://dublincore.org/documents/2002/07/23/dc-xml-guidelines/

?? **Sur les DTD**

Références sur DTD et Document électronique :
http://www.info.unicaen.fr/bnum/jelec/Solaris/d06/6benoit.html

La DTD ISO 12083 pour les articles de périodiques :
http://www.xmlxperts.com/xmlarticledtd.htm

DocBook (1999). *DocBook the definitive guide*. Editions 0'Reilly. La version en ligne et en anglais : http://www.oreilly.com/catalog/docbook/chapter/book/docbook.html

DocBookDTD (2002). *Page d'accueil du comité DocBook de Oasis.* Accès à la version4.2 de juillet 2002 de la DTD DocBook. of 2002-06-14. Oasis-Open.org. http://www.oasis-open.org/committees/docbook/

?? **Sur les OAI et les identifiants**

OAImetadata (2002). *The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14*. Document Version 2002/07/05. *Open Archive Initiative*: http://www.openarchives.org/OAI/openarchivesprotocol.html

OpenURLbib (2001). *Bibliography on OpenURL Issues*. NISO Committee AX:
http://library.caltech.edu/openurl/Bibliography.htm

OPenURLsyntax (2000). *OpenURL syntax description*. version: OpenURL/1.0f - 2000-05-16. SFX et ExLibris : E-U.: http:/www.sfxit.com/openurl/openurl.html

GLOSSAIRE

ALSIC : ALSIC (Apprentissage des Langues et Système d'Information et de Communication, ALSIC.org) est une revue numérique qui regroupe une communauté de chercheurs et de praticiens à la fois auteurs et lecteurs sur le thème de l'apprentissage des langues. www.alsic.org

Articlesciences : Moteur de recherche et de commande en ligne de copies d'articles scientifiques et techniques de l'INIST, disponibles en 4 langues avec possibilité de paiement en ligne par carte bancaire.

Article@INIST : Catalogue en ligne du fonds INIST contenant 7 millions de références d'articles et monographies.

Base de données : « Ensemble de données organisées en vue de leur utilisation par des programmes correspondant à des applications distinctes et de manière à faciliter l'évolution indépendante des données et des programmes. » (JO 17 janvier 1982)

Base de données bibliographique : Ensemble de références bibliographiques regroupées et organisées en fichier informatique.

Bibliométrie : Procédés mathématiques et statistiques utilisés pour mesurer les modes d'utilisation et de publication du matériel documentaire.

Biblioscіences : Accès mutualisé à l'information scientifique et technique regroupant plusieurs bases de données (Pascal, FRANCIS, Biomed, Medline…) dans une interface Silver Platter. Réservé aux laboratoires CNRS.

CGI : *Common Gateway Interface*. Petit programme informatique, écrit en langage de 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.

ConnectSciences : Portail CNRS d'Information Scientifique et Technique en ligne qui met gratuitement à la disposition du public un ensemble de ressources documentaires et de services produits par l'INIST.

DILIB : *Documentation and 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.

DOI: *Digital Object Identifier*. Le DOI est un système d'identification unique d'un document permettant de localiser des objets de manière sûre en s'appuyant sur la mise en place d'un répertoire international d'identifiants de publications électroniques et sur lequel se penche le groupe Elsevier. Le prix d'un identifiant est de 1000 $.

DTD : *Document Type Definition.* Il s'agit d'un ensemble de règles qui établit un modèle de structure logique pour un document SGML ou XML et définit les éléments qui entrent dans la composition du document (vocabulaire et grammaire).

Dublin Core : Elaboré depuis 1995, le *Dublin Core* est 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.

Espace de nom : La définition d'un espace de nom dans un document XML balisé permet d'associer toutes les balises d'un langage à un groupe (être capable de dissocier les éléments de HTML contenus dans le document des balises XML, ou mieux : pouvoir mettre du HTML, MathML, et CML dans un même document).

Fichier d'associations : Fichier permettant la mise en relation de termes co-occurrents 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.

FRANCIS: Base de données de l'INIST spécialisée dans les Sciences Humaines et Sociales (SHS) avec près de 2,5 millions de références bibliographiques depuis 1972.

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 consistant à décrire et caractériser un document à l'aide d'une analyse des concepts contenus dans ce document.

Infométrie : Terme adopté par la FID (*International Federation of Documentation*) pour désigner l'ensemble des activités métriques relatives à l'information, couvrant la bibliométrie.

INIST : INstitut de l'Information Scientifique et Technique (au sein du CNRS) qui a pour mission de collecter, traiter et diffuser les résultats de la recherche scientifique et technique.

IST: Information Scientifique et Technique qui recouvre l'ensemble des connaissances produites dans les divers secteurs de la recherche.

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.

LORIA: Laboratoire LOrrain de Recherches en Informatique et ses Applications. Unité mixte du CNRS, de l'INRIA (Institut National de Recherche en Informatique et en Automatique),

de l'INPL (Institut National Polytechnique de Lorraine), et des universités Henri Poincaré, Nancy 1 et Nancy 2.

LLTJ : *Language and Learning Technology Journal*. Revue américaine très proche d'ALSIC au niveau thématique (http://llt.msu.edu/default.html).

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 des moteurs de recherche sur Internet.

Miriad : Module intranet de recherche dans les bases Pascal et FRANCIS de l'INIST. Permet la consultation de notices déchargeables dans plusieurs formats dont SGML.

OAI : *Open Archive Initiative*. Protocole permettant l'identification des documents sur Internet.

Pascal : Base de données bibliographiques multidisciplinaire et multilingue produite par l'INIST et signalant près de 14 millions de références en sciences, technologies et medecine depuis 1973.

PDF : *Portable Document Format*. Format de document électronique développé par Adobe. Ce format conserve l'allure originale de la forme imprimée du document (textes, graphiques, couleurs) peu importe la plate-forme utilisée.

RDF : *Resource Description Framework* (RDF) (Structure de description des ressources) est un modèle ou un formalisme indépendant des systèmes d'exploitation qui fournit un cadre conceptuel pour définir et utiliser les métadonnées. Les éléments en Dublin Core devront ainsi être compris comme des propriétés RDF qui leur donne un cadre ou une syntaxe particulière. La syntaxe sera exprimée en langage XML mais la sémantique sera définie selon les besoins des usagers.

REVELEC : REVues ELECtroniques. Projet issu d'un partenariat entre ALSIC et l'INIST consistant en la réalisation d'un prototype de plate forme de revues électroniques en Sciences Humaines et Sociales.

RTF : *Rich Text Format*. Ce format issu de *word*? a été créé par *Microsoft*? comme norme pour échanger les documents entre les différentes applications Windows.

