

ISSN - 2170-0656

CERIST NEWS

Bulletin d'information trimestriel

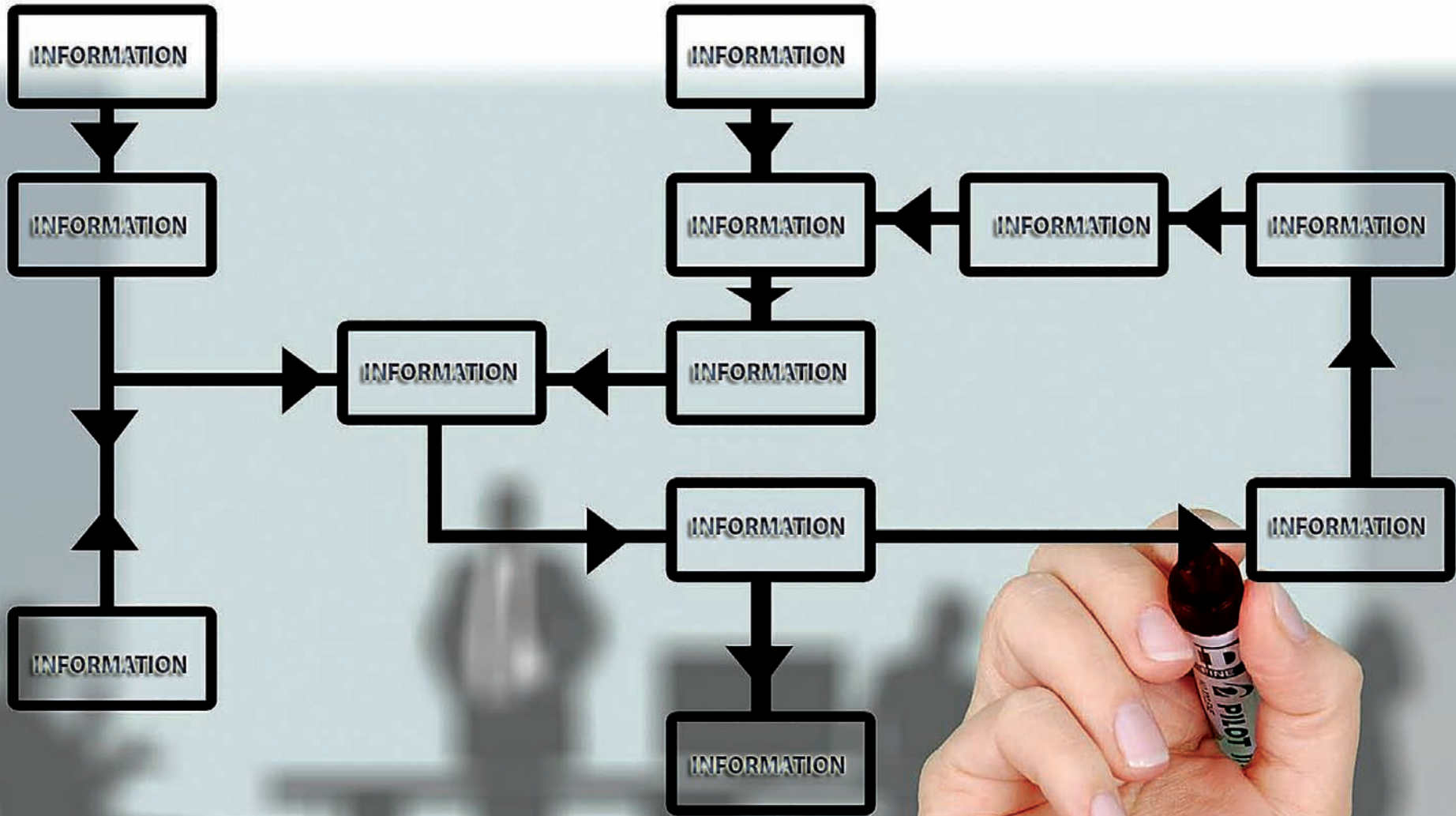
Dix septième numéro - Mars 2015

DOSSIER

**RECHERCHE D'INFORMATIONS
STRUCTURÉES**

CENTRE DE RECHERCHE
SUR L'INFORMATION
SCIENTIFIQUE ET TECHNIQUE





**Fatma Zohra****Bessai-Mechmache**

Maître de Recherche Classe B

Responsable Equipe

Recherche d'Information

Division TISI (Théorie et

Ingénierie des Systèmes

Informatiques)

