

Rapport de Recherche Bibliographique

XML Schema

Marie Philippe

Sous la direction de Jean-Marie Pinon
 Directeur du Département d'Informatique et Professeur au LISI
 I.N.S.A
 7, Avenue Jean Capelle 69621 Villeurbanne Cedex
 Tel : 04-72-43-84-81 E-mail : pinon@if.insa-lyon.fr

Table des matières

XML SCHEMA : RÉSUMÉ ET DESCRIPTEURS.....	4
PARTIE 1 : MÉTHODOLOGIE.....	5
1. Choix du sujet	5
2. Cibler les spécificités du sujet	6
3. Localiser les sources d'information	6
3.1. Consultation des catalogues.....	6
3.1.1. Les catalogues des bibliothèques : OPAC et Sudoc	6
3.1.2. Les catalogues des librairies et éditeurs scientifiques.....	8
3.2. Interrogation des banques de données	9
3.2.1. Consultation de Dialog	9
3.2.1.1. <i>Détail des équations de recherche</i>	9
3.2.1.2. <i>Analyse des résultats</i>	10
3.2.2. Consultation de ScienceDirect	10
3.2.2.1. <i>Présentation</i>	10
3.2.2.2. <i>Détail des équations de recherche</i>	10
3.2.2.3. <i>Analyse des résultats</i>	11
3.3. Consultation des sommaires des revues et périodiques	11
3.3.1. Le Service de Sommaires de l'ENSSIB (SSE)	11
3.3.2. Consultation manuelle des sommaires des périodiques de Doc'INSA.....	12
4. Elargir la recherche via Internet	12
4.1. Interrogation des moteurs de recherche.....	12
4.1.1. Google	12
4.1.1.1. <i>Présentation</i>	12
4.1.1.2. <i>Equations de recherche</i>	13
4.1.1.3. <i>Analyse des résultats</i>	14
4.1.2. Copernic 2000 Pro.....	14
4.1.2.1. <i>Présentation</i>	14
4.1.2.2. <i>Equations de recherche</i>	14
4.1.2.3. <i>Analyse des résultats</i>	15
4.2. Kartoo	15
4.2.1. Méthodologie de recherche et visualisation des résultats	15
4.2.2. Kartoo: en conclusion	18
4.3. Les sociétés commerciales.....	18
4.3.1. Information Quest.....	18
4.3.2. Intelligence Process.....	18
4.4. Les sites ressources	21
4.4.1. Présentation des principaux portails dédiés à XML	21
4.4.1.1. <i>Les portails francophones</i>	21
4.4.1.2. <i>Les portails anglophones</i>	22
4.5. Les listes de diffusion et les FAQs	22
5. Conclusion	23
PARTIE 2 : SYNTHÈSE.....	25

1. Introduction	25
2. XML Schema : Recommandation du W3C	26
2.1. XML Schema : dans quel but ?	26
2.2. Point de vue technique : l'écriture d'un schéma XML	28
2.2.1. Déclarations d'éléments et d'attributs : type simple et type complexe	28
2.2.1.1. Bibliothèque de types, Types listes, Unions	28
2.2.1.2. Modèle de contenu et connecteurs	29
2.2.1.3. Contraintes d'occurrences sur les éléments	29
2.2.2. XML Schema et les espaces de nommage (namespaces)	29
2.2.3. La dérivation de types	30
2.2.3.1. La dérivation par restriction	30
2.2.3.2. La dérivation par extension	30
2.2.4. Réutiliser et documenter des schémas existants	31
Les concurrents de XML Schema	31
2.3. XML Schema comme alternative aux DTDs	31
2.3.1. Brève présentation des DTDs	31
2.3.2. Les limites des DTDs et la puissance de XML Schema	32
2.4. Les versions « allégées » de XML Schema	33
3. Les outils et applications compatibles avec XML Schema	34
4. XML Schema et les bases de données	35
4.1. Un point commun : les schémas	35
4.2. Etat des lieux des projets en cours	36
4.2.1. Les Systèmes de Gestion de Bases de Données relationnels adoptent XML	36
4.2.2. L'émergence de nouveaux SGBD	36
5. Conclusion	38
BIBLIOGRAPHIE	39
TABLE DES ANNEXES	48

XML Schema : Résumé et descripteurs

Auteur : Marie Philippe

Résumé

XML Schema est un langage de schéma qui se différencie de la Définition de Type de Document (DTD) par sa syntaxe écrite en XML (eXtensible Markup Language). Développé par le World Wide Web Consortium (W3C), il a obtenu depuis mai 2001 le statut de Recommandation. Destiné à décrire la structure et la sémantique d'un document XML, l'objectif de XML Schema est avant tout de combler certaines des lacunes que présentent la DTD afin d'offrir aux développeurs la possibilité de définir avec une plus grande précision des classes de documents XML. Nombreux sont les enjeux liés à XML Schema qui doit donc convaincre l'ensemble de la communauté pour acquérir, à terme, le statut de norme.

Descripteurs primaires : Schéma XML - Langage de définition de schéma XML (XSD) - Définition de Type de Document (DTD) - W3C XML Schema

Descripteurs secondaires : Structure de document XML - Classe de document XML - Document XML valide

Abstract

XML Schema is a schema language which differentiates itself from the Document Type Definition (DTD) with the introduction of XML (eXtensible Markup Language) to describe its syntax. Developed by the World Wide Web Consortium (W3C), it has obtained the status of Recommendation since may 2001. Conceived to define the structure and the semantics of an XML document, its primary goal is to fill the gaps existing in a DTD in order to enable developers to define classes of XML documents more precisely. The issues at stake are numerous and consequently, XML Schema has to convince the whole community in order to acquire in the long run, the norm status.

Primary keywords : XML schema - XML Schema Language Definition (XSD) - Document Type Definition (DTD) - W3C XML Schema

Secondary keywords : XML document structure - XML document class - Valid XML document

Partie 1 : Méthodologie

1. Choix du sujet

Consciente des nouveaux enjeux liés au développement du langage de programmation XML (eXtensible Markup Language), il me semblait judicieux de consacrer mon travail de recherche bibliographique à ce domaine. J'ai donc pris contact avec Mr Pinon, Directeur du département informatique de l'INSA¹ de LYON qui, de par ses fonctions et ses travaux de recherche au sein du LISI², me semblait être la personne la plus appropriée.

Il a su répondre à mes attentes en me proposant d'étudier XML Schema et j'espère avoir moi aussi, au travers de ce rapport, atteint les objectifs qu'il m'avait fixés.

Ce travail de collaboration a été très enrichissant et Mr Pinon a su m'orienter et me conseiller dans mes recherches, afin de cibler au mieux les spécificités du sujet. N'étant pas une « spécialiste » du domaine, je me suis tout d'abord intéressée à XML dans son point de vue général pour être en mesure ensuite de me focaliser sur les particularités de ce langage.

A l'issue de cette recherche, j'ai le sentiment d'avoir beaucoup appris et il me tarde de mettre en pratique les connaissances théoriques que je présente dans cette étude.

¹ INSA : Institut National des Sciences Appliquées

² LISI : Laboratoire d'Ingénierie des Systèmes d'Information

2. Cibler les spécificités du sujet

Comme ce sujet englobe un nombre important de notions, il a fallu définir les différentes problématiques liées à XML Schema afin d'être en mesure d'en sélectionner la plus pertinente.

Ainsi, lors de ma première entrevue avec Mr Pinon, nous avons décidé de nous focaliser plus particulièrement sur les débats et les enjeux relatifs à XML Schema plutôt que de nous intéresser uniquement aux particularités techniques que présentent ce langage.

Nous avons donc convenu de dédier cette recherche à l'étude de l'impact de XML Schema sur le monde des développeurs à savoir, comment ce langage est actuellement perçu, quels avantages il offre et à quelles objections il doit faire face. L'objectif final étant d'évaluer les probabilités que ce langage de schéma obtienne le statut de norme, il semblait également intéressant d'étudier les applications existantes ou en développement permettant l'implémentation de XML Schema.

3. Localiser les sources d'information

3.1. Consultation des catalogues

3.1.1. Les catalogues des bibliothèques : OPAC et Sudoc

Afin d'obtenir un aperçu des ressources disponibles sur mon sujet, j'ai donc successivement consulté les catalogues de :

- La bibliothèque de l'ENSSIB : http://www.enssib.fr/bibliotheque/trouver/cadre_catalogue.html
- La Bibliothèque Municipale de Lyon : <http://sbibbh.si.bm-lyon.fr/>
- La bibliothèque de l'ENS Sciences : <http://140.77.1.21:4505/ALEPH/>
- Doc'INSA : <http://csidoc.insa-lyon.fr/opac/opac2.htm>

- Les bibliothèques des 4 sites de l'INRIA³ : <http://www.inria.fr/publications/centredoc.fr.html>
- La bibliothèque de l'Université Claude Bernard Lyon : <http://buweb.univ-lyon1.fr/webpac-1.2-bin/wgbroker?new+-access+top.bulyon1>
- La Bibliothèque Nationale de France : <http://www.bnf.fr>
- The Library of Congress (USA) : <http://www.loc.gov>
- The British Library (UK) : <http://portico.bl.uk>

Pour chacun des catalogues, j'ai tout d'abord effectué une recherche dans l'index des mots du titre avec comme premier mot-clé « XML ». Cette initiative s'est révélée être fructueuse avec un bon nombre de références proposées. En ajoutant à cette première entrée le terme « schéma », je n'ai obtenu aucun résultat. J'ai donc réitéré cette tentative mais en incluant cette fois-ci, mon mot-clé « schéma » dans l'index des mots du sujet tout en conservant le terme « XML » dans l'index des mots du titre afin d'éviter toute interférence avec les schémas de bases de données. Le croisement de ces deux termes ne m'a fourni aucune référence.

La consultation du catalogue de Sudoc⁴, qui permet d'effectuer des recherches bibliographiques sur les collections des bibliothèques françaises d'enseignement supérieur, ne m'a pas non plus fourni les résultats escomptés. En effet, je me suis heurtée au même phénomène que lors de la consultation des OPACs⁵ à savoir, beaucoup d'ouvrages consacrés à XML mais aucun ne répondant directement à ma problématique.

³ INRIA : Institut National de Recherche en Informatique et Automatique

⁴ Sudoc : Système Universitaire de Documentation

⁵ OPAC : Online Public Access Catalogue

3.1.2. Les catalogues des librairies et éditeurs scientifiques

Compte tenu du caractère très récent de ma problématique et sans oublier le fait que XML Schema n'en n'est qu'au stade d'expérimentation, j'ai donc eu recours aux catalogues en ligne des librairies spécialisées qui sont le plus à même de proposer les ouvrages les plus récents dédiés à mon sujet.

Cette démarche s'est avérée beaucoup plus efficace que les précédentes car la plupart des références proposées dans les catalogues sont accompagnés d'un résumé, de la table des matières détaillée ainsi que, pour certains, des extraits de chapitres consultables en ligne. Ainsi, il m'était plus facile de sélectionner des ouvrages généraux sur XML traitant des schémas XML.

