

5. Conclusion générale et perspective

Les moyens et modes de recherche de l'information en général, et des manuscrits en particulier, ont beaucoup évolué grâce aux nouvelles technologies.

Ces dernières années, plusieurs bibliothèques dans le monde entier ont donné accès à leurs collections au moyen de catalogues électroniques en ligne. Ce nouvel outil offre de nombreux avantages pour les chercheurs, les professionnels de l'information, les érudits, etc. En effet, il facilite :

- ❑ le travail des chercheurs en permettant la localisation à distance des livres et/ou des collections dans une ou plusieurs bibliothèques ;
- ❑ la tâche des catalogueurs (puisqu'il fournit des notices complètes sur le livre);
- ❑ le prêt inter-bibliothécaire;
- ❑ l'échange des notices entre les bibliothèques qui utilisent la même norme de catalogage.

Par ailleurs, les catalogues électroniques imposant de nouveaux matériaux de recherche (l'ordinateur ayant remplacé les fiches cartonnées), de nouvelles pratiques de recherche ont émergé : ainsi la recherche sur ordinateur, via la saisie des mots, s'est-elle substituée à la recherche devant les tiroirs du fichier manuel et par le feuilletage des notices bibliographiques.

L'accès aux manuscrits était à ce jour encore plus ardu. En effet, le catalogue imprimé est resté jusqu'à présent et dans certaines grandes bibliothèques telle la Bibliothèque Nationale de France, le seul moyen de recherche. On trouve cependant des catalogues en ligne dans d'autres bibliothèques possédant des manuscrits, notamment l'Université de Princeton aux Etats-Unis, le Wellcome Institut à Londres, le département « Oriental and India » de la British Library, etc¹³⁷. Ceci dit, l'indexation de ces manuscrits n'étant pas toujours faite, la recherche par mots-clés est impossible.

L'objet de cette thèse était donc de créer un outil de recherche électronique permettant et la localisation d'un manuscrit dans une bibliothèque et l'accès au contenu de ces manuscrits numérisés en mode image.

¹³⁷ Cf. <http://mywebpage.netscape.com/atanasiuvlad/emss/bookmark.html> pour une liste des catalogues en ligne des collections de manuscrits arabes dans d'autres bibliothèques.

La démarche adoptée dans notre travail a été d'approfondir deux pistes théoriques différentes pour aboutir à la création d'une troisième voie aux implications très concrètes (voir le schéma ci-dessous).

5.1. La partie théorique

Notre travail impliquait la maîtrise de deux disciplines radicalement différentes : la première étant l'étude des manuscrits arabes du point de vue codicologique et paléographique, et la deuxième s'inscrivant dans l'étude des nouvelles technologies de l'information.

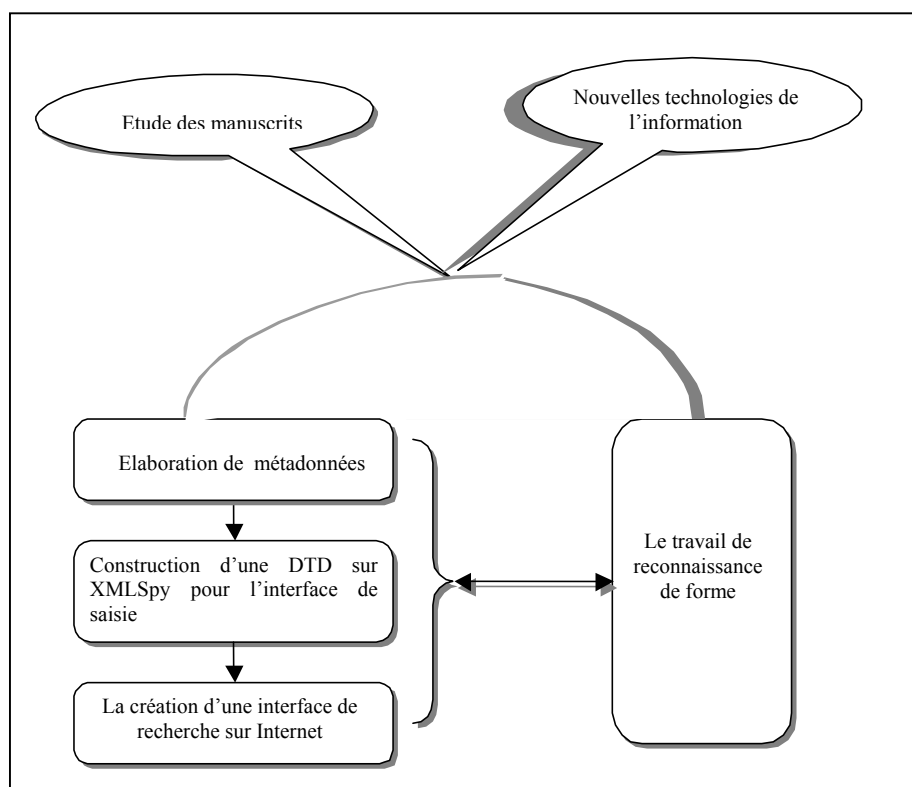


Figure n°.64 : Le schéma du travail

5.1.1. L'étude des manuscrits

Pour amorcer notre travail de thèse, il nous a été indispensable d'étudier les caractéristiques générales des manuscrits arabes, sachant que nos connaissances dans ce domaine étaient initialement peu élaborées. Afin de rattraper cette lacune, nous avons eu

trois approches qui ont consisté en l'étude de la littérature spécialisée dans ce domaine, la réalisation d'une enquête auprès des experts en manuscrits arabes et, finalement, l'examen détaillé des caractéristiques générales de notre échantillon de vingt-et-un manuscrits arabes (voir le document afférent en annexe).