La Recherche d'Information (RI) automatisée permet à un utilisateur de formuler un besoin d'information, à l'aide d'une requête, pour obtenir une réponse issue d'un ensemble de documents. L'idée est apparue dès la naissance des premiers ordinateurs. Depuis, avec la gigantesque augmentation des connaissances produites et conservées numériquement, elle est devenue un secteur stratégique pour beaucoup d'entreprises et elle a indiscutablement bouleversé, par l'intermédiaire de l'Internet, le rapport de tout un chacun avec l'information. Aussi, vu la croissance sans cesse de la quantité d'information et des utilisateurs qui tentent de l'exploiter, les enjeux de la recherche d'information sont devenus considérables.

Un des aspects importants auquel nous nous intéressons, est que l'information est de plus en plus formulée de façon structurée. Cette structure est formalisée explicitement dans le texte lui-même par des balises, à l'aide de langages de représentations spécifiques. Parmi ces langages, XML (eXtended Markup Language) est désormais devenu un standard universellement utilisé. Il permet l'échange et le stockage de l'information avec une certaine flexibilité qui l'a rendu extrêmement populaire. Les documents obtenus sont dits semi structurés, car ils sont en quelque sorte un intermédiaire entre les documents classiques et les bases de données. La problématique engendrée,

en RI, par ce type de document est liée à la nature de leur contenu. En effet, comme ces documents comportent de l'information (du texte) et des contraintes structurelles (des balises), ils ne peuvent pas être efficacement exploités par les techniques classiques de RI, qui considèrent le document comme un granule d'information indivisible. Aussi, dans un document XML toute partie (élément dans le jargon XML) du document peut être considérée comme réponse potentielle à la requête de l'utilisateur. La partie concernée peut être spécifiée directement dans la requête de l'utilisateur ou calculée automatiquement par le système de recherche d'information. Les requêtes dans les systèmes de Recherche d'Information Structurée de type XML peuvent en effet avoir deux formes : une forme « contenu seulement », la requête est dans ce cas composée que de mots clés et une forme combinant la structure et le contenu. Afin d'exploiter au mieux l'ensemble des informations disponibles, les méthodes existantes de recherche d'information doivent être adaptées ou de nouvelles méthodes doivent être proposées. C'est dans ce contexte de recherche d'information structurée que se situent nos travaux. Notre objectif est de proposer un modèle permettant de sélectionner automatiquement l'élément (ou l'ensemble d'éléments) du document qui répond le mieux au besoin de l'utilisateur formulé à travers une liste de mots clés.

Sommaire

5 Actualités

- Le CERIST leader du classement Webometrics au niveau national
- COMARES : Concours design graphique
- Rencontre internationale sur la télémédecine
- Participation du CERIST à la 16ème édition du Salon international du Futur Technologique SIFTECH

9 Dossier - RECHERCHE D'INFORMATIONS STRUCTURÉES

Document spécial de 20 pages : 09/29

Un dossier élaboré par : **Dr F.Z. Bessai-Mechmache**

Maître de Recherche B,
Équipe Recherche d'Information, DTISI

30 Les Conseils de DZ - CERT

- Bonnes pratiques pour le déploiement sécurisé du navigateur Google Chrome

35 Zoom sur un Projet

Reinforcement of ICT Regulations and ICTs for tackling Societal Challenges links in Europe and Mediterranean countries (ClusMED)

40 CERIST Recherche & Formation

- Rapports de recherche internes

42 CERIST Bases de Données Documentaires

- SNDL

Le CERIST dans le classement Webometrics

Selon l'édition de janvier 2015 du classement mondial des centres de recherche « The Ranking Web of Research Centers Webometrics », le site Web du Centre de Recherche en Information Scientifique et Technique (CERIST) maintient sa position de leader parmi les sites Web des centres de recherche au niveau national.