J'ai donc successivement consulté les sites de :

- Wrox Press : <http://www.wrox.com/>
- Eyrolles : <http://www.eyrolles.com>
- Osborne Mc Graw Hill : <http://www.osborne.com>
- Prentice Hall : <http://web21.prenhall.com>
- O'Reilly : <http://xml.oreilly.com/>

Comme précédemment, j'ai utilisé comme mot-clé XML Schema mais cette fois-ci en tant qu'expression. Pour chacun des catalogues interrogés, j'ai obtenu des réponses tout à fait pertinentes qui ont grandement contribué à l'élaboration de ma bibliographie.

3.2. Interrogation des banques de données

3.2.1. Consultation de Dialog

Grâce aux comptes de connexion mis à notre disposition par l'ENSSIB, j'ai consulté Dialog via son serveur web, Dialogweb⁶. Sur les 600 bases de données interrogeables, j'ai sélectionné celles qui me semblaient être les plus pertinentes pour ma recherche. Je me suis donc orientée vers des bases de données qui présentent un fort contenu scientifique.

Afin d'obtenir à la fois des références bibliographiques de monographies, d'articles de périodiques, de rapports de recherche ou encore de pré-publication, j'ai privilégié l'interrogation croisée des bases de données suivantes :

- Pascal: CNRS (Centre National de la Recherche Scientifique)
- Inspec: IEE (Institution of Electrical Engineers)
- EI Compendex: COS (Community of Science)

3.2.1.1. *Détail des équations de recherche*

Voici un aperçu des requêtes et des réponses obtenues :

Requêtes	Syntaxe	Résultats
S1	XML/ti,de	3419
S2	Schema? ?/ti,de	87163
S3	XML(w)schema ? ?/ti,de	77
S4	RD	60
S5	S4 and W3C	8
S6	Dtd or document(w)type(w)definition	646
S7	RD	534
S8	S4 and S7	14
S9	Database? ? and S4	26
S10	RD	21

⁶ Ce site est disponible à l'adresse suivante : <http://www.dialogweb.com/>

3.2.1.2. *Analyse des résultats*

Dans un souci de clarté, j'ai synthétisé les requêtes et les étapes afin de mettre en valeur les résultats obtenus et j'ai eu recours à la commande RD qui permet d'éliminer les doublons. De plus, j'ai appliqué des contraintes au niveau des index en effectuant une recherche sur les index du titre et des descripteurs ce qui permet de garantir que le contenu des références sélectionnées traite de façon conséquente mon sujet et ne fait pas que mentionner le mot-clé. J'ai également limité ma recherche par le biais d'une restriction temporelle en privilégiant les documents édités depuis 1999 et j'ai aussi appliqué une contrainte linguistique en précisant, lorsque c'était possible, que je souhaitais des documents de langue anglaise ou française. Dans la majorité des cas, j'ai choisi le format 7 qui permet de visualiser le titre, l'auteur, les sources et le résumé. Pour les références qui me semblaient particulièrement pertinentes et qui étaient accompagnées de la mention « Full text Available », j'ai privilégié le format 9.

3.2.2. Consultation de ScienceDirect

3.2.2.1. *Présentation*

Edité par Elsevier, cette base de données spécialisée dans les domaines scientifiques propose un grand nombre de références bibliographiques et d'articles de périodiques.

3.2.2.2. *Détail des équations de recherche*

Afin de réduire le bruit au niveau des résultats, j'ai concentré ma recherche sur l'interrogation des journaux de la catégorie « Computer Science » et j'ai limité l'interrogation dans le temps à partir de 1999 jusqu'à aujourd'hui.

J'ai ensuite effectué 3 requêtes, chacune portant à la fois sur l'index du titre, du résumé et des descripteurs :

Requêtes	Syntaxe	Résultats
S1	XML and schema*	11
S2	"XML schema*"	1
S3	"W3C XML schema"	0

3.2.2.3. *Analyse des résultats*

Contrairement à mes attentes, j'ai obtenu peu de résultats. Cependant, les 12 références qui m'ont été proposées se sont révélées être pertinentes et pour certaines, l'accès au texte intégral était permis ce qui m'a servi de base pour ma synthèse.

3.3. Consultation des sommaires des revues et périodiques

3.3.1. Le Service de Sommaires de l'ENSSIB (SSE)

Ce service permet d'effectuer une recherche sur les sommaires de l'ensemble des revues et périodiques des deux bibliothèques de l'ENSSIB. Très efficace, il m'a permis de localiser rapidement les articles correspondant à ma recherche. En effet, sur 18278 articles disponibles, 3 références répondaient à mon interrogation portant sur les mots du titre de l'article avec comme mot-clé « XML Schéma ».

3.3.2. Consultation manuelle des sommaires des périodiques de Doc'INSA

Beaucoup plus fastidieuse, cette démarche est très coûteuse en temps et ne garantit pas de résultats pertinents.

J'ai tout d'abord consulté, via le catalogue des périodiques, l'ensemble des revues spécialisées en informatique disponible à Doc'INSA et j'ai ensuite sélectionné les plus récentes, c'est à dire celles publiées entre 1999 et 2001. En effet, il aurait été inutile de remonter plus avant dans le temps compte tenu que XML Schema fait l'objet d'un développement récent.

Un dépouillement méthodique m'a tout de même permis de compléter ma bibliographie.

4. Elargir la recherche via Internet

4.1. Interrogation des moteurs de recherche

4.1.1. Google⁷

4.1.1.1. *Présentation*

Ce moteur de recherche à couverture mondiale effectue son classement sur une notion d'indice de popularité à deux niveaux, c'est à dire qu'il prend en compte non seulement le nombre de mots-clés présents dans une même page internet mais aussi le nombre de pages qui pointent vers celle-ci.

⁷ Ce site est consultable à l'adresse suivante : <http://www.google.com>

4.1.1.2. *Equations de recherche*

Ma démarche a été de partir d'un point de vue général pour spécifier ensuite, de plus en plus, mes requêtes.

J'ai donc, dans un premier temps, utilisé l'opérateur par défaut « ET » symbolisé par le signe « + ». La requête formulée est la suivante :

- **+XML +schema***
 - Le caractère « * » permet d'élargir ma recherche en obtenant le terme « schema » au pluriel.
 - ➔ Suite à cette requête, Google a sélectionné 418000 références.

Afin de limiter le nombre de ressources proposées, j'ai eu recours aux guillemets pour obtenir dans mes résultats l'expression exacte.

La syntaxe de ma requête est donc:

- **"XML schema"** : Google me propose alors 194000 références
- Avec **"W3C XML schema"**, j'ai obtenu 12500 références.

J'ai ensuite classé ces références par noms de domaines pour avoir un meilleur aperçu du type de ressources proposées :

- **"W3c xml schema" site:.org** :
 - Cette requête m'a retourné 6950 références contenant exclusivement des sites d'organismes.
- **"W3c xml schema" site:.com** :
 - 3960 références dédiées à des sites commerciaux sont répertoriées dans Google.
- **"W3c xml schema" site:.edu**
 - J'ai obtenu avec cette équation, 170 références de sites universitaires.
- **"W3c xml schema" site:.fr**
 - Par cette requête, je souhaitais obtenir des références en langue française sur mon sujet. Google m'en a proposé 73.

4.1.1.3. *Analyse des résultats*

Cet outil s'est révélé être particulièrement performant au niveau des ressources proposées suite à mes requêtes. Malgré le grand nombre de références citées, j'ai pu cibler les plus pertinentes grâce au système de classement de Google. J'ai donc sélectionné, pour chacune des interrogations, les vingt premiers résultats. De plus, en appliquant une restriction sur les noms de domaines, j'ai obtenu un bon aperçu du type de ressources disponibles.

Cependant, l'un des inconvénients majeurs que présente Google est qu'il n'offre pas de possibilité de réduire les doublons au niveau des sites.

4.1.2. Copernic 2000 Pro

4.1.2.1. *Présentation*

Ce méta-moteur permet d'interroger successivement des moteurs de recherche tels que : *Alta Vista*, *Hot Bot*, *All the Web*, *Lycos*, *Yahoo*, *MSN Web Search* et *Open Directory Project*. Il propose également des domaines de recherche spécialisés dans les sciences et les technologies et offre la possibilité d'interroger les groupes de discussion via *Google Usenet Service* et *Topica*. Différentes méthodes de classification des résultats sont proposées, parmi lesquelles j'ai privilégié la classification par pertinence.

4.1.2.2. *Equations de recherche*

Lors de la consultation de Copernic, j'ai effectué une recherche sur 2 expressions : « **XML schema** » et « **W3C XML schema** » en sélectionnant tout d'abord, un domaine de recherche général, c'est à dire *le Web*, puis en précisant ma recherche sur des domaines particuliers proposés par Copernic, à savoir, les *nouvelles sur les technologies*, les *technologies de l'information* et les *groupes de discussion*.

Les résultats obtenus sont représentés au travers du tableau suivant :

	"XML schema"	"W3C XML Schema"
Web	32	22
Nouvelles sur les technologies	46	6
Technologies de l'information	11	8
Groupes de discussion	10	10

4.1.2.3. *Analyse des résultats*

Copernic est un outil de recherche vraiment efficace qui permet d'optimiser le temps de consultation des moteurs de recherche. Ainsi, on obtient rapidement un aperçu global des références disponibles sur un sujet et l'interface de présentation est étudiée pour améliorer leur visualisation.

La stratégie de recherche que j'ai adoptée m'a permis de croiser les références générales avec des références plus spécifiques et d'éliminer les doublons.