Services@inist : Plate forme d'application de l'INIST proposant des services liés à la fourniture de documents et regroupant article@inist, connectsciences, francis…

SGML : *Standard Generalized Mark up 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. Il est moins contraignant que XML car ne nécessite pas la fermeture des balises mais reste moins propre que celui-ci.

TICE : Technologie de l'Information et de la Communication dans l'Enseignement.

URN : *Uniform Resource Names.* Fournit des identifiants uniques des ressources sur le réseau.

XML : *eXtensible Markup Language.* Évolution du langage SGML permettant de décrire n'importe quel type de données et de modéliser un document grâce à des balises qui en décrivent la structure et le contenu.

XMLSpy? **:** Outil « intégral » qui permet de créer, *parser*, valider un document XML ainsi que de générer une DTD associée.

XSL : *eXtensible Stylesheet Language.* Langage qui permet en effet de développer des feuilles de styles pour la représentation à l'écran des documents XML au format HTML. SXL-FO qui dérive de XSL permet quant à lui d'afficher des documents en PDF.

W3C : *World Wide Web Consortium.* Les industriels s'y accordent sur les standards du Web.

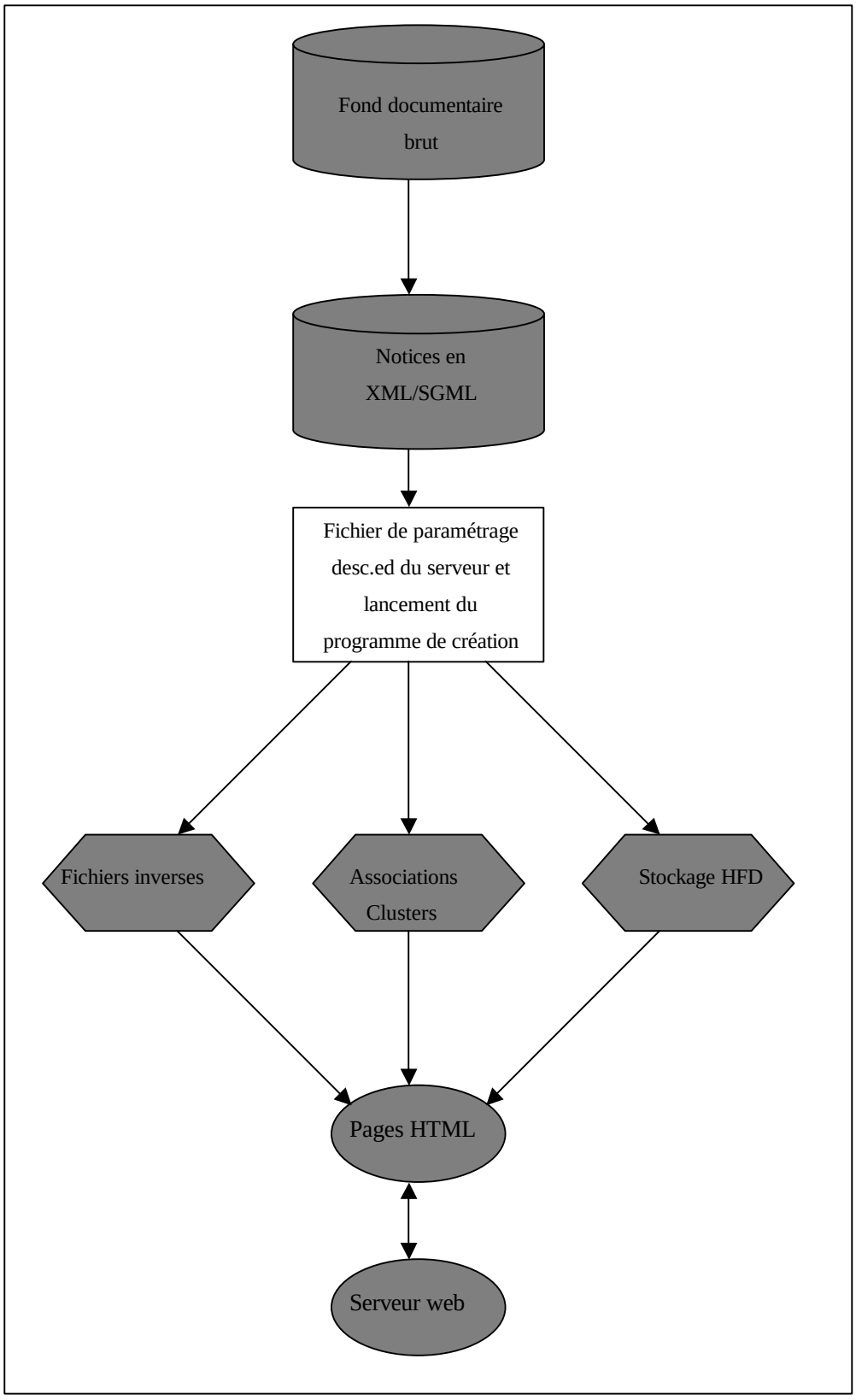
TABLE DES ILLUSTRATIONS

Figure 1 : Organigramme de l'INIST.....8
Figure 2 : Le modèle de navigation classique.....20
Figure 3 : Modèle de navigation avec texte intégral en XML.....21
Figure 4 : La structure des notices en RDF et Dublin Core 24
Figure 5 : le fichier desc.ed ou document paramètre.....26
Figure 6 : Les bases et index du serveur REVELEC. Exemple du multi.....27
Figure 7 : L'accès au texte intégral …… 27
Figure 8 : Page Web pour Connect Sciences.....30
Figure 9 : Passage de la revue existante en XML.....31
Figure 10 : Arborescence permettant de localiser et archiver les articles.....41
Figure 11 : Schéma de construction et de navigation du serveur 43
Figure 12 : La chaîne de traitement d'ALSIC 47

ANNEXES

ANNEXE 1 : ETAPES DE CREATION DU SERVEUR DILIB 58
ANNEXE 2 : CONVERSION DES ARTICLES DE HTML VERS XML 59
- Extrait de l'article en HTML.....59
- La DTD 62
- Extrait d'un des programmes LEX de conversion en XML 64
- Extrait de l'article XML généré.....66
- La feuille de style XSL 68
- Le document final.....72

ANNEXE 1 : Etapes de création du serveur DILIB



ANNEXE 2 : CONVERSION DES ARTICLES EN XML

Dans cette annexe, nous illustrons nos propos de la partie III concernant la conversion des articles d'ALSIC en XML. Ainsi, dans l'ordre logique du déroulement des opérations nous avons sélectionné un extrait de l'article au format HTML initial, un extrait de la DTD basée sur la Docbook à partir duquel le texte a été structuré, un extrait d'un programme LEX permettant d'automatiser la conversion, un extrait de l'article en XML généré, un extrait de la feuille de style XSL et enfin un extrait du document final HTML. Nous avons à chaque fois sélectionné la même partie de l'article afin de montrer les différentes étapes de la conversion.