En effet, pour la troisième fois consécutive le site du CERIST occupe la première place parmi les sites web des centres de recherche au niveau national. Par ailleurs, il maintient aussi sa position de leader au niveau arabe et nord africain tandis qu'il se hisse à la 3ème place au niveau africain et à la 306ème place au niveau mondial.

Webometrics, dont le siège se trouve en Espagne, effectue tous les six mois un classement mondial des centres de recherche scientifiques et universités. Ce classement propose

une approche différente des autres classements mondiaux. En effet, celui-ci a pour but d'évaluer la présence des centres de recherche et universités sur internet. Il se base avant tout sur la quantité du contenu présent sur le net, et l'accessibilité du site de l'institution en question. La pertinence du classement Webometrics réside dans ces indicateurs qui traduisent les efforts d'investissement des centres de Recherche en matière de production scientifique, de valorisation des résultats de la recherche et de visibilité.

Algeria

ranking	World Rank	Institute	Size	Visibility	Rich Files	scholar
1	306	Centre de Recherche sur l'Information Scientifique et Technique	652	1107	2493	16
2	886	Centre de Développement des Energies Renouvelables	1782	2141	487	251
3	1258	Centre de Recherche Scientifique et Technique sur les Régions Arides	1912	2185	2169	422
4	1357	Centre National de Recherche en Anthropologie Sociale et Culturelle	1345	1907	1739	1102
5	3849	Centre de Développement des Technologies Avancées	4842	4371	3842	2411
6	5353	Algerian Petroleum Institute	5419	4931	5141	4897
7	5433	Centre de Recherche en Astronomie Astrophysique et Géophysique	4903	5597	4966	4556
8	6279	Centre de Recherche en Economie Appliquée pour le Développement	4842	5748	6811	4897
9	6327	Centre National de Recherche Appliquée en Génie Parasismique	5062	6116	6464	4897
10	7055	Centre de Recherche Scientifique et Technique en Analyses Physico Chimiques	4671	7168	6579	4897



COMARES Concours design graphique

M^{me} Ouazzoug née Benabderrahim Kahina, infographe au service des relations extérieures et de la communication au CERIST, a été la lauréate du concours design graphique : « Créez le logo de la COMARES » lancé par l'Agence Universitaire de la Francophonie (AUF).

En effet, Mme Ouazzoug s'est portée candidate au concours sur la conception d'un logo pour symboliser la Conférence Maghrébine des Responsables d'Établissements d'Enseignement Supérieur (COMARES), lancé par le bureau Maghreb de l'AUF en direction des étudiants des institutions tunisiennes, algériennes et marocaines, membres de l'Agence Universitaire de la Francophonie. Ces candidats devaient s'inspirer de la carte du « Maghreb sans frontières », image de marque au Bureau Maghreb de l'AUF. Lors de l'Assemblée Générale de la COMARES qui s'est tenue au Maroc, du 09 au 11 février 2015, Mme Ouazzoug a été honorée et sa proposition de logo a été retenue parmi une trentaine d'autres propositions.

Rencontre internationale sur la télémédecine

Placée sous le slogan « la télésanté est en mesure d'offrir une qualité de soins à l'ensemble des citoyens, là où ils se trouvent », une rencontre internationale sur la télémédecine a rassemblé les entreprises des secteurs technologique et des Sciences et médicale le 26 mars 2015 au CERIST.

Organisée par le Catel, un réseau français multidisciplinaire de compétences en télésanté, en partenariat avec le Centre de Recherche sur l'Information Scientifique et Technique (CERIST) et en présence d'un panel de participants étrangers, cette rencontre avait pour objectif de montrer l'intérêt de la télémédecine.