4.2. **Kartoo⁸**

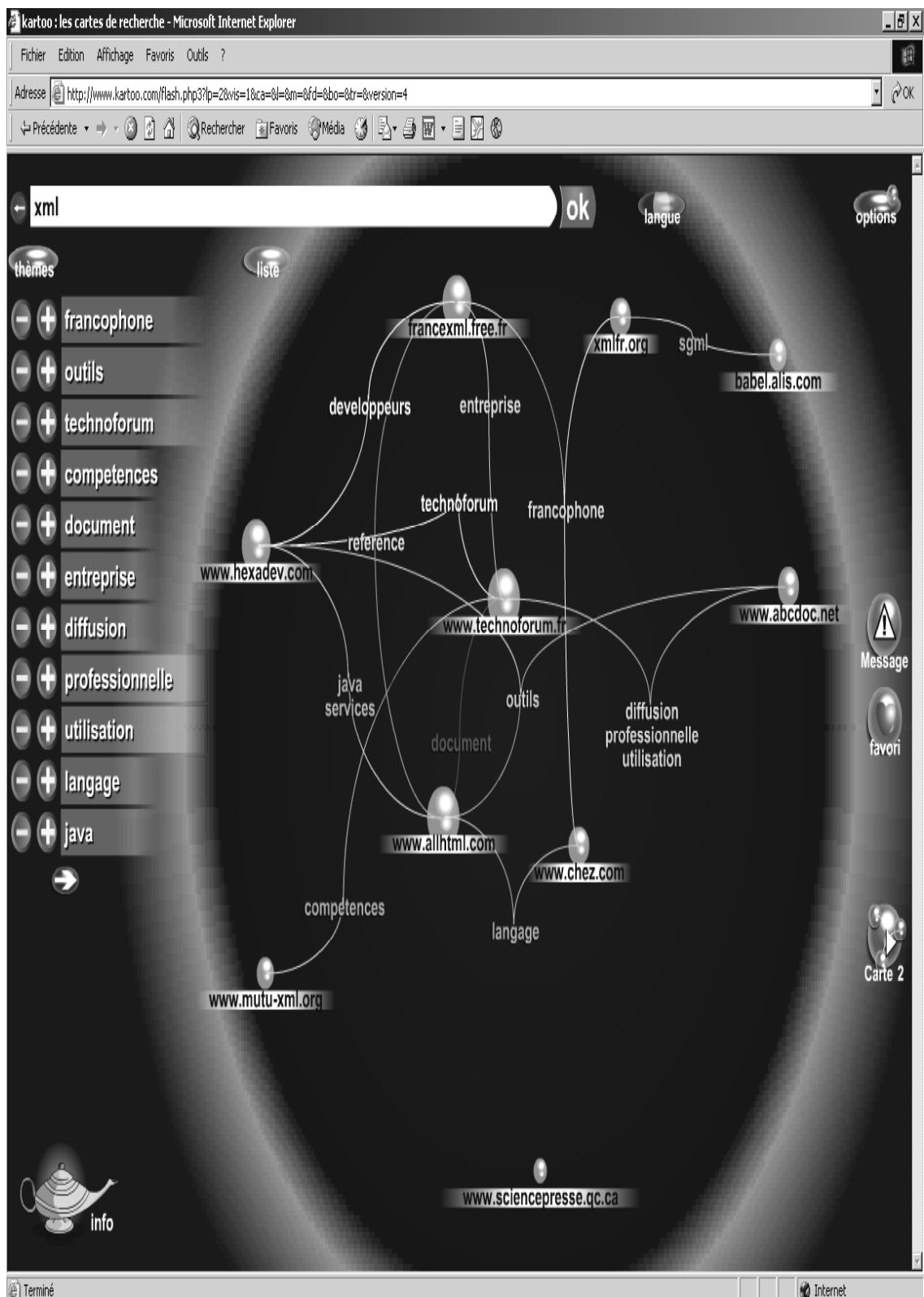
Kartoo est un outil de cartographie qui propose tout un ensemble de thématiques liées à une recherche, visualisables sur une « carte » qui représente les liens sémantiques reliés entre eux.

4.2.1. Méthodologie de recherche et visualisation des résultats

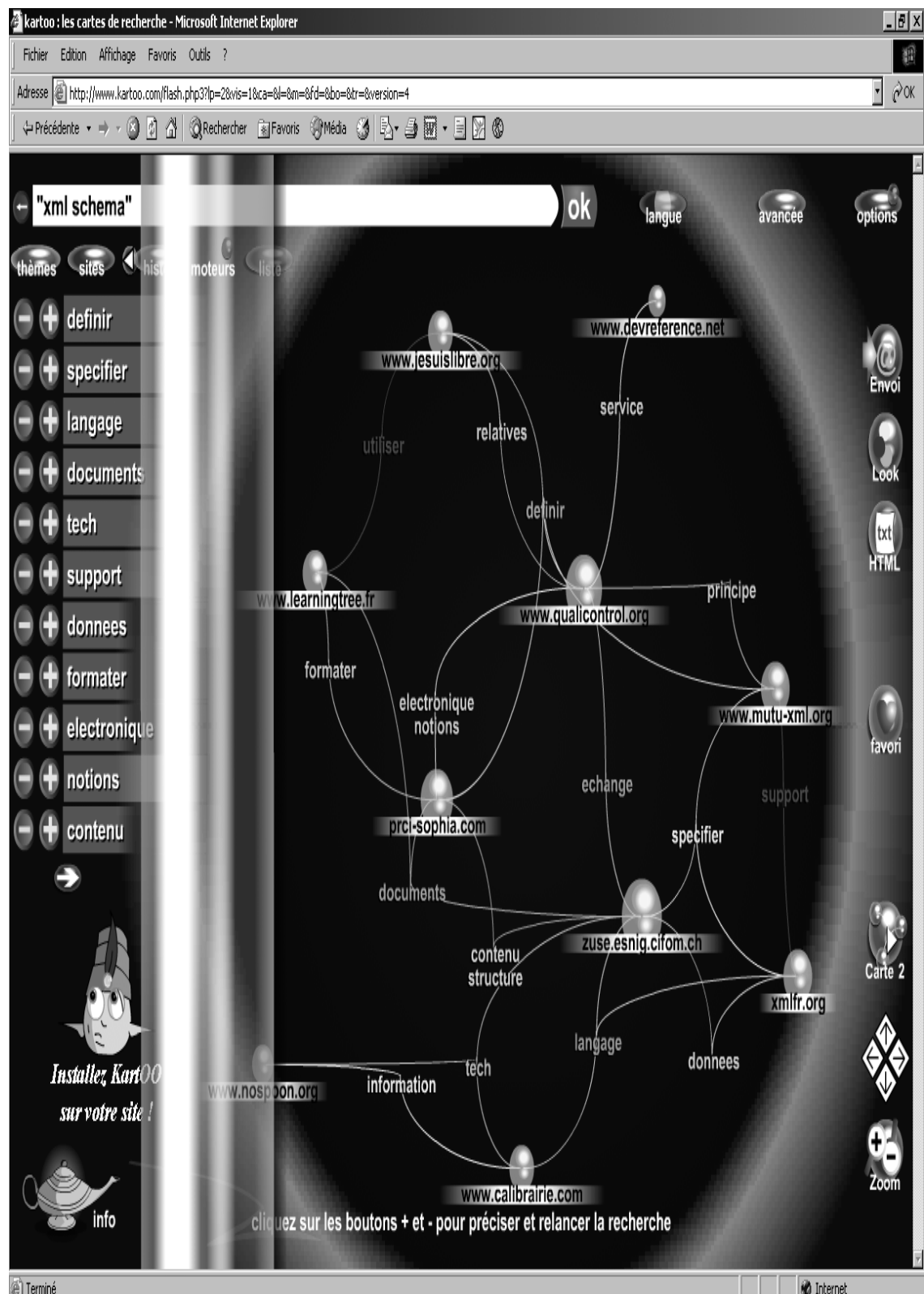
Pour une meilleure compréhension de cette technologie, j'associe à chaque étape de ma recherche les résultats proposés par Kartoo.

⁸ Ce site est consultable à l'adresse suivante : <http://www.kartoo.com>

- Etape 1 : XML



- Etape 2 : « XML schema »



4.2.2. Kartoo: en conclusion

Cet outil innovant offre de nouvelles perspectives quant à la recherche sur Internet. Il m'a permis de délimiter les spécificités de mon sujet et de mieux appréhender les thèmes qui lui sont associés. De plus, l'interface de recherche est soignée et intuitive.

4.3. Les sociétés commerciales

4.3.1. Information Quest⁹

Spécialisée dans la recherche d'information sur Internet, cette entreprise propose un service d'essai afin de faire valoir ses compétences. J'ai donc profité de la gratuité de cette offre pour tester la qualité et la pertinence des références sélectionnées en relation avec mon sujet.

- Ma requête a porté sur l'expression « **W3C XML Schema** » :

J'ai utilisé l'interface de recherche avancée qui permet de classer les résultats par pertinence. J'ai obtenu 200 références qui, pour la plupart, ont répondu tout à fait à mes attentes. En outre, la description de chaque article est très complète et pour certains le texte intégral est disponible.

4.3.2. Intelligence Process¹⁰

J'ai profité du même type d'offre que celle proposée par son concurrent à savoir, la possibilité de tester gratuitement les services de cette entreprise. Elle met donc à disposition deux outils : *My SearchProcess*, un service de recherche d'information sur Internet et *SearchProcessPro* qui est un outil de veille. Pour chacune des requêtes, les réponses sont présentées de façon à permettre de détecter les sites les plus récents et les portails d'information, d'effectuer une analyse synthétique du sujet ou alors, de rechercher des informations précises via l'utilisation du moteur de recherche interne.

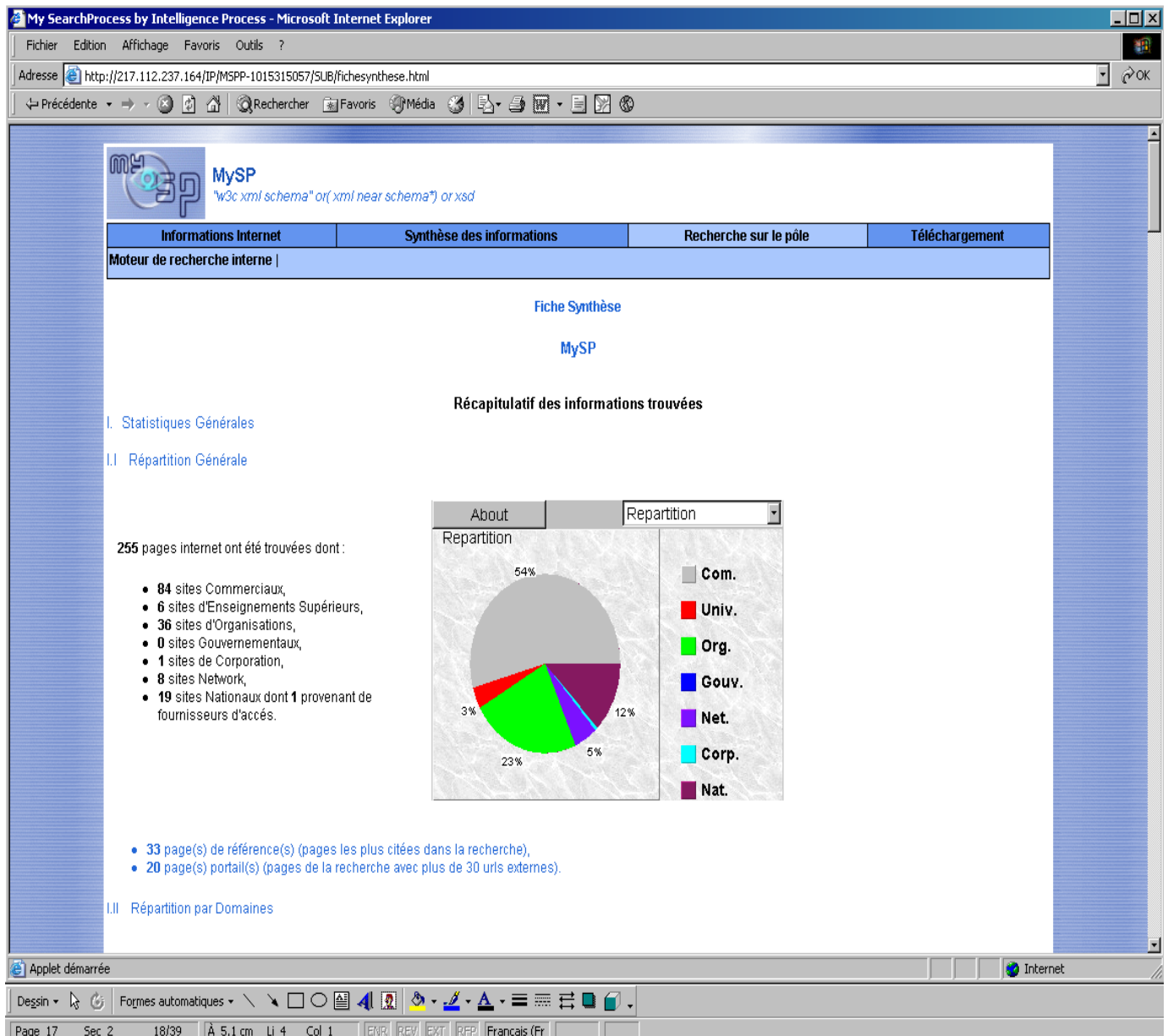
⁹ Ce site est consultable à l'adresse suivante : <http://www.informationquest.com/>