✍Extrait de l'article en HTML

Nous avons sélectionné une partie du code source d'un article ALSIC balisé en HTML, et en particulier l'en-tête, le sommaire, la table des matières et le résumé, ainsi qu'un extrait de l'introduction.

```
<HTML><HEAD>
<LINK REL="stylesheet" TYPE="text/css"HREF="../../Style.css">
<META HTTP-EQUIV="Content-Type"CONTENT="text/html;charset=ISO-8859-1">
<META http-equiv="Content-Style-Type"content="text/css">
<TITLE>ALSIC : &eacute;criture didactique du multim&eacu  te;dia</TITLE>
   <STYLE type="text/css">
   <!--
   -->
   </STYLE>
</HEAD>

<BODY>
<A NAME="Haut">
<!-------------------------------- EN TETE ------------------------ >
<TABLE BORDER=0 WIDTH="95%">
<TR>
<TD VALIGN=TOP><A HREF="../sommaire.htm"><IMG
SRC="../../Images/loband12.gif"
ALT="Vers sommaire"BORDER=0 HEIGHT=72 WIDTH=240 ALIGN="left"></A>
</TD>
<TD ALIGN="right"VALIGN=TOP>
<SPAN class="ar80"><A HREF="http://alsic.u -strasbg.fr">http://alsic.u -
strasbg.fr</A><BR><SPAN
class="rar">Vol. 1, Num&eacute;ro  1, juin 1998</SPAN><BR><SPAN
class="var">pp 3  - 25<BR></SPAN><SPAN class="rb16">Recherche</SPAN>
</TR>
</TABLE>

<!-------------------------------- TITRE ------------------------ >
```

```
<BR><BR><BR><BR>
<BLOCKQUOTE>
<H2 Align="center">R&eacute;flexions linguist  iques et s&eacute;miologiques
pour une
&eacute;criture didactique du multim&eacute;dia de langues</H2>
</BLOCKQUOTE>


<!------------------------------   Premier para  ------------------------- >
<table border="0" WIDTH="95%">
<TR>
<TD
WIDTH="120"> &nbs p;         &nb
sp;</TD>
<TD><BR><A HREF="#auteur">Anne -Laure FOUCHER</A><BR>Universit&eacute; Paris
6
                                , France
                    <BR><BR>
</TD>
</TR>

<TR>
<TD WIDTH="120"VALIGN="top">
   & nbsp;         </T
D>
<TD VALIGN=TOP WIDTH="*">
<P><SPAN class="var"><B>R&eacute;sum&eacute; : </B></SPAN><I>Dans une
combinaison originale
de domaines souvent envisag&eacute;s de
mani&egrave;re ferm&eacute;e que s ont la didactique des langues, la
linguistique, la s&eacute;miologie et l'ALAO (Apprentissage des Langues
Assist&eacute; par Ordinateur), nous mettons en avant la
n&eacute;cessit&eacute; de concevoir l'image multim&eacute;dia comme un
objet
composite, lieu  v&eacute;ritable d'interactions entre modes iconique,
auditif
et linguistique. De cette option th&eacute;orique d&eacute;coule, selon
nous,
une philosophie qui pr&eacute;side &agrave; la conception d'un
multim&eacute;dia pour l'apprentissage des langues.   Nous en donnons les
grandes
lignes sous forme de jalons m&eacute;thodologiques destin&eacute;s &agrave;
une
&eacute;criture du multim&eacute;dia en langues.<I><p>
</TD></TR>

<TR>
<TD WIDTH="120"VALIGN="top">
          ;    </T
D>
<TD VALIGN=TOP WIDTH="*"><BR>
<UL>
<LI><A HREF="#Heading1">1. Introduction</A>
<LI><A HREF="#Heading2">2.  Enseignement/apprentissage des Langues et
multim&eacute;dia</A>
<LI><A HREF="#Heading3">3.  "L'image multim&eacute;dia"  est un lieu
d'interactions multimodales</A>
<LI><A HREF="#Heading9">4.  Implications pour une philosophie de conception
du
```

```
          multim&eacute;dia de langues</A>
<LI><A HREF="#Heading12">5.  Remarques m&eacute;thodologiques pour une
&eacute;criture
    du multim&eacute;dia</A>
<LI><A HREF="#Heading15">6.  Pr&eacute;alables m&eacute;thodologiques
relatifs &agrave; l'introduction de l'image dans l'image
                  multim&eacute;dia</A>
<LI><A HREF="#Heading22">7.  Quelques jalons m&eacute;thodologique s pour un
cadre d'&eacute;criture didactique
   du multim&eacute;dia de langues</A>
<LI><A HREF="#Heading37">8. Conclusion</A>
<LI><A HREF="#Heading38">R&eacute;f&eacute;rences</A>
</UL>
<BR><BR>
</TD
</TR>
</TABLE>

<!-----------------------------   texte  ------------------------ >
<BLOCKQUOTE>
<h2>
<A NAME="Heading1">1.  Introduction</A></h2>
<p>
<IMG SRC="../../Images/alpha_l.gif"border=O>ors d'une table ronde en 1996
autour du th&egrave;me "Environnements
interactifs pour l'apprentissage des langues", &  eacute;tait
soulign&eacute;e
l'importance, pour qui d&eacute;sirait exploiter p&eacute;dagogiquement des
documents hyperm&eacute;dias, d'un passage par les trois &eacute;tapes
suivantes : la premi&egrave;re &eacute;tant la recherche de produits
susceptibles de plaire aux apprenants ; la seconde consistant en un
rep&eacute;rage dans les produits des parties "<i>dans lesquelles le
rapport
texte/son/image [favoriserait] une compr&eacute;hension ais&eacute;e et
l'envie
de s'exprimer ; ce rep&eacute;rage [s upposant] un minimum de connaissances
s&eacute;miologiques</i>."  (Chanier &amp; al, 1996 : p. 252) ; la
derni&egrave;re &eacute;tant consacr&eacute;e &agrave; la conception de
"<i>t&acirc;ches amenant l'apprenant &agrave; manipuler (en
compr&eacute;he nsion
comme en production) des donn&eacute;es sonores, visuelles et
graphiques</i>"
(Idem : p. 252).  Parall&egrave;lement &agrave; ces
consid&eacute;rations et
plus r&eacute;cemment, T. Lancien sugg&eacute;rait qu'on optimise la
multicanalit&eacute;<a  name="fnB0"href="#fn0">[1]</a> dans le
multim&eacute;dia<a name="fnB1"href="#fn1">[2]</a>, multicanalit&eacute;
qu'il
d&eacute;finit comme "<i>le fait que coexistent sur un m&ecirc;me support
diff&eacute;rents canaux de communication</i>" et qui "<i>ne pr  end de sens
que selon les choix que fait la personne qui les consulte</i>" (Lancien,
1998
: p. 24), cela afin de favoriser l'apprentissage de la langue. Il ajoutait
que
la multicanalit&eacute; d'un multim&eacute;dia devait &ecirc;tre
&eacute;valu&eacute;e  en consid&eacute;rant la nature, la pertinence et la
richesse des liens entre les m&eacute;dias.

</P>             ▪▪▪
```