Cette triple démarche nous a ainsi permis de nous familiariser avec ce domaine d'étude et d'élaborer des métadonnées propres à ce genre de document, ces métadonnées et la DTD étant le canevas à partir duquel nous avons construit notre base de données d'une part et l'interface de recherche sur Internet de l'autre.

5.1.2. *Les nouvelles technologies de l'information*

Nous avons dû étudier les technologies de l'information les plus pointues, notamment en matière de numérisation. La numérisation inclut toute une chaîne de processus débutant par le choix de la collection à numériser et se concluant par le stockage et l'indexation des images produites. Cela nous a permis d'acquérir une bonne connaissance des formats de stockage (Gif, JPEG, etc.) et des langues à balises (SGML, HTML, XML).

5.2. La partie pratique

La combinaison de ces deux approches théoriques nous a conduit à la partie pratique de notre travail qui s'est construit autour de trois axes : l'élaboration des métadonnées, la création de l'interface de recherche sur Internet et le travail sur la reconnaissance de forme dans les manuscrits arabes.

5.2.1. *L'élaboration des métadonnées*

Ce sont les 173 éléments qui ont été définis pour décrire les manuscrits arabes. On peut classer les métadonnées en deux grandes catégories:

La première permet au catalogueur de saisir l'information bibliographique sur un document (titre, auteur, lieu, date, etc.) et donne aux chercheurs la possibilité de localiser très précisément le document dont ils ont besoin.

La deuxième décrit l'information sur le contenu du document (les chapitres, les illustrations, les graphiques, etc.). Chaque élément de ce deuxième groupe de

métadonnées est étroitement lié à un URL. Cette dernière pointe sur les pages dans la base d'images qui contiennent les descriptions correspondant à l'élément en question.

5.2.2. *La création de l'interface de recherche*

Les documents saisis dans XML Spy ont été transférés sur le logiciel SDX. Cette plateforme permet la diffusion du document XML sur Internet ainsi que son stockage. SDX donne également la possibilité d'effectuer des recherches simples ou expertes sur le document. Ce travail de création d'interface a été accompli en coopération avec Guillaume Bourgeois, un stagiaire d'IUP à Lyon, présent à l'Enssib d'avril au juin 2003.

Cette interface de recherche n'est actuellement disponible que sur le serveur de l'Enssib*, en attendant que les détenteurs des manuscrits originaux nous accordent leur autorisation de diffusion. Ces autorisations acquises, la base de données sera consultable en ligne sur le site de l'Enssib.

5.2.3. *La reconnaissance de forme.*

Ce processus vise à identifier les divers éléments à l'intérieur du manuscrit arabe numérisé en mode image. Ainsi ont été distingués les titres, les chapitres, les illustrations, les décors séparant les paragraphes et/ou les chapitres, et les cachets, toute indication permettant aux chercheurs de trouver aisément les informations dont ils ont besoin.

Jusqu'à présent la reconnaissance de forme de manuscrits arabes n'a fait l'objet de pratiquement aucune étude. Les résultats obtenus à l'issue de ce travail fait en collaboration avec le laboratoire RFV (Reconnaissance de Forme et de Vision) de l'INSA à Lyon sont par conséquent tout à fait novateurs et de fait, assez encourageants et prometteurs.

5.3. Le partie application

Après la numérisation du document et afin d'obtenir une version exploitable de celui-ci dans la base de données, il faut passer par les étapes suivantes:

* <http://orphee.enssib.fr/SDX>

- La saisie des informations concernant un manuscrit sur XML Spy (qu'on appellera un document XML) ;
- Le transfert des documents XML (contenant toutes les informations sur chaque manuscrit) sur le logiciel SDX qui permet l'affichage et la consultation sur Internet. Evidemment, les images correspondant aux manuscrits seront aussi téléchargées sur SDX.
- Le traitement de l'image se fait séparément grâce à un *exécuteur* conçu spécialement pour notre projet.
- Nous soulignons par ailleurs que la description des éléments des manuscrits - telle qu'elle apparaît dans la base de données actuelle - a été produite manuellement, de même que les liens entre ces éléments descriptifs et les images.

5.4. Les Perspectives

La réalisation de cette base de données ouvre de larges perspectives, intéressantes et nouvelles, dans l'étude des manuscrits arabes. Elle doit cependant encore évoluer, grâce à la collaboration entre experts en manuscrits arabes et ingénieurs informatiques, afin de permettre l'intégration sur une seule plate-forme des trois applications mentionnées plus haut.

Par ailleurs, la recherche en langue arabe devrait être facilitée par l'installation dans SDX d'applications intégrant les caractères arabes, applications qui jusqu'à présent font défaut, tant pour l'arabe que les autres alphabets non latins.

Nous espérons que ce travail s'inscrira à terme dans un projet plus vaste de numérisation des manuscrits arabes de par le monde. Cette base pourrait également servir de prototype au développement de catalogues numériques des manuscrits en caractères non latins, du moment que SDX intégrera des applications pour ces alphabets.

Cette base de donnée se veut la première pierre d'un édifice plus complexe qui exige une collaboration étroite entre le monde de la recherche et celui des nouvelles technologies afin de répondre au mieux aux espérances des chercheurs.