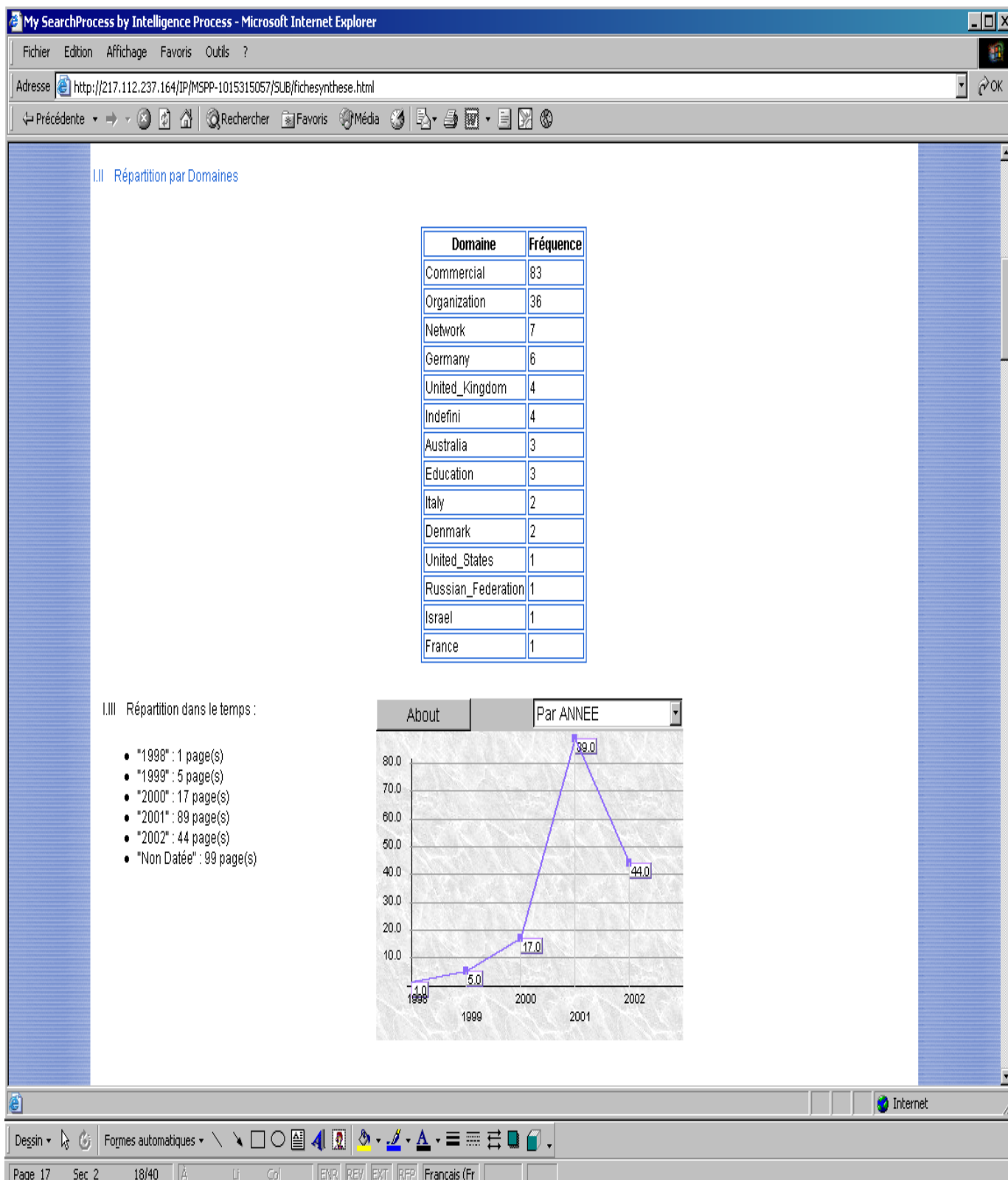
¹⁰ Ce site est consultable à l'adresse suivante : <http://www.sp-portal.com/>

Ma première requête a porté sur le couplage d'expressions suivant :

"w3c xml schema" or(xml near schema*) or xsd

→ 255 pages internet ont été synthétisées sous cette forme :





Ma deuxième interrogation concernait:

("xml schema") and ("base* de données") or sgbd*

→ 413 pages ont été sélectionnées selon le même principe.

Cet outil témoigne des efforts et des évolutions faites en matière de recherche d'information. La possibilité de visualiser les différentes réponses par catégories, diagrammes et graphes donne de la valeur ajoutée aux grands nombres de sélections fournies lors d'une requête et permet de cibler très rapidement les orientations qu'on souhaite lui donner. Cependant, il est bien dommage que cet outil soit payant.

4.4. Les sites ressources

4.4.1. Présentation des principaux portails dédiés à XML

4.4.1.1. Les portails francophones

L'essor de XML et ses perpétuelles évolutions a favorisé l'émergence de portails dédiés à mettre en valeur ce nouveau langage de programmation.

<XML> fr¹¹ en est l'illustration. Ce site relate l'actualité et les nouveaux développements de XML. Il propose, entre autre, des articles en ligne, des tutoriaux ainsi que de nombreuses listes de discussion qui permettent d'appréhender pleinement les enjeux et les débats liés à ce langage.

Mutualiser l'effort de montée en compétences sur XML¹² est un site dédié à la mise en valeur des technologies XML. Comme son nom l'indique, il a pour principal objectif de promouvoir les actions d'études, d'information et de formations sur XML. Très complet, ce site présente des informations à la fois techniques et théoriques et répond aussi bien aux attentes des néophytes que des professionnels.

¹¹ Ce site est disponible à l'adresse suivante : <http://xmlfr.org/>

¹² Ce site est disponible à l'adresse suivante : <http://www.mutu-xml.org/>

4.4.1.2. *Les portails anglophones*

Ces sites ont la même vocation que leurs homologues francophones, c'est à dire, fournir une information détaillée concernant XML. Il en existe un grand nombre, parmi lesquels on peut citer **XML.org**¹³, qui se décrit comme étant le portail de l'industrie XML. **PerfectXML**¹⁴ en est un autre exemple et propose en plus, des ressources bibliographiques et un aperçu détaillé de l'ensemble des logiciels et des applications compatibles avec ce langage.

4.5. **Les listes de diffusion et les FAQs**

Basées sur la technologie «push», les listes de diffusion permettent d'effectuer des actions de veille sur un sujet bien précis. Dans le cadre de ma recherche, je me suis abonnée à la liste de discussion **Atout XML**¹⁵ qui est gérée par l'ATICA (Agence pour les Télécommunications de l'Information et de la Communication dans l'Administration) et a pour objectif la mise en œuvre des technologies XML dans les administrations françaises. J'ai ainsi pu suivre les débats en cours et mesurer les possibilités d'intégration de XML et des technologies qui lui sont associées sur des projets à long terme. J'ai également consulté les rubriques consacrées aux FAQs (Frequently Asked Questions) et plus particulièrement un site internet intitulé, « **XMLSchema FAQ**¹⁶ », qui reprend de façon très organisée l'ensemble des questions relatives à l'écriture d'un schéma XML et tente également de répondre aux principales interrogations concernant les évolutions de XML Schema.

¹³ Ce site est disponible à l'adresse suivante : <http://www.xml.org/>

¹⁴ Pour consulter ce site, rendez vous à l'adresse suivante : <http://www.perfectxml.com/>

¹⁵ Atout XML : http://www.atica.pm.gouv.fr/XML/inscription_atout_xml.shtml

¹⁶ Ce site est en ligne à l'adresse suivante : <http://www.schemavalid.com/faq/xml-schema.html>

5. Conclusion

La stratégie de recherche que j'ai explicitée tout au long de cette partie m'a permis de hiérarchiser et de diversifier les sources d'information relatives à mon sujet. En effet, la consultation des catalogues et l'interrogation des bases de données m'ont fourni des réponses fiables et pertinentes tandis qu'Internet représente un moyen efficace d'élargir la recherche et de bénéficier de sources d'information récentes. En outre, grâce à des outils de plus en plus élaborés, il est devenu possible d'évaluer concrètement la validité d'une information en effectuant notamment un contrôle sur l'auteur d'un site internet, sur le nombre de liens qui font références à cette page et sur sa date de publication. Cette technique de recoupement, mise en œuvre de façon explicite par les sociétés commerciales, se révèle donc être un atout majeur pour valoriser la recherche d'information sur le Web.

Comme je l'ai expliqué, il était primordiale dans cette recherche de bien délimiter mon sujet, afin d'éviter les trop nombreuses références concernant le langage XML dans sa globalité. En effet, cette technologie fait couler beaucoup d'encre et j'aurais été vite dépassée si je n'avais pas choisi de contraindre ma syntaxe de requête soit à des expressions figées comme « W3C XML Schema » ou « XML Schema », soit à la combinaison de deux mots-clés, à savoir « XML » et « Schema », en tenant compte bien sûr des spécificités d'interrogation propre à chacun des outils que j'ai utilisés.

La lecture des références sélectionnées m'a ensuite permis d'évaluer la pertinence de l'information proposée et son adéquation avec mon sujet pour finalement, élaborer une bibliographie la plus complète possible et présenter la synthèse suivante.

Ce travail représente un investissement personnel conséquent tant au niveau du temps que je lui ai consacré, qu'au niveau du coût engendré par la consultation d'outils de recherche payants.

Voici donc une évaluation du temps de travail et des investissements financiers, synthétisés dans le tableau suivant :

	Consultation des catalogues	Consultation de Dialog	Consultation de ScienceDirect	Recherche sur Internet	Elaboration Méthodologie	Elaboration Synthèse	TOTAL
Temps en heures	2	6	2	15	8	10	43
Coût	Gratuit	40 \$	Connexion fournie par l'ENSSIB	Connexion fournie par l'ENSSIB			40 \$

Partie 2 : Synthèse

1. Introduction

Afin de mieux comprendre XML Schema et ses implications, il est tout d'abord nécessaire de présenter brièvement le langage XML et les différentes fonctionnalités qu'il propose.