La DTD

La DTD est beaucoup plus longue (compte-tenu de la déclaration des entités) mais nous n'en proposons qu'un extrait. Nous n'avons pas mis la liste de toutes les entités.

```
<?xml version="1.0" encoding="ISO -8859-1"?>
<!-- edited with XML Spy v4.4 (http://www.xmlspy.com) by GRESILLAUD (CNRS)
-->
<!--DTD generated by XML Spy v4.4 (http://www.xmlspy.com)  -->
<!ELEMENT abstract (para)>
<!ELEMENT affilia tion (orgname?, address?)>
<!ELEMENT orgname (#PCDATA)>
<!ELEMENT address (#PCDATA)>
<!ELEMENT anchor EMPTY>
<!ATTLIST anchor
      id (auteur | fn0 | fn1 | fn2 | fn3 | fn4 | fn5 | fn6 | fn7 | fn8 |
fnB0 | fnB1 | fnB2 | fnB3 | fnB4 | fnB5 | fnB6 | fnB7 | fnB8)   #REQUIRED
>
<!ELEMENT article (articleinfo, section+, bibliography, note, author)>
<!ELEMENT articleinfo (volumenum, issuenum, pubdate, artpagenums, rubrique,
title, author, abstract, toc)>
<!ELEMENT artpagenums (#PCDATA)>
<!ELEMENT attribution (#PCDATA)>
<!ELEMENT author (#PCDATA | name | affiliation | anchor | authorblurb |
email | title | subtitle | ulink)*>
<!ELEMENT authorblurb (#PCDATA)>
<!ELEMENT biblioentry (#PCDATA | author | copyright | biblioset | pagenums
| seriesinfo | bibliomisc | title)*>
<!ELEMENT bibliography (section+)>
<!ELEMENT bibliomisc (#PCDATA)>
<!ELEMENT biblioset (#PCDATA | title)*>
<!ATTLIST biblioset
      relation (article | revue) #REQUIRED
>
<!ELEMENT blockquote (#PCDATA | attribution | informalexample)*>
<!ELEMENT citation (blockqu ote |attribution)*>
<!ELEMENT citetitle (#PCDATA)>
<!ELEMENT copyright (year)>
<!ELEMENT email (#PCDATA)>
<!ELEMENT informalexample (#PCDATA | citetitle | itemizedlist)*>
<!ELEMENT issuenum (#PCDATA)>
<!ELEMENT itemizedlist (listitem+)>
<!ELEMENT link (#PC DATA)>
<!ATTLIST link
      linkend (auteur | fn0 | fn1 | fn2 | fn3 | fn4 | fn5 | fn6 | fn7 | fn8
| fnB0 | fnB1 | fnB2 | fnB3 | fnB4 | fnB5 | fnB6 | fnB7 | fnB8) #REQUIRED
>
<!ELEMENT listitem (#PCDATA)>
<!ELEMENT name (#PCDATA | link)*>
<!ELEMENT note (#PCDATA  | title | noteentry)*>
<!ELEMENT noteentry (anchor | link | para | citation | quote)*>
<!ELEMENT pagenums (#PCDATA)>
<!ELEMENT para (#PCDATA | anchor | link | para | quote | citation)*>
<!ELEMENT pubdate (#PCDATA)>
```

```
<!ELEMENT quote (#PCDATA)>
<!ELEMENT rub rique (#PCDATA)>
<!ELEMENT section (#PCDATA | title | para | citation | anchor | link |
blockquote | itemizedlist | biblioentry)*>
<!ATTLIST section
      id (S1 | S10 | S11 | S12 | S13 | S14 | S15 | S16 | S17 | S18 | S19 |
S2 | S20 | S21 | S22 | S23 | S24 | S2 5 | S26 | S27 | S28 | S29 | S3 | S30 |
S31 | S32 | S33 | S34 | S35 | S36 | S37 | S38 | S39 | S4 | S5 | S6 | S7 |
S8 | S9) #REQUIRED
>
<!ELEMENT seriesinfo (#PCDATA)>
<!ELEMENT subtitle (ulink?, email?, affiliation?)>
<!ATTLIST subtitle
      type CDATA #REQUIRE D
>
<!ELEMENT title (#PCDATA)>
<!ATTLIST title niveau (niveau1 | niveau2 | niveau3) #IMPLIED >
<!ELEMENT toc (tocentry+)>
<!ELEMENT tocentry (#PCDATA | xref)*>
<!ELEMENT ulink EMPTY>
<!ATTLIST ulink
      url CDATA #REQUIRED
>
<!ELEMENT volumenum (#PCDATA)>
<!ELEMENT xref EMPTY>
<!ATTLIST xref
      linkend (S1 | S12 | S15 | S2 | S22 | S3 | S37 | S38 | S9) #REQUIRED
>
<!ELEMENT year (#PCDATA)>
<!ENTITY aacute "&#x00E1;">
<!ENTITY Aacute "&#x00C1;">
<!ENTITY acirc "&#x00E2;">

<!ENTITY Acirc "&#x00C2;">
<!ENTITY agrav e "&#x00E0;">
<!ENTITY Agrave "&#x00C0;">
<!ENTITY aring "&#x00E5;">
<!ENTITY Aring "&#x00C5;">
<!ENTITY atilde "&#x00E3;">
<!ENTITY Atilde "&#x00C3;">
<!ENTITY auml "&#x00E4;">
<!ENTITY Auml "&#x00C4;">
<!ENTITY aelig "&#x00E6;">
<!ENTITY AElig "&#x00C6;" >
<!ENTITY ccedil "&#x00E7;">
<!ENTITY Ccedil "&#x00C7;">
<!ENTITY eth "&#x00F0;">
<!ENTITY ETH "&#x00D0;">
<!ENTITY eacute "&#x00E9;">
<!ENTITY Eacute "&#x00C9;">
```