Les participants à cette conférence ont souligné la nécessité de normes fiables et interopérables pour fournir aux patients et aux professionnels de la santé les moyens d'utiliser des services de télémédecine comme la consultation à distance, les procédures

de diagnostic collaboratives ou des services d'information électronique médicale. Le secrétaire général de la société algérienne de télémédecine, le Pr Mohamed Zeroug, a

pourraient se faire suivre au moyen de la télémédecine depuis chez eux. Par ailleurs, M. Abdelkrim Meziane, Chef de division systèmes d'information et systèmes multimedia au CE-



plaidé que la télémédecine devrait être envisagée comme une «solution d'avenir». Il a aussi insisté sur les avantages que présente cette option qui s'appuie sur les Technologies de l'Information et de la Communication (NTIC) en appuyant ses propos par l'exemple des patients résidant dans des régions enclavées du pays et qui, en raison de l'éloignement et de l'absence de structures de santé,

RIST, a souligné que si les moyens existent, il reste à coordonner les actions pour faire de la télémédecine l'alternative qui offrira des soins de qualité. Il a fait savoir à ce sujet que des expériences avaient été tentées depuis une dizaine d'années par le Centre de développement des technologies avancées (CDTA) d'Alger avec un nombre de structures hospitalières, mais qu'elles sont demeurées «ponctuelles».

- • • Pour inscrire véritablement la rencontre dans l'esprit de son contenu, les organisateurs ont fait connecter une quinzaine de villes Françaises et une dizaine de pays Africains dont le Maroc et la Tunisie .

Participation du CERIST à la 16^{ème} édition du Salon International du Futur Technologique SIFTECH

Le CERIST a participé à la 16^{ème} édition du Salon International du Futur Technologique qui s'est déroulée du 24 au 26 février 2015 au palais de la culture Moufdi Zakaria à Alger.

Pour cette année, le thème principal était consacré à la sécurité des réseaux et des données informatiques "DATA & NETWORK SECURITY". Le CERIST a saisi cette occasion pour promouvoir et valoriser les activités du centre dans ce domaine. A cet effet, des représentants de l'équipe Sécurité des Données & Privacy de la division Sécurité Informatique du CERIST étaient présent pour expliquer la mission du centre relative à la prise en charge de la composante sécurité informatique afin d'assurer le bon fonctionnement des systèmes informatiques et leur protection contre les intrusions. Sans oublier l'assistance que le centre fournit aux institutions algériennes dans la prise en charge

des incidents de sécurité ou d'attaques. C'est ainsi qu'un, document anonyme représentant un audit de vulnérabilité a été présenté et une brochure de bonnes pratiques pour mieux se prémunir des risques inhérents à l'usage des Technologies de l'Information et de la Communication a été distribuée aux visiteurs du stand.



LE DOSSIER

Document spécial de 20 pages : 09/29

Un dossier élaboré par :

Dr F.Z. Bessai-Mechmache

Maître de Recherche B,

Équipe Recherche d'Information,

Division Théorie et Ingénierie des Systèmes informatiques - CERIST -

RECHERCHE D'INFORMATIONS STRUCTURÉES



1. Introduction

Le développement du document électronique et du Web ont vu émerger puis s'imposer des formats de données structurés, tel que le XML (eXtensible Markup Language), conçus à l'origine pour faciliter l'échange et la standardisation des données. Aujourd'hui, l'utilisation du langage XML est de plus en plus répandue dans de nombreuses applications dans le cadre de la représentation, du stockage et de l'échange de documents et de données, en particulier sur le Web. Par conséquent, le type des documents mis à la disposition des utilisateurs a évolué du simple document texte "plat" aux documents structurés ou semi structurés, permettant de représenter l'information sous une forme plus riche que le simple contenu et adaptée à des besoins spécifiques. En effet, les documents semi structurés permettent de représenter conjointement l'information textuelle et l'information de structure d'un document. La connaissance de la structure des documents est une ressource additionnelle qui devrait être exploitée pendant la recherche d'information afin de mieux exprimer un besoin d'information. Dans le contexte de la Recherche d'Information (RI), la question majeure soulevée par ce type de document concerne la manière de manipuler efficacement la structure et le contenu du document pour mieux répondre aux besoins de l'utilisateur. Ces besoins peuvent être formulés par le biais de requêtes

formées que de mots clé ou par des requêtes comportant des mots clés et des contraintes structurelles (des balises).