Créé en 1998 par le W3C (World Wide Web Consortium), XML ou eXtensible Markup Language est une technologie qui permet la description et la structuration des données. Dérivé du SGML (Standard Generalized Markup Language), ce méta-langage est particulièrement performant pour l'échange et le stockage de tous types de données et de leurs structures. Les balises permettent une structuration hiérarchique du document et offrent donc la possibilité d'identifier de façon logique la structure et l'organisation de l'information échangée.

XML se révèle donc comme étant un langage à la fois puissant et flexible et permet de plus, d'effectuer un certain contrôle sur la validité des données et de la structure qui composent un document XML. Ainsi, lorsque les règles de la syntaxe XML sont respectées, le document est dit « bien-formé » et peut alors être représenté sous la forme d'un arbre. Un document est dit « valide » lorsqu'il est structuré conformément au schéma XML qui lui correspond. Dans ce cas, le document XML est une instance d'une classe de documents.

D'un point de vue global, les schémas XML facilitent l'échange des données et permettent une meilleure homogénéisation des documents.

Il existe un grand nombre de langages de schémas, parmi lesquels la DTD (Document Type Definition) et le XML Schema, ou XSD (XML Schema Definition Language) qui fait l'objet d'une Recommandation par le W3C.

Jusqu'à présent, la DTD est le langage le plus fréquemment utilisé pour définir la structure d'une classe de document XML mais, en raison de certaines limites que nous expliciterons en détail, le XML Schema est en passe de devenir la norme de référence.

Nombreux sont les débats qui opposent l'usage de la DTD à XML Schema.

Cette étude propose donc dans un premier temps, de décrire la genèse de XML Schema et de définir les objectifs fixés par le W3C pour détailler ensuite, les spécificités techniques relatives à l'écriture de ce langage. Ces précisions établies nous permettront alors de comparer les différentes fonctionnalités offertes par ces deux langages de schéma aujourd'hui en concurrence.

Parfois critiqué en raison de sa complexité, XML Schema doit s'imposer également à d'autres versions dites « simplifiées » dont nous étudierons les principales caractéristiques.

Afin d'appréhender pleinement les enjeux liés à XML Schema, une description des principaux outils et applications compatibles est également proposée tandis que nous ne manquerons pas d'évoquer les possibilités d'interaction entre XML Schema et les bases de données.

2. XML Schema : Recommandation du W3C

2.1. XML Schema : dans quel but ?

Les objectifs de XML Schema sont décrits dans un rapport du W3C intitulé « *XML Schema Requirements*¹⁷ » en date du 15 février 1999.

Le langage XML Schema doit donc répondre à un certain nombre d'exigences. Il doit, entre autre, être plus performant que les DTDs, auto-descriptif, utiliser la syntaxe XML et être compatible avec un grand nombre d'applications qui utilisent elles-mêmes XML. Les buts fixés concernent également la description de la structure d'un document XML, de son contenu et du type de données qui le constitue.

¹⁷ Ce rapport est disponible sur Internet à l'adresse suivante : <http://www.w3.org/TR/NOTE-xml-schema-req> (CF Bibliographie)

En d'autres termes, XML Schema doit donc permettre de décrire l'imbrication et l'ordre d'apparition des éléments et de leurs attributs dans le but de définir une classe de documents XML, offrir la possibilité de contrôler l'existence d'éléments et d'attributs ainsi que leur contenu et autoriser le typage de données spécifiques.

Le 2 mai 2001, le W3C a rendu publique les spécifications concernant le langage XML Schema. Elles se présentent en trois parties distinctes :

La recommandation « *XML Schema Part 0 : Primer*¹⁸ » est une introduction qui présente de façon générale les fonctions offertes par XML Schema. Ce document qui n'est pas normatif permet de comprendre néanmoins les procédés utilisés pour créer des schémas avec le langage XML Schema.

Par contre, les recommandations « *XML Schema Part 1 : Structures*¹⁹ » et « *XML Schema Part 2 : Datatypes*²⁰ » contiennent une description complète et formelle de la norme. Le tome 1 définit la nature des schémas XML et des parties qui les composent quant au tome 2, il se focalise sur la description des types de données. Il est important de préciser qu'une recommandation est un document validé par le comité directeur du W3C. Il s'agit donc d'un document stable qui peut être utilisé comme document de référence.

¹⁸ Ce rapport est disponible sur Internet à l'adresse suivante : <http://www.w3.org/TR/xmlschema-0> (CF Bibliographie)

¹⁹ Ce rapport est disponible sur Internet à l'adresse suivante : <http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/> (CF Bibliographie)

²⁰ Ce rapport est disponible sur Internet à l'adresse suivante : <http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/> (CF Bibliographie)

2.2. Point de vue technique : l'écriture d'un schéma XML

2.2.1. Déclarations d'éléments et d'attributs : type simple et type complexe

Comme pour une DTD, un schéma XML est composé principalement d'éléments qui sont représentés par les balises. Ces éléments peuvent comporter des attributs et sont associés à un type de données qui peut être défini comme « type simple » ou « type complexe ». Un type simple est un élément ne contenant que du texte, tandis qu'un type complexe est un élément qui peut contenir d'autres éléments ou des attributs. Ces derniers servent à qualifier un élément et ne peuvent être que de type simple. Cependant, il est possible de rendre un attribut optionnel ou obligatoire et de lui attribuer une valeur par défaut. C'est ce que l'on appelle les contraintes d'occurrences.

XML Schema offre également la possibilité de constituer des « groupes d'attributs » afin de faciliter à la fois la lisibilité et la maintenance d'un schéma.

2.2.1.1. *Bibliothèque de types, Types listes, Unions*

Contrairement aux DTDs, XML Schema comporte une bibliothèque intégrée de types simples (ou atomiques) très complète. Pour des raisons de compatibilité, cette bibliothèque reprend les principaux types de données des DTDs mais en ajoute bien d'autres. L'objectif est de permettre un meilleur contrôle des données lorsque le document est traité par un processeur XML. Les types listes, quant à eux, sont des suites de types atomiques. XML Schema possède 3 types de listes intégrés qui sont : NMTOKENS, ENTITIES et IDREFS. Il est également possible de créer de nouveaux types de listes par dérivation de types atomiques existants.

L'élément union est un autre type simple qui permet à un élément ou à un attribut d'être une ou plusieurs instances d'un type formé de la réunion de plusieurs types atomiques ou listes.

2.2.1.2. *Modèle de contenu et connecteurs*

Représenté par des connecteurs, le modèle de contenu permet de définir un type de données complexe en précisant de quelle façon sont imbriqués les éléments qu'il contient.

Le connecteur de séquence permet d'imposer une contrainte sur l'ordre d'apparition des éléments fils (ou sous-éléments) au sein d'un élément de type complexe. Pour être valide, le document instance devra donc être fidèle à l'organisation séquentielle décrite dans le schéma.

Le connecteur de choix permet, dans l'instance de document, de limiter la présence d'un seul élément fils parmi ceux décrits dans le schéma tandis que l'élément « All » permet aux éléments fils d'apparaître une fois ou pas du tout et dans un ordre indifférent.

2.2.1.3. *Contraintes d'occurrences sur les éléments*

XML Schema permet d'imposer des contraintes sur le nombre de fois qu'un élément peut apparaître dans le document instance. Cependant, contrairement aux DTDs, la définition des contraintes dans XML Schema se fait de façon beaucoup plus souple puisque tout nombre entier non négatif peut être utilisé pour fixer la valeur du nombre maximum et/ou minimum d'occurrences.

2.2.2. XML Schema et les espaces de nommage (namespaces)

Comme nous l'avons vu précédemment, un schéma est constitué d'un ensemble de définitions de types et de déclarations d'éléments. Pour éviter tout conflit, chacun des noms qui leur sont respectivement attribués appartient à un espace de nom particulier appelé espace de nom cible (target namespace). Les espaces de nommage permettent donc de fournir un contexte à un vocabulaire, ce qui facilite la création de schémas et la validation de documents.

2.2.3. La dérivation de types

Compte tenu de l'essor des langages de programmation orientés-objet, XML Schema a donc été conçu de façon à supporter les mécanismes d'héritage. La dérivation par restriction et la dérivation par extension sont les deux techniques mises à disposition et s'appliquent aussi bien aux types de données simples que complexes.

Cependant, afin d'éviter toute incohérence, XML Schema permet de contrôler les dérivations effectuées. Ainsi, l'auteur d'un schéma peut donc préciser, pour un type simple ou complexe, s'il est dérivable ou non et si oui quel type de dérivation peut lui être affecté.

2.2.3.1. *La dérivation par restriction*

Cette technique permet de créer de nouveaux types simples, à partir de la bibliothèque de types simples intégrée à XML Schema. Ainsi, grâce aux facettes, il est donc possible de restreindre l'ensemble des valeurs légales d'un type de base. Parmi les différentes facettes qui existent les facettes « enumeration » et « pattern » sont particulièrement utiles. La première permet d'énumérer l'ensemble des valeurs possibles pouvant être attribuées à un type simple, l'autre permet de définir des expressions régulières.

Dans le cas de la restriction de types complexes, les valeurs représentées par le nouveau type sont un sous-ensemble des valeurs du type de base.

2.2.3.2. *La dérivation par extension*

L'extension de types simples consiste en fait, à créer des types complexes dont le contenu est un type simple pouvant porter des attributs. Pour ce qui est de l'extension de types complexes, elle permet de rajouter des éléments ou des attributs au modèle complexe de base.

2.2.4. Réutiliser et documenter des schémas existants

Pour répondre à ces objectifs, différentes techniques ont été développées.

Le mécanisme d'inclusion permet d'utiliser dans un schéma les définitions et les déclarations d'un schéma externe ayant le même espace de nom cible. La même contrainte s'applique au mécanisme de redéfinition qui permet, quant à lui, de redéfinir des types simples ou complexes, ainsi que des groupes d'attributs issus de schémas externes. Par contre, si l'on souhaite intégrer à un schéma des composants de schémas provenant de différents espaces de nom cibles il faut alors faire appel à la technique d'importation de types et associer alors un préfixe à chaque espace de nom auquel le schéma fait référence.

Les annotations permettent de documenter de façon complète un schéma soit à destination d'un lecteur ou d'une application.

Les concurrents de XML Schema

2.3. XML Schema comme alternative aux DTDs

2.3.1. Brève présentation des DTDs

Comme XML Schema, le rôle d'une DTD est de définir des contraintes sur la structure et le contenu d'une instance de document XML. Elle permet de spécifier les éléments et leurs attributs ainsi que leurs relations, leur ordre et leur fréquence d'apparition dans le document XML.

Une DTD est écrite selon la norme EBNF (Extended Backus Naur Form) et, à ce titre, elle est donc héritière du langage SGML (Standard Generalized Markup Language).

Dans la Recommandation du W3C en date du 6 octobre 2000, dédiée à la description de la norme XML version 1.0, la section intitulée « *Prolog and Document Type Declaration*²¹ » explique en détail la grammaire d'une DTD.