…

Extrait d'un des programmes LEX de conversion de HTML vers XML

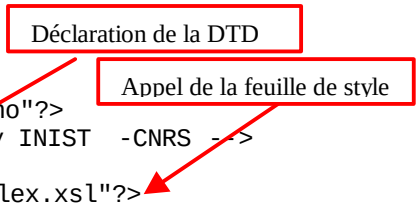
```
%START HE
%START TE TE
%START AUTEUR
%START ORGANISME
%START ORG
%START RESUM
%START RES
%START SOMMAIRE
%START SOMM
%START TEXTE
%START TITRE
%START CITATION
%START LIGNE
%START EXEMPLE
%START VERSNOTE
%%
"<HEAD>"  {printf("<!DOCTYPE article PUBLIC  -//OASIS//DTD DocBook XML
V4.1.2//EN\\http://www.oasis -
open.org/docbook/xml/4.1.2/docbookx.dtd><article>"); BEGIN HE;}
<HE>.  ;
<HE>"</HEAD>" {; BEGIN 0;}
"<A NAME=\"Haut\">"  {printf("<anchor id= \"HAUT\">"); BEGIN TETE;}
<TETE>.  ;
<TETE>"ol. "[0-9]  {printf("<volumenum>"); printf ("Vol.");
printf(yytext+4); printf("</volumenum>");}
<TETE>"ro "[0-9]*  {printf("<issuenum>"); printf("Num."); printf(yytext+3);
printf("</issuenum>");}
<TETE>", "[a-z]*" "[0-9][0-9][0-9][0-9]  {printf("<pubdate>");
printf(yytext+1); printf("</pubdate>");}
<TETE>"pp "[0-9]*" - "[0-9]*  {printf("<artpagenums>"); printf("pp.  ");
printf(yytext+3); printf("</artpagenums>");}
<TETE>"rb16\">"[A-Z]*[a-z]*[ &]*  {printf("<rubrique>"); printf(yytext+6);
printf("</rubrique>");}
<TETE>"center\">"  {; printf("<title>") ; BEGIN 0;}
"</H2>"[ ]*"</BLOCKQUOTE>"  {; printf("</title>"); BEGIN AUTEUR;}
<AUTEUR>.  ;
<AUTEUR>"auteur\">"  {printf("<author>"); printf("<para><name>");
printf("<link linkend= \"auteur\">"); BEGIN 0; BEGIN ORGANISME;}
<ORGANISME>"</A>"  {;
printf("</lin k></name></para><affiliation><orgname>"); BEGIN 0; BEGIN ORG;}
<ORG>.  ;
<ORG>"<I>"  {printf("</orgname></affiliation></author><abstract><para>");
BEGIN 0; BEGIN RESUM;}
<RESUM>.  ECHO;
<RESUM>[</I ]*"<"[/]*"p>"  {printf("</para></abstract>"); BEGIN 0; BEG  IN
RES;}
<RES>.  ;
<RES>"<UL>"  {printf("<toc>"); BEGIN 0; BEGIN SOMMAIRE;}
<SOMMAIRE>"<LI>"[ ]*"<A HREF= \"#Heading"  printf("<tocentry><xref
linkend=\"S");
<SOMMAIRE>"<LI>"[ ]*"<A HREF= \"#"  printf("<tocentry><xref linkend= \"");
<SOMMAIRE>"</A>"  printf(" </tocentry>");
```

```
<SOMMAIRE>"</UL>"  {printf("</toc></articleinfo>"); BEGIN 0; BEGIN SOMM;}
<SOMM>.  ;
<SOMM>"<BLOCKQUOTE>"  {BEGIN 0; BEGIN TEXTE;}
<TEXTE>"<"[hH][0-9]*">"[ ]*"<A NAME=\"Heading"  {printf("<section id= \"S");
BEGIN 0; BEGIN TITRE;}
<TEXTE>"<A NAME=\"Heading"  {printf("<section id= \"S"); BEGIN 0; BEGIN
TITRE;}
<TITRE>[1-9]"."  {printf("<title>"); printf(yytext); BEGIN 0, BEGIN TEXTE;}
<TITRE>">"[^0-9]  {printf("><title>"); printf(yytext+1); BEGIN 0, BEGIN
TEXTE;}
<TITRE>"</"[Hh][0-9]">"  {printf ("</title>"); BEGIN 0; BEGIN TEXTE;}
<TEXTE>"<A href"  {printf("<link linkend"); BEGIN 0; BEGIN VERSNOTE;}
<TEXTE>"</A>"[ ]*"</"[hH][0-9]">"  printf("</title>");
<TEXTE>"</"[Pp]">"  printf("</para>");
<TEXTE>"<"[Pp]">"  printf("<para>");
<TEXTE>"</"[Hh][0-9]">"  printf("</title>");
<TEXTE>"<"[hH][0-9]*">"  printf("<title>");
<TEXTE>"<IMG SRC"  printf("<imageobject><imagedata fileref");
<TEXTE>"\"<i>" {printf("<citation><blockquote><para> \""); BEGIN 0; BEGIN
CITATION;}
<TEXTE>"<i>\"" {printf("<citation><bloc kquote><para>\""); BEGIN 0; BEGIN
CITATION;}
<TEXTE>"<i>"  printf("<quote>");
<TEXTE>"</i>"  printf("</quote>");
<TEXTE>"<BLOCKQUOTE><i>" {printf("<citation><blockquote><para>  \""); BEGIN
0; BEGIN CITATION;}
<TEXTE>"<BLOCKQUOTE>"  ;
<CITATION>"</i>"  printf ("</para>");
<CITATION>"("[A-Z]  {printf("<attribution>("); printf(yytext+1);}
<CITATION>"</BLOCKQUOTE>"  printf(" ");
<CITATION>[0-9]")" {printf(yytext);
printf("</attribution></blockquote></citation>"); BEGIN 0; BEGIN TEXTE;}
<CITATION>"\"<i>"
printf("</blockquote></citation><citation><blockquote><para>  \"");
<TEXTE>"</BLOCKQUOTE>"  printf(" ");
<TEXTE>"<BLOCKQUOTE>"[ ]*"<FONT SIZE= -1>"
{printf("<blockquote><informalexample>"); BEGIN 0; BEGIN EXEMPLE;}
<TEXTE>"<!"[-*]*"ligne"[-*!]*">"  {printf("<!--***ligne***--!>"); BEGIN
0; BEGIN LIGNE;}
<TEXTE>"<"[UuLL]*">"[ ]*"<"[LliI]*">"  printf("<itemizedlist><listitem>");
<TEXTE>"<"[UuLL]*">"  printf("<itemizedlist><listitem>");
<TEXTE>"<"[lLiI]*">"  printf("</listitem><listitem>");
<TEXTE>"</"[uUlL]*">"  printf ("</listitem></itemizedlist>");
<EXEMPLE>"<i>"  printf("<citetitle>");
<EXEMPLE>"</i>"  printf("</citetitle>");
<EXEMPLE>"<"[uUlL]*">"[ ]*"<"[lLiI]*">"
printf("<itemizedlist><listitem>");
<EXEMPLE>"<"[lLiI]*">"  printf("</listitem><listitem");
<EXEMPLE>"</"[UuLl]*">"  printf("</listitem></itemizedlist>");
<EXEMPLE>"</FONT>"[.]*"</BLOCKQUOTE>"
{printf("</informalexample></blockquote>"); BEGIN 0; BEGIN TEXTE;}
<LIGNE>.  ECHO;
<LIGNE>"<A HREF=\"#"  printf("<link linkend= \"");
<LIGNE>"</A>"  printf("</link>") ;
<LIGNE>"<IMG SRC"  printf("<imageobject><imagedata fileref");
<LIGNE>"</"[pP]*">"  {printf("</para>"); BEGIN 0; BEGIN TEXTE;}
<LIGNE>"<"[pP]*" class=ligne>"   printf("<para>");
<TEXTE>"<"[Aa]*" name" {printf("<anchor id"); BEGIN 0; BEGIN VERSNOTE;}
<VERSNOTE>"#"  ;
```

```
<VERSNOTE>"href= \"#"  printf("><link linkend= \"");
<VERSNOTE>.  ECHO;
<VERSNOTE>"</"[Aa]*">" {printf("</link>"); BEGIN 0; BEGIN TEXTE;}
. ECHO;
%%
main()
{
 yylex();
 exit (0);
}
```

Extrait de l'article XML généré

Par rapport au texte HTML de départ, nous ne conservons aucun élément de présentation, puisque nous ferons une feuille de style pour cela. Nous ne conservons ni tableau ni pagination dans le document XML. Nous voyons que le document est complètement structuré et notamment les citations et références ce qui nous permettra de faire des traitements approfondis sur ces éléments.




```
<?xml version="1.0" encoding="ISO -8859-1" standalone="no"?>
<!-- edited with XML Spy v4.4 (http://www.xmlspy.com) by INIST  -CNRS -->
<!DOCTYPE article SYSTEM "F: \Revelec\dtdalex.dtd">
<?xml-stylesheet type="text/xsl" href="F: \Revelec\xslalex.xsl"?>
<article>
      <articleinfo>
            <volumenum>Vol.1</volumenum>
            <issuenum>Num.1</issuenum>
            <pubdate> juin 1998</pubdate>
            <artpagenums>pp. 3  - 25</artpagenums>
            <rubrique>Recherche</rubrique>
            <title>R&eacute;flexions linguistiques et s&eacute;miologiques
pour une &eacute;criture didactique du multim&eacute;dia de langues</title>
            <author>
                  <name>
                        <link linkend="auteur">Anne -Laure FOUCHER</link>
                  </name>
                  <affiliation>
                        <orgname>Université Paris 6</orgname>
                        <address>France</address>
                  </affiliation>
            </author>
            <abstract>
                  <para>Dans une combinaison originale de domaines souvent
envisag&eacute;s de mani&egrave;re ferm&eacute;e que sont la didactique   des
langues, la linguistique, la s&eacute;miologie et l'ALAO (Apprentissage des
LanguesAssist&eacute; par Ordinateur), nous mettons en avant la
n&eacute;cessit&eacute; de concevoir l'image multim&eacute;dia comme un
objet composite, lieu v&eacute;ritable d 'interactions entre modes iconique,
auditif et linguistique. De cette option th&eacute;orique d&eacute;coule,
selon nous, une philosophie qui pr&eacute;side &agrave; la conception d'un
```

```
multim&eacute;dia pour l'apprentissage des langues. Nous en donnons les
grandes lignes sous forme de jalons m&eacute;thodologiques destin&eacute;s
&agrave; une&eacute;criture du multim&eacute;dia en langues.</para>
            </abstract>
            <toc>
                  <tocentry>
                        <xref linkend="S1"/>1.  Introduction</tocentry>
                  <tocentry>
                        <xref linkend="S2"/>2.  Enseignement/apprentissage
des Langues et multim&eacute;dia</tocentry>
                  <tocentry>
                        <xref linkend="S3"/>3.  "L'image multim&eacute;dia"
est un lieu d'interactions multimodales</tocentry>
                  <tocentry>
                        <xref linkend="S9"/>4.  Implications pour une
philosophie de conception du multim&eacute;dia de langues</tocentry>
                  <tocentry>
                        <xref linkend="S12"/>5.  Remarques
m&eacute;thodologiques pour une &eacute;criture    du
multim&eacute;dia</tocentry>
                  <tocentry>
                        <xref linkend="S15 "/>6. Pr&eacute;alables
m&eacute;thodologiques relatifs &agrave; l'introduction de l'image dans
l'image multim&eacute;dia</tocentry>
                  <tocentry>
                        <xref linkend="S22"/>7.  Quelques jalons
m&eacute;thodologiques pour un cadre d'&eacute;criture didactique     du
multim&eacute;dia de langues</tocentry>
                  <tocentry>
                        <xref linkend="S37"/>8.  Conclusion</tocentry>
                  <tocentry>
                        <xref
linkend="S38"/>R&eacute;f&eacute;rences</tocentry>
            </toc>
      </articleinfo>
      <section id="S1">
            <title niveau="niveau1" >1. Introduction</title>
            <para>Lors d'une table ronde en 1996 autour du th&egrave;me
"Environnements interactifs pour l'apprentissage des langues", &eacute;tait
soulign&eacute;e l'importance, pour qui d&eacute;sirait exploiter
p&eacute;dagogiquement des  documents hyperm&eacute;dias, d'un passage par
les trois &eacute;tapes suivantes: la premi&egrave;re &eacute;tant la
recherche de produits susceptibles de plaire aux apprenants ; la seconde
consistant en un rep&eacute;rage dans les produits des parties
            <citation>
                        <blockquote>"dans lesquelles le rapport
texte/son/image [favoriserait] une compr&eacute;hension ais&eacute;e et
l'envie de s'exprimer ; ce rep&eacute;rage [supposant] un minimum de
connaissances s&eacute;miologiques." </blockquote>
                        <attribution>(Chanier &amp; al, 1996 p.
252)</attribution>
                  </citation>
                   ; la derni&egrave;re &eacute;tant consacr&eacute;e
&agrave; la conception de
            <citation>
```

Xref linkend permet de créer un lien croisé dans le document

Début de section et lien avec le sommaire

Marquage de la citation

...