2.3.2. Les limites des DTDs et la puissance de XML Schema

On peut définir quatre points majeurs qui différencient ces deux langages de schéma et mettent en avant les performances de XML Schema.

Tout d'abord, au niveau de la syntaxe puisque XML Schema utilise le langage XML. Cela évite d'une part, à l'auteur d'un schéma d'apprendre un nouveau langage et permet également aux schémas XML d'être exploités via différentes applications développées pour XML comme, par exemple : DOM (Document Object Model), SAX (Simple Api for XML) et XSLT (Extensible Stylesheet Language Transformation). Les outils permettant d'exploiter EBNF sont, quant à eux, assez onéreux et leur utilisation est complexe.

De plus, le support étendu du typage des éléments, via la bibliothèque de types intégrée, donne un avantage décisif à XML Schema par rapport aux DTDs, dont les types de données sont plutôt limités (PCDATA, CDATA...).

Nombreuses également sont les différences observées sur le plan du modèle de contenu. En effet, avec le développement de nouvelles applications basées sur XML comme les espaces de nommage, les DTDs se sont révélées trop rigides et trop limitées. XML Schema comble donc cette lacune. En outre, le mécanisme de contraintes d'occurrences d'éléments de XML Schema est beaucoup plus puissant que celui des DTDs, puisqu'il permet de définir de façon précise le nombre d'occurrences d'un élément. Les contraintes d'unicité ont été également révisées afin d'offrir la possibilité de créer des clés et des références à partir de combinaisons d'éléments ou d'attributs, ce qui est impossible avec les DTDs.

²¹ Ce rapport est disponible sur Internet à l'adresse suivante : <http://www.w3.org/TR/REC-xml#sec-prolog-dtd> , Section 2.8 « Prolog and Document Type Declaration » (CF Bibliographie)

Le W3C a aussi veillé à optimiser l'extensibilité des schémas en développant notamment les mécanismes d'importation et de dérivation. Le but étant de favoriser l'évolution de schémas existants par la fusion de schémas ou encore la réutilisation de parties de schémas.

XML Schema offre donc un grand nombre de fonctionnalités supplémentaires comparé aux DTDs. Décrit comme un langage de schéma à la fois puissant et flexible, il lève néanmoins des objections quant à sa complexité.

2.4. Les versions « allégées » de XML Schema

L'objectif principal est de créer une spécification simplifiée d'un langage de schéma pour XML qui offre cependant les mêmes atouts que XML Schema. Parmi les différentes propositions mises en avant, on retrouve principalement Relax NG. Ce langage résulte de la fusion de deux langages de schéma : TREX (Tree Regular Expressions for XML) et RELAX (REgular LAnguage description for XML). La spécification concernant ce langage est explicitée dans le rapport intitulé « Relax NG Spécification²² » élaboré par le Comité d'OASIS (Organization for the Advancement of Structured Information Standards) daté du 3 Décembre 2001.

D'autres langages ont été également développés parmi lesquels Schematron ou encore Minimal XML.

²² Ce rapport est disponible sur Internet à l'adresse suivante: <http://www.oasis-open.org/committees/relax-ng/spec-20011203.html> (CF Bibliographie)

3. Les outils et applications compatibles avec XML Schema

Différents outils ont été développés afin de permettre la lecture et la validation des schémas XML.

Mis en œuvre par le W3C, XSV (XML Schema Validator) en est la principale référence.

Microsoft a également implémenté des outils compatibles avec XML Schema, notamment l'application « Microsoft XML Core Services 4.0 RTM » ou encore « Microsoft XML Well-Formedness Checking and Validation Tool ». Tous deux sont des parseurs XML qui permettent la validation de données XML, d'après un schéma XML.

Oracle est aussi dans la course et commercialise plusieurs applications destinées non seulement à traiter, analyser et valider les schémas conformes au W3C XML Schema, mais aussi à permettre de récupérer l'information sur la structure des données d'un document.

IBM propose quant à lui, « XML Schema Quality Checker 2.0 » qui traite un schéma XML et établit une liste des éléments qui ne sont pas conformes aux spécifications du W3C.

Développé par la société Altova, la version 4.1 de « XML Spy » est un outil très polyvalent. En effet, il permet la saisie graphique des schémas, la validation des documents, la génération de documentation et offre également la possibilité de convertir une DTD en un schéma XML. Cette liste n'est bien sûr pas exhaustive et nombreuses sont les applications encore au stade d'implémentation. De plus, la grande diversité de l'offre en matière d'applications compatibles au W3C XML Schema laisse supposer du succès imminent de cette norme.

4. XML Schema et les bases de données

4.1. Un point commun : les schémas

Selon les auteurs de l'ouvrage « XML et les bases de données » un schéma de base de données est : « *l'organisation ou la structure d'une base de données, provenant généralement de la modélisation de données. Cette structure est décrite à l'aide d'un vocabulaire contrôlé qui nomme des éléments de données et répertorie les contraintes pouvant s'appliquer.*²³ »

Cette définition souligne donc les nombreuses similitudes existant entre un schéma XML et un schéma de bases de données. En effet, comme nous l'avons étudié précédemment, XML Schema est particulièrement performant au niveau de la définition de contraintes et de la création de types de données. De plus, grâce aux contraintes d'unicité, il offre la possibilité de garantir des valeurs uniques, élément clé dans la conception d'une base de données. La puissance de XML Schema repose également sur la possibilité de modéliser des relations tandis que les mécanismes d'héritage offrent des perspectives très intéressantes dans le domaine de la programmation orientée-objet et donc, des bases de données qui correspondent.

En conséquence, cette proximité structurelle entre XML Schema et les bases de données permet d'envisager de nouvelles techniques de traitement, de stockage et d'échange des données. XML Schema rend possible l'intégration de données XML à des bases de données ou à l'inverse, l'extraction de données provenant d'une base de données et leur présentation au format XML. Beaucoup de projets sont en cours actuellement pour veiller à la mise en œuvre de ses nouvelles possibilités offertes par XML Schema et la partie suivante en présente quelques uns.

²³ Williams K., Brundage M., Dengler P., Gabriel J., « XML et les bases de données », Eyrolles, Avril 2001, p.152 (CF Bibliographie)

4.2. Etat des lieux des projets en cours

4.2.1. Les Systèmes de Gestion de Bases de Données relationnels adoptent XML

Dans leurs dernières versions, la plupart des SGBD relationnels offrent la possibilité de traiter les données XML.

C'est la cas notamment d'Oracle qui, dans la version 9i, propose un nouvel environnement destiné à l'intégration, au stockage, à la mise à jour et à l'échange de documents XML. Via son langage d'interrogation SQL, il est désormais possible d'accéder et de générer des documents XML.

On peut citer également « DB2 XML Extender » d'IBM qui propose sensiblement les mêmes fonctionnalités que son concurrent.

Objectivity, quant à lui, compte parmi les rares développeurs à s'être penché sur l'intégration de documents XML à des bases de données orientées-objet. Avec « Objectivity/DB » il est donc possible à la fois d'extraire des objets et de les formater en XML et d'intégrer des données XML aux tables.

4.2.2. L'émergence de nouveaux SGBD

Le stockage XML natal est une alternative au stockage des documents XML dans les bases de données dites « traditionnelles ». « Ipedo XML Database » et « Tamino » sont deux exemples d'une nouvelle génération de SGBD.

Développé par Software AG, « Tamino » a été conçu pour stocker des documents XML sans les convertir dans un autre format (ce qui n'est pas le cas des bases de données relationnelles) et intégrer des sources de données externes en les transformant instantanément au format XML. Cette base de données est la première à permettre de stocker, d'intégrer et d'échanger directement des documents XML, ce qui optimise grandement le partage des données et est tout particulièrement adapté aux applications de commerce électronique.

Pour ce qui est de Ipedo, cette base de données se révèle être particulièrement efficace quant à l'accès et au stockage des documents XML. En effet, elle propose un accès intégral ou partiel aux documents XML et fait preuve d'une bonne rapidité d'exécution des requêtes grâce à un puissant moteur de recherche et de conversion, à un système de caches, et à ses index. Sa grande capacité de stockage (jusqu'à plusieurs giga bytes de documents) ajouté à l'organisation des documents XML en collections font de « Ipedo » un outil performant et prometteur.

Compte tenu de l'essor de XML, beaucoup d'autres développeurs de bases de données proposent également des solutions de stockage et de traitement des documents XML. Les fonctionnalités modulaires et structurales offertes par XML Schema devrait, en outre, faciliter considérablement la gestion des documents XML.

5. Conclusion

A ses débuts, XML Schema s'est heurté à bon nombres d'objections émanant de la part des développeurs. Par certains, jugé trop complexe et par d'autres trop simple, XML Schema a dû répondre aux principales attentes de chacun. Le groupe de travail qui lui est dédié semble avoir aujourd'hui accompli cette lourde tâche et le fait qu'il ait le statut de Recommandation depuis le 2 Mai 2001, laisse présager de l'évolution favorable de ce nouveau langage de schéma.

L'un des atouts majeurs de XML Schema repose sur sa syntaxe XML, ce qui permet à de nombreuses applications développées pour ce langage de pouvoir également gérer les schémas XML. Cependant, il est important de souligner que les DTDs ne sont pas vouées à disparaître et restent très utiles pour concevoir des modèles de documents XML relativement simples. Comme nous l'avons précisé tout au long de cette étude, XML Schema est une alternative à la DTD et a pour objectif d'apporter des fonctionnalités supplémentaires comparé à son prédécesseur. De plus, les applications comme « XML Spy » sont un bon exemple de compromis entre ces deux langages en offrant de précieuses possibilités de conversion.

La programmation en langage XML offre de nouvelles perspectives au niveau de la gestion et de la diffusion des données et sa force, comparé à HTML, repose entièrement sur ses capacités à décrire la structure d'un document. En ce sens, XML Schema en est son principal pilier.

Bibliographie

MONOGRAPHS

- 1. CAGLE K., DUCKETT J., GRIFFIN O. et al.** *Professional XML Schemas*. Chicago : Wrox Press, 2001, 800p. ISBN: 1-861005-47-4.
- 2. DYKES L., VALENTINE C., TITTEL E.** *XML Schemas*. Londres : Sybex, 2002, 632p. ISBN : 0-7821-4045-9.
- 3. GULBRANSEN D.** *Special edition using XML Schema*. Paris : Eyrolles, 2001, 482p. ISBN: 0-7897-2607-6.
- 4. PRONK R.** *XML Schema Developer's Guide*. Berkeley : Osborne Macgraw-Hill, 2001, 1008p. ISBN : 0-07219-030-2.
- 5. VAN DER VLIST E.** *XML.com: The Guide to the W3C XML Schema*. Paris : Editions O'Reilly, 2001, 66p. ISBN : 0596002645.
- 6. WALMSLEY P.** *Definitive Xml schema*. New Jersey : Prentice Hall, 2001, 556p. ISBN : 0-13-065567-8.