Un exemple d'entrée dans la bibliographie :

…

```
            <!--***biblio*** -->
      </section>
      <bibliography>
            <section id="S38">
                  <title niveau="niveau1">R&eacute;f&eacute;rences</title>
                  <title niveau="niveau2"> R&eacute;f&eacute;rences
bibliographiques</title>
                  <biblioentry>
                        <author>Barthes R. </author>
                        <copyright>
                              <year>(1964)</year>
                        </copyright>. <biblioset relation="article">
                              <title>"Rh&eacute;torique de l'image</title>
                        </biblioset>
                        <biblioset relation="revue">
Communications</biblioset>
                        <seriesinfo> 4,"Recherches s&eacute;miologiques,
Seuil,</seriesin fo>
                        <pagenums>pp. 40 -51.</pagenums>
                  </biblioentry>
```

…

La feuille de style XSL

Nous reproduisons la feuille de style entière qui a été générée avec XMLSpy? et qui permet d'afficher le document au format HTML. La technologie XSL et notamment l'utilisation des Xpath s'est avéré très importante pour tous les liens à créer dans le document (notes, auteur…)

```
<?xml version="1.0" encoding="iso-8859-1"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:fo="http://www.w3.org/1999/XSL/Format">
    <xsl:output method="html" encoding="ISO-8859-1"/>
    <xsl:template match="/">
        <html>
        <link rel="stylesheet" type="text/css" href="http://alsic.u-strasbg.fr/Style.css"/>
                    <head>
                <TITLE>Alsic : éciture diacritique du multimédia</TITLE>
            </head>
            <body>
                <xsl:apply-templates/>
            </body>
        </html>
    </xsl:template>
    <!--création de l'en-tête -->
    <xsl:template match="articleinfo">
        <table border="0" width="95%">
            <tbody>
```

```
                <tr>
                    <td VALIGN="TOP">
                        <a HREF="http://alsic.u-strasbg.fr/sommaire.htm">
                            <img SRC="http://alsic.u-strasbg.fr/Images/loband12.gif" ALT="Vers sommaire" border="0"
HEIGHT="72" WIDTH="240" ALIGN="left"/>
                        </a>
                    </td>
                    <td ALIGN="right" VALIGN="TOP">
                        <span class="ar80">
                            <A HREF="http://alsic.u-strasbg.fr">http://alsic.u-strasbg.fr</A>
                        </span>
                        <br/>
                        <span class="rar">
                            <xsl:apply-templates select="volumenum"/>,
    <xsl:apply-templates select="issuenum"/>,
    <xsl:apply-templates select="pubdate"/>
                        </span>
                        <br/>
                        <span class="var">
                            <xsl:apply-templates select="artpagenums"/>
                        </span>
                        <br/>
                        <span class="rb16">
                            <xsl:apply-templates select="rubrique"/>
                        </span>
                    </td>
                </tr>
            </tbody>
        </table>
        <!-- Création du titre -->
        <h2 align="center">
            <xsl:apply-templates select="title"/>
        </h2>
        <br/>
        <!-- Positionnement des auteurs-->
        <table border="0" width="95%">
            <tbody>
                <tr>
                    <td width="120">&amp;#160;</td>
                    <td valign="top">
                        <xsl:apply-templates select="author"/>
                    </td>
                </tr>
            </tbody>
        </table>
        <br/>
        <!-- positionnement du résumé et du sommaire-->
        <table border="0" width="95%">
            <tbody>
                <tr>
                    <td width="120">&amp;#160;</td>
                    <td valign="top" width="*">
                        <p>
                        <span class="var"><b> Résumé : </b></span>
                        <i><xsl:apply-templates select="abstract"/></i>
                        </p>
                    </td>
                </tr>
                <tr>
                    <td width="120">&amp;#160;</td>
                    <td valign="top">
                        <xsl:apply-templates select="toc"/>
                    </td>
                </tr>
            </tbody>
        </table>
    </xsl:template>
```

```
	<!--Mise en forme des auteurs-->
	<xsl:template match="article/articleinfo/author">
		<xsl:apply-templates select="name"/>
		<br/>
		<xsl:apply-templates select="affiliation"/>
	</xsl:template>

	<!--création du lien "a propos de l'auteur"-->
	<xsl:template match="article/articleinfo/author/name">
		<xsl:for-each select="link[@linkend='auteur']">
	<a href="#{@linkend}"><xsl:apply-templates select="node()"/></a>
	</xsl:for-each>
	</xsl:template>

	<xsl:template match="article/author">
			<a name="{anchor/@id}"><h2><xsl:apply-templates select="title"/></h2></a>
			<b><xsl:apply-templates select=".//name"/></b>
			 : 
			<xsl:apply-templates select="authorblurb"/>
			<br/>
			<xsl:for-each select="subtitle"><p><b><xsl:apply-templates select="@type"/></b>  <xsl:apply-
templates select="node()"/></p>
</xsl:for-each>
	</xsl:template>

	<!--lien vers mel de l'auteur-->
	<xsl:template match="article/author/subtitle/email">
	<a href="{../ulink/@url}">
	<xsl:apply-templates select="node()"/>
	</a>
	</xsl:template>

	<!--complément pour l'auteur-->
	<xsl:template match="affiliation">
		<xsl:apply-templates select="orgname"/>,   <xsl:apply-templates select="address"/>
	</xsl:template>

<!-- Création du sommaire de l'article-->

<xsl:template match="toc">
<br/>
<p>
<xsl:for-each select="tocentry">
<li>
<xsl:variable name="linkend" select=".//xref/@linkend"/>
<a href="#{$linkend}">
<xsl:apply-templates select="node()" mode="crossref"/>
</a>
</li>
</xsl:for-each>
</p>
</xsl:template>


<xsl:template match="section">
<p>
<a name="{@id}"><xsl:apply-templates select="title"/></a>
</p>
<xsl:for-each select="para"><p><xsl:apply-templates select="node()"/>
</p></xsl:for-each>

<!--Mise en forme des références bibliographiques-->
<font size="-1"><xsl:for-each select="biblioentry"><p><xsl:apply-templates select="node()"/></p></xsl:for-each></font>
</xsl:template>

<xsl:template match="section/title[@niveau='niveau1']">
<h2>
<xsl:apply-templates select="node()"/>
```

```
</h2>
</xsl:template>


<xsl:template match="section/title[@niveau='niveau2']">
<h3>
<xsl:apply-templates select="node()"/>
</h3>
</xsl:template>

<xsl:template match="section/title[@niveau='niveau3']">
<h4>
<xsl:apply-templates select="node()"/>
</h4>
</xsl:template>



<!-- mise en forme des citations-->
<xsl:template match="para/citation">
<i><xsl:for-each select="blockquote"><xsl:apply-templates select="node()"/></xsl:for-each></i>
<font style="normal"><xsl:for-each select="attribution"><xsl:apply-templates select="node()"/></xsl:for-each></font>
</xsl:template>




<!--mise en place des liens vers les notes-->
<xsl:template match="note">
<h2><xsl:apply-templates select="title"/></h2>
<font size="-1">
<xsl:for-each select="noteentry">
<p>
<xsl:variable name="linkend" select=".//link/@linkend"/>
<a name="{anchor/@id}" href="#{$linkend}"><xsl:apply-templates select="link"/></a><xsl:apply-templates
select=".//para"/>
</p>
</xsl:for-each>
</font>
</xsl:template>

<!--mise en forme des liens vers les notes-->
<xsl:template match="section/para/link"><xsl:variable name="linkend" select="@linkend"/>
<a name="{../anchor/@id}" href="#{$linkend}"><xsl:apply-templates select="node()"/></a>
</xsl:template>


</xsl:stylesheet>
```