CHAPITRES D'OUVRAGES

- 7. BIRBECK M., DUCKETT J., GUDMUNDSSON O.G. et al.** *Professional XML*. 2nde édition. Chicago : Wrox Press, 2001. Chapitre 6, Introducing XML Schema, p 191-240. Chapitre 7, Schema Alternatives, p 241-327. ISBN 1-861005-05-9.
- 8. CAGLE K., DIX C., HUNTER D., et al.** *Beginning XML*. 2nde édition. Chicago : Wrox Press, 2001. Chapitre 10, Valid XML: Schemas, p 379-420. ISBN 1-861005-59-8.
- 9. DENGLER P., HOSKINSON A., BRUNDAGE M., et al.** *Professional XML Databases*. Chicago : Wrox Press, 2000. Chapitre 5, XML Schemas, p 143-190. ISBN : 1-861003-58-7.
- 10. MICHARD A.** *XML : langage et applications*. 2nde édition. Paris : Eyrolles, 2001. Chapitre 3, DTD, domaines nominaux et schémas XML. ISBN 2-212-09206-7.

11. YOUNG M.J. *XML Step by Step*. 2nde édition. Portland : Microsoft Press, 2001. Chapitre 7, Creating Valid XML Documents Using XML Schemas, p 163-192. ISBN : 0-7356-1465-2.

OUVRAGES À VENIR...

12. BINSTOCK C., CLEARY D., RAPOZA J., et al. *XML Schema*. Boston : Addison Wesley Professional, Juillet 2002, 912p. ISBN : 0672323745.

13. JELLIFFE R. *XML Schema Interlinear*. Greenwich : Manning Publications, Mars 2002, 600p. 1-930110-26-X.

14. LOWERY J. *XML Schema Elucidated*. Greenwich : Manning Publications, Mars 2002, 350p. ISBN : 1930110154.

15. VAN DER VLIST E. *XML Schema*. Paris : Editions O'Reilly, Mai 2002, 350p. ISBN : 0596002521.

16. WYKE R.A., WATT A. *XML Schema Essentials*. New York : John Wiley & Sons, Avril 2002, 304p. ISBN : 0-471-41259-7.

ARTICLES DE PÉRIODIQUES – PUBLICATIONS EN SÉRIE – COMPTES-RENDUS DE CONFÉRENCES

- ARTICLES DE PÉRIODIQUES

17. DYCK T. *Translating XML Schema*. eWeek, 28 Mai 2001, p 1. ISSN : 1530-6283.

18. DYCK T. *XML Schema*. IT Week, 2001, Vol.4, N°5, p 21.

19. GILLMOR S., GALLAGHER S. *The Status of Schemas*. XML Magazine, 2001, Vol.2, N°1.

20. ELSON M. *Validation with MSXML and XML Schema*. Windows Developer Magazine, Vol.13, N°1, p 35.

21. HOLLAND R. *XML Schema catches heat*. eWeek, 23Avril 2001, ISSN : 1530-6283.

22. KIELY D. *XML Schema Let Diverse Apps More Easily Share Data*. InformationWeek, 16 Octobre 2000. ISSN : 8750-6874.

23. RAPOZA J. *A better way to share information: XML Schema permits rich data descriptions*. eWeek, 2001, Vol.18, N°18, p 18. ISSN : 0740-1604.

24. SIMON S. H. *XML Schemas Get the Nod*. Intelligent Enterprise, 2001, Vol.4, N°11, p 54.

25. SLIWA C. *W3C readies long-awaited XML Schema specification*. Computerworld, 2000, Vol.34, N°44, p 12. ISSN : 0010-4841.

26. TAUBER J. *The evolution of XML Schema*. E-Business Advisor, 2000, Vol.18, N°5, p 26-28. ISSN : 1098-8912.

- PUBLICATIONS EN SÉRIE

27. BHOWMICK S.S., WEE K.N., SANJAY M.B. *Schemas for web data: a reverse engineering approach*. Vol.39, N°2. Data & Knowledge Engineering, 2000. p 105-142.

28. IOANNIDES D. *XML schema languages : beyond DTD*. Vol.18, N°1. United Kingdom : Library Hi Tech, 2000. p 9-14. ISSN : 0737-8831.

29. LEE D., CHU W.W. *Comparative Analysis of Six XML Schema Language*. Vol.29, N°3. New-York : ACM SIGMOD Record, September 2000. p 76-87.

30. MURATA M. *XML Schema and Relax*. Vol.84, N°12. Japon : Journal of the Institute of Electronics, Information and Communication Engineers, 2001. p 890-894. ISSN : 0913-5693.

31. NEEDLEMAN M.H. *XML Schema Language: The New Way of Coding XML*. Vol.27, N°2. Serials Review, August 2001. p 49-52.

32. VORTHMANN S., ROBIE J. *Beyond schemas: schema adjuncts and the outside world*. Edited by MIT Press. Vol.2, N°3. Chapel Hill : Markup Languages: Theory & Practice, 2000. p 281-294. ISSN : 1099-6621.

- **COMPTES-RENDUS DE CONFERENCES**

33. BIRD L., GOODCHILD A., HALPIN T. *Object Role Modelling and XML Schema*. International conference on conceptual modeling. In Lecture notes in computer science. Vol.1920. New-York : Springer-Verlag, 2000. p 309-322. ISBN: 3-540-41072-4.

34. BEERI C., MILO T. *Schemas for Integration and Translation of Structured and Semi-structured Data*. 7th International Conference in Jerusalem. In Lecture Notes in Computer Science. Vol.1540. New-York : Springer-Verlag, 1999. p 296-313. ISBN : 3-540-65452-6.

35. BEHRENS R. *A grammar based model for XML Schema integration*. 17th British National Conference on Databases. In Lecture Notes in Computer Science. Vol.1832. New-York : Springer-Verlag, 2000. p 172-190. ISBN : 3- 540-67743-7.

36. BROWN A., FUCHS M., ROBIE J., et al. *MSL: A Model for W3C XML Schema*. Int'l World Wide Web Conference. Hong-Kong : VLDB, May 2001. ISBN : 1-58113-3480.

37. CHIDLOVSKII B. *Using regular tree automata as XML Schemas*. Proceedings IEEE Advances in Digital Libraries. Los Alamitos : IEEE Computer Society, 2000. p 89-98. ISBN : 0-7695-0659-3.

38. KLARLUND N., MÖLLER A., SCHWARTZBACH M.I. *DSD: A Schema Language for XML*. Proceedings of the 3rd Workshop on Formal Methods in Software Practice. New-York : ACM Press, 2000. p 101-111.

39. KOTSAKIS E., BÖHM K. *XML Schema Directory: A Data Structure for XML Data Processing*. Proceedings of WISE 2000: 1st International Conference on Web Information Systems Engineering. Vol.1. Los Alamitos : DBLP, 2000. p 62-69.

40. MANI M., LEE D., MUNTZ R. R. *Semantic data modeling using XML Schemas*. International conference on conceptual modeling. In Lecture notes in computer science. Vol.2224. New-York : Springer-Verlag, 2001. p 149-163. ISBN : 3-540-42866-6.

41. RAJEEV A., MOHANIA M.K., KAMBAYASHI Y., et al. *An Architecture for XML Schema Integration*. Kyoto International Conference on Digital Libraries. Kyoto : DBLP, 2000. p 240-244.

42. XIAO R., DILLON T. S., CHANG E., FENG L. *Modeling and transformation of object-oriented conceptual models into XML Schema*. DEXA 2001 : database and expert systems applications. In Lecture Notes in Computer Science. Vol.2113. New-York : Springer-Verlag, 2001. p 795-804. ISBN : 3-540-42527-6.

SITES INTERNET

43. BIRON P.V., MALHOTRA A. *XML Schema Part 2: Datatypes*. W3C Recommendation. [On-line]. May 2001. [06.12.2001] Available from Internet : <URL : <http://www.w3.org/TR/xmlschema-2/>>.

44. BOURRET R. *Mapping W3C Schemas to Object Schemas to Relational Schemas*. [On-line]. March 2001. [05.01.2002] Available from Internet : <URL : <http://www.rpbourret.com/xml/SchemaMap.htm>>.

45. CLARK J., MURATA M. *RELAX NG Specification*. [On-line]. December 2001. [08.01.2002] Available from Internet : <URL : <http://www.oasis-open.org/committees/relax-ng/spec-20011203.html>>.

46. COSTELLO R.L., SCHNEIDER J.C. *Challenge of XML Schemas: Schema Evolution*. [On-line]. May 2000. [09.12.2001] Available from Internet : <URL : <http://www.xfront.org/EvolvableSchemas.html>>.

47. DODDS L. *Schemas revisited*. [On-line]. July 2000. [08.12.2001] Available from Internet : <URL : <http://www.xml.com/pub/a/2000/07/12/deviant.html>>.

48. FALLSIDE D.C. *XML Schema Part 0: Primer*. W3C Recommendation. [On-line]. May 2001. [06.12.2001] Available from Internet : <URL : <http://www.w3.org/TR/xmlschema-0>>.

49. FITZGERALD M. *XML Schema and RELAX NG Element Comparison*. [On-line]. February 2002. [01.03.2002] Available from Internet : <URL : <http://www.wyeast.net/compare.html>>.

50. JELLIFFE R. *The XML Schema Specification in Context*. [On-line]. February 2000. [06.12.2001] Available from Internet : <URL : <http://www.ascc.net/~ricko/XMLSchemaInContext.html>>.

51. MALHOTRA A., MALONEY M. *XML Schema Requirements*. [On-line]. February 1999. [05.12.2001] Available from Internet : <URL : <http://www.w3.org/TR/NOTE-xml-schema-req>>.

52. **MERTZ D.** *Comparing W3C XML Schemas and Document Type Definitions*. [On-line]. March 2001. [08.12.2001] Available from Internet :
<URL : <http://www106.ibm.com/developerworks/library/x-matters7.html>>.
53. **PETERSON D.** *What's a Schema Anyway?* [On-line]. December 2001. [16.12.2001] Available from Internet :
<URL : <http://www.idealliance.org/xmlfiles/issue32/schema.asp>>.
54. **ROBERTSSON E.** *Combining the Power of W3C XML Schema and Schematron*. [On-line]. February 2002. [01.03.2002] Available from Internet : <URL : http://www.topologi.com/public/Schtrn_XSD/Paper.html>.
55. **SAINT LAURENT S.** *Information Supply Chain: XML Schemas Get the Nod*. Intelligent Enterprise [On-line]. July 23, 2001. [09.12.2001] Available from Internet :
<URL : http://www.intelligententerprise.com/010723/411infosc1_1.shtml>.
56. **SAINT. LAURENT S.** *Describing Your Data: DTDs and XML Schemas*. [On-line]. December 1999. [09.12.2001] Available from Internet :
<URL : <http://www.xml.com/pub/1999/12/dtd/>>.
57. **SMITH D.** *Understanding W3C Schema Complex Types*. [On-line]. August 2001. [10.12.2001] Available from Internet :
<URL : <http://www.xml.com/pub/a/2001/08/22/easyschema.html>>.
58. **THOMPSON H.S., BEECH D., MALONEY M., et al.** *XML Schema Part 1: Structures*. W3C Recommendation. [On-line]. May 2001. [06.12.2001] Available from Internet :
<URL : <http://www.w3.org/TR/xmlschema-1/>>.
59. **VAN DER VLIST E.** *Comparing XML Schema*. [On-line]. December 2001. [05.01.2002] Available from Internet :
<URL : <http://www.xml.com/pub/a/2001/12/12/schemacompare.html> >.
60. **VAN DER VLIST E.** *Using W3C XML Schema*. [On-line]. October 2001. [08.12.2001] Available from Internet :
<URL : <http://www.xml.com/pub/a/2000/11/29/schemas/part1.html>>.
61. **WALKER WHITLOCK N.** *XML Schema becomes W3C Recommendation: What This Means*. [On-line]. May 2001. [06.12.2001] Available from Internet :
<URL : <http://www-106.ibm.com/developerworks/xml/library/xschrec.html>>.
62. **WALSH N.** *Understanding XML Schemas*. [On-line]. July 1999. [08.12.2001] Available from Internet :
<URL : <http://www.xml.com/pub/a/1999/07/schemas/> >.