✍Le document final HTML

Nous avons sélectionné un extrait du document final correspondant au même passage que le document HTML avant transformation.

Apprentissage des Langues et Systèmes d'Information et de Communication

http://alsic.u-strasbg.fr/
Vol.1, Num.1, juin 1998
pp. 3 - 25
Recherche

Réflexions linguistiques et sémiologiques pour une écriture didactique du multimédia de langues

Anne-Laure FOUCHER

Université Paris 6, France

Résumé : *Dans une combinaison originale de domaines souvent envisagés de manière fermée que sont la didactique des langues, la linguistique, la sémiologie et l'ALAO (Apprentissage des LanguesAssisté par Ordinateur), nous mettons en avant la nécessité de concevoir l'image multimédia comme un objet composite, lieu véritable d'interactions entre modes iconique, auditif et linguistique. De cette option théorique découle, selon nous, une philosophie qui préside à la conception d'un multimédia pour l'apprentissage des langues. Nous en donnons les grandes lignes sous forme de jalons méthodologiques destinés à uneécriture du multimédia en langues.*

? 1. Introduction

? 2. Enseignement/apprentissage des Langues et multimédia

? 3. "L'image multimédia" est un lieu d'interactions multimodales

? 4. Implications pour une philosophie de conception du multimédia de langues

? 5. Remarques méthodologiques pour une écriture du multimédia

? 6. Préalables méthodologiques relatifs à l'introduction de l'image dans l'image multimédia

? 7. Quelques jalons méthodologiques pour un cadre d'écriture didactique du multimédia de langues

? 8. Conclusion

? Références

1. Introduction

Lors d'une table ronde en 1996 autour du thème "Environnements interactifs pour l'apprentissage des langues", était soulignée l'importance, pour qui désirait exploiter pédagogiquement des documents hypermédias, d'un passage par les trois étapes suivantes: la première étant la recherche de produits susceptibles de plaire aux apprenants ; la seconde consistant en un repérage dans les produits des parties *"dans lesquelles le rapport texte/son/image [favoriserait] une compréhension aisée et l'envie de s'exprimer ; ce repérage [supposant] un minimum de connaissances sémiologiques."* (Chanier & al, 1996 p. 252) ; la dernière étant consacrée à la conception de *"tâches amenant l'apprenant à manipuler (en compréhension comme en production) des données sonores, visuelles et graphiques"*(Idem: p. 252). Parallèlement à ces considérations et plus récemment, T. Lancien suggérait qu'on optimise la multicanalité[1] dans le multimédia[2], multicanalité qu'il définit comme *"le fait que coexistent sur un même support différents canaux de communication"* et qui *"ne prend de sens que selon les choix que fait la personne qui les consulte"* (Lancien, 1998: p. 24), cela afin de favoriser l'apprentissage de la langue.

• • •

Lien vers la note

Notes

[1] Nous préférons le terme de"multimodalité" qui a l'avantage, selon nous, de prendre en compte la nature et les spécificités des modes en jeu contrairementà "multicanalité" qui renvoie plutôt aux supports véhiculant ces modes.

[2]Dans tout ce texte, le termes"multimédia" renverra à*"des informations stockées sur des supports multiples et diffusées par le médiaélectronique des systèmes d'information. Ainsi un système multimédia favorise la communication interactive d'informations dans un format intégrant des ressources non restreintes aux textes, soit des ressources verbales (texte + audio), soit des ressources verbales et non verbales (diagrammes, images fixes ou animées, vidéo). Cette possibilité de jouer sur des canaux de communication variés(visuel pour le texte et l'image ; oral pour l'audio et les sons) a ouvert des perspectives particulièrement intéressantes en apprentissage des langues, en permettant à l'apprenant de coupler des procédures cognitives de traitement basées sur les aspects verbaux et non-verbaux du langage."* (Chanier, 1998 : p. 5). Plus précisément, nous parlerons du "multimédia intégré", où toutes les ressources sont disponibles sur un support unique, un cédérom par exemple (par opposition au "multimédia réparti", où les ressources sont dispatchées sur la Toile (WWW). (Bodenreider & al,1996 : p. 282).

• • •

Résumé

L'INIST (Institut National de l'Information Scientifique et Technique du CNRS) se tourne vers les nouvelles technologies dans l'édition électronique. Le projet REVELEC sur lequel a porté mon stage, issu d'un partenariat entre la revue ALSIC et l'INIST, définit en quelque sorte la place et le rôle de l'INIST par rapport aux revues électroniques et les problématiques et technologies liées (RDF, XML…). La réalisation d'un prototype de plate forme de revues électroniques a en effet permis de faire une étude de faisabilité et d'évaluer les besoins et les moyens ou services à mettre en œuvre dans ce type de dossier. L'INIST peut alors devenir un **"opérateur technique"** dans le cadre de la conception de revues électroniques en SHS, **intégrant l'accès aux documents électroniques** au format XML dans le portail **CONNECTSCIENCES et proposant des outils de navigation** dans ces ressources électroniques.

Mots clés

Revues électroniques ; XML ; DTD ; RDF ; Dublin Core ; Serveur de navigation ; Chaîne de publication électronique ; DocBook