63. WILLIAMS K. *Soapbox: Why XML Schema beats DTDs hands-down for data : A look at some data features of XML Schema.* [On-line]. June 2001. [10.01.2002] Available from Internet :
<URL : <http://www-106.ibm.com/developerworks/xml/library/x-sbsch.html>>.

A LA PÉRIPHÉRIE DE XML SCHEMA : LES BASES DE DONNÉES

MONOGRAPHIES

- 1. ABITEBOUL S., BUNEMAN P., SUCIU D.** *Data on the Web: From Relations to Semistructured Data and XML.* San Fransisco : Morgan-Kaufman, 1999. 257p. ISBN : 1-55860-622-X.
- 2. ARCINIEGAS F.** *XML Developer's Guide.* Berkeley : McGraw-Hill Osborne Companies, 2000. 200p. ISBN : 0-07-212648-5.
- 3. CHANG B., SCARDINA M., KIRITZOV S.** *Oracle9i XML Handbook.* Berkeley : McGraw-Hill Osborne Companies, 2001. 528p. ISBN : 007213495X.
- 4. GRAVES M.** *Designing XML Databases.* New Jersey : Prentice Hall, 2002. 680p. ISBN : 0-13-088901-6.
- 5. MUENCH S.** *Building Oracle XML applications.* Sebastopol CA : O'Reilly & Associates, 2000. 792p. ISBN : 1565926919.
- 6. QUIN L.** *Open source XML Database Toolkit: Resources and Techniques for Improved Development.* New York : Wiley, 2000. 434p. ISBN : 0-471-37522-5.

ARTICLES DE PÉRIODIQUES ET PUBLICATIONS EN SÉRIE

- PUBLICATIONS EN SÉRIE

- 7. BARU C.K.** *XViews: XML Views of Relational Schemas.* Proceedings of the 10th International Workshop on Database & Expert Systems Applications. Washington : IEEE Computer Society, 1999. p700-705.

8. **CAREY M.J., WIDOM J., BOSWORTH A., et al.** *Of XML and databases (panel session): where's the beef?* Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. Vol.29, N°2. New-York : ACM Press, 2000. p 576. ISSN : 01635808.
9. **DONGWON L., WESLEY W.C.** *Constraints-Preserving Transformation from XML Document Type Definition to Relational Schema.* 19th International Conference on Conceptual Modeling. In Lecture Notes in Computer Science. Vol.1920. New-York : Springer-Verlag, 2000. p 323-338. ISBN : 3-540-41072-4.
10. **DEUTSCH A., FERNANDEZ M., SUCIU D.** *Storing semistructured data with STORED.* Proceedings of the international conference on Management of data. Vol.28, Issue 2. New York : ACM SIGMOD Record, 1999. p. 431-442.
11. **FLORESCU D., KOSSMANN D.** *Storing and Querying XML Data Using an RDBMS.* Vol.22, N°3. Washington : IEEE Data Eng. Bulletin, 1999. p 27-34.
12. **JONES M.B., BERKLEY C., BOJILOVA J., et al.** *Managing scientific metadata.* Vol.5, N° 5. Washington : IEEE Internet computing, 2001, p 59-68.
13. **MALAIKA S.** *Using XML in Relational Database Applications.* 15th International Conference on Data Engineering proceedings. Silver Spring : IEEE Computer Society Press, 1999. p167. ISSN : 1063-6382.
14. **RENNER A.** *XML Data and Object Databases: The Perfect Couple?* 17th International Conference on Data Engineering. Washington : IEEE, April 2001. p 143-148. ISBN : 0769510019.
15. **SHANMUGASUNDARAM J., SHEKITA E., BARR R., and al.** *Efficiently publishing relational data as XML documents.* VLDB Journal. Vol.10, N°2. New-York : Springer-Verlag, 2001. p 133-154. ISSN : 1066-8888.
16. **SHIMURA T., YOSHIKAWA M., UEMURA S.** *Storage and Retrieval of XML Documents using Object-Relational Databases.* Int'l Conf. on Database and Expert Systems Applications. In Lecture Notes in Computer Science. Vol.1677. New-York : Springer-Verlag, 1999. p 206-217. ISSN : 03029743.
17. **SCHÖNING H.** *Tamino - A DBMS Designed for XML.* 17th International Conference on Data Engineering. Washington : IEEE, April 2001. p 149-154. ISBN : 0-7695-1001-9.

18. WÄSCH J., SCHÖNING H. *Tamino - An Internet Database System*. 7th International Conference on Extending Database Technology. In Lecture Notes in Computer Science. Vol.1777. New-York : Springer-Verlag, 2000. p 383-387. ISBN : 3-540-67227-3.

19. YAMANE Y., IGATA N., NAMBA I. *High-performance XML storage and retrieval system*. Vol.36, N° 2. Fujitsu Scientific and Technical Journal, 2000. p 185-192.

- ARTICLES DE PÉRIODIQUES

20. DOUGHERTY M.S. *More XML Fundamentals*. DB2 Magazine, 2001, Quarter 3, Vol.6, Issue 3.

21. GOLDMAN R., Mchugh J., WIDOM J. *Lore: A Database Management System for XML*. Dr. Dobb's Journal of Software Tools, April 2000, Vol.25, N°4. p76-80. ISSN : 1044-789X.

22. WHITE K. *DBMS: Past, Present, and Future: Where databases have been, where they are going, and what it means to you*. Dr.Dobb's Journal of Software Tools. 2001, Vol.26, N°8, p 21-29.

SITES INTERNET

23. BOURRET R. *XML and Databases*. [On-line]. September 1999 [05.12.2001] Available from Internet :
<URL : <http://www.rpbourret.com/xml/XMLAndDatabases.htm>>.

24. MERTZ D. *Putting XML in context with hierarchical, relational, and object-oriented models*. [On-line]. April 2001. [08.12.2001] Available from Internet : <URL:<http://www106.ibm.com/developerworks/library/xmatters8/index.html>>

25. OBASANJO D. *An Exporation of XML in Database Management ystems*. [On-line]. October 2001. [08.12.2001] Available from Internet :
<URL : <http://www.25hoursaday.com/StoringAndQueryingXML.html>>.

26. STAKEN K. *Introduction to Native XML Databases*. [On-line]. October 2001. [05.12.2001] Available from Internet :
<URL:<http://www.xml.com/pub/a/2001/10/31/nativexmlldb.html>>.

27. WILLIAMS K. *XML for Data: Native XML databases: a bad idea for data?* [On-line]. October 2001. [05.12.2001] Available from Internet :
<URL : <http://www-106.ibm.com/developerworks/xml/library/x-xdnat.html>>.

Table des annexes

L'ÉCRITURE D'UN SCHÉMA : UN EXEMPLE PRATIQUE	I
BIBLIOTHÈQUE DE TYPE INTÉGRÉE À XML SCHEMA	II

L'écriture d'un schéma : un exemple pratique

Soit une instance de document XML décrivant un fichier client²⁴ :

```
<?xml version = "1.0" ?>
  <Customer>
    <FirstName>Raymond</FirstName>
    <MiddleInitial>G</MiddleInitial>
    <LastName>Bayliss</LastName>
  </Customer>
```

Le schéma qui lui correspond et donc qui décrit sa structure est formalisé de la façon suivante :

```
<?xml version = "1.0" ?>
<schema xmlns = "http://www.w3.org/2001/XMLSchema">
  <element name = "Customer">
    <complexType>
      <sequence>
        <element name = "FirstName" type = "string" />
        <element name = "MiddleInitial" type = "string" />
        <element name = "LastName" type = "string" />
      </sequence>
    </complexType>
  </element>
```

²⁴ Cet exemple est extrait de « Professional XML Schema », p 13

Bibliothèque de type intégrée à XML Schema

Les tableaux suivants sont extraits du site internet : **mutu-xml.org**²⁵

Chaînes de caractères :

Type	Signification	Modèles de contenus	Attributs
string	Chaîne	O	O
Name	Nom XML	O	O
QName	Nom XML qualifié	O	O
NCName	Nom XML non qualifié	O	O

Temps :

Type	Signification	Modèles de contenus	Attributs
timeInstant	Date+Time	O	O
timeDuration	Progress time	O	O
recurringInstant		O	O
date	Date	O	O
time	Time	O	O

²⁵ Ce tableau est en ligne à l'adresse suivante : <http://www.mutu-xml.org/xml-base/shared/KEY-SCHEMA.html>

XML1.0 compatibilité :

Type	Signification	Modèles de contenus	Attributs
ID	XML1.0 ID	-	O
IDREF	XML1.0 IDREF	-	O
ENTITY	XML1.0 ENTITY	-	O
NOTATION	XML1.0 NOTATION	-	O
IDREFS	XML1.0 IDREFS	-	O
ENTITIES	XML 1.0 ENTITIES	-	O
NMTOKEN	XML 1.0 NMTOKEN	-	O
NMTOKENS	XML 1.0 NMTOKENS	-	O

Nombres :

Type	Signification	Modèles de contenus	Attributs
decimal	Décimal	O	O
boolean	Valeur booléenne	O	O
float	16bit floating point number	O	O
double	32bit floating point number	O	O
integer	Infinite accuracy integer	O	O
nonPositiveInteger	Infinite accuracy integer less than 0	O	O
negativeInteger	Infinite accuracy integer less than 0	O	O
long	64bit integer	O	O
int	32bit integer	O	O
short	16bit integer	O	O
byte	8bit integer	O	O
nonNegativeInteger	Infinite accuracy integer more than 0	O	O
positiveInteger	Infinite accuracy integer more than 1	O	O
unsignedLong	Unsigned 64bit integer	O	O
unsignedInt	Unsigned 32bit integer	O	O
unsignedShort	Unsigned 16bit integer	O	O
unsignedByte	Unsigned 8bit integer	O	O

Autres:

Type	Signification	Modèles de contenus	Attributs
binary	Binary	O	O
uriReference	URI	O	O
language	Language		