Nous assistons aujourd'hui à l'expansion du domaine de la recherche d'informations structurées (généralement de type XML), il s'est principalement développé ces dernières années notamment avec l'apparition de l'initiative **INEX¹** et de plusieurs Workshop dans différentes conférences de RI (**SIGIR²** par exemple).

Les approches proposées dans le cadre de la recherche d'information XML peuvent être classées en deux principales catégories : (I) **l'approche orientée données** utilise des techniques développées par la communauté des Bases de Données (BD), (II) **l'approche orientée documents** est prise en charge par la communauté recherche d'information.

Les **approches orientées données** s'intéressent davantage à la structure du document. Plusieurs langages ont été définis, LOREL, XML-QL, QUILT, XML-GL. Les **approches orientées documents** considèrent les documents XML comme une collection de documents textes comportant des balises (éléments) et des relations entre ces balises. Les balises sont utilisées comme moyen pour mieux identifier la pertinence d'une partie de document vis-à-vis d'une autre partie. La majorité des travaux ont en fait adapté les modèles de RI reconnus pour traiter les documents XML.

¹ INEX : Initiative for the evaluation of XML Retrieval

² SIGIR : Special Interest Group of Information Retrieval



• • • Si les approches orientées BD permettent de traiter efficacement la structure des documents XML, elles sont cependant limitées pour le traitement de la partie textuelle des documents. Les mots clés sont en effet traités de façon binaire (présent /absent), or il a été démontré en RI textuelle que la prise en compte des poids des mots-clés dans un document est primordiale, voire nécessaire. Ceci permet de mesurer un degré de pertinence d'un document (ou d'une partie de document) vis-à-vis d'une requête et donc de renvoyer à l'utilisateur une liste triée de résultats, comme le proposent les approches de RI. Les enjeux actuels pour le domaine de la recherche d'informations structurées, sont le développement de modèles de RI pour une prise en compte simultanée de la structure et du contenu, ainsi que la définition et l'évaluation de nouvelles problématiques spécifiques aux données semi structurées.

2. Les documents semi structurés

Les documents semi structurés sont des documents électroniques représentées selon un format structuré, c'est-à-dire un format qui utilise des balises pour décrire la structure logique des documents. Cette structure logique d'un document correspond à un type de structuration permettant de présenter de façon cohérente et logique le contenu des documents. Souvent, la structure logique d'un

document sera sa division en parties, chapitres, sections, etc., de même que certaines autres unités telles des notes de bas de page ou des références bibliographiques.

Décrire la structure d'un document, consiste à identifier et décrire chacun des éléments textuels – ou non textuels – qui le constituent. Ceci dit, cette description peut prendre plusieurs formes. En effet, nous distinguons, en général, deux types de structure : la structure physique et la structure logique.

Si l'on s'attache à retrouver **la structure physique** d'un document, on décrira sa mise en page, on définira les différentes zones de texte, leur agencement les unes par rapport aux autres, ainsi que l'ensemble de leurs caractéristiques typographiques : police, couleur, gras, italique, etc.

Si l'on s'intéresse maintenant à **la structure logique**, nous décrirons plutôt le rôle et la nature de chaque élément d'un document ainsi l'ensemble des liens hiérarchiques et/ou logiques qui les lient les uns aux autres (titre, auteur, introduction, résumé, mots clés, chapitre, sous chapitre, section, sous section, conclusion).

Nous noterons qu'une structure physique bien conçue aura pour principale qualité de rendre lisible la structure logique du document ce qui peut amener à confondre les deux. Cependant, il est important de bien faire la distinction entre ces deux types de structuration, car même si elles peuvent conduire toutes deux à un découpage du document équivalent, la manipulation des éléments du document et l'accès à chacun d'eux, sont très différents.

- • • En langage XML, la structure logique des documents est définie par des balises (tags ou labels) encadrant des portions d'informations. Ce type de document représente un compromis entre les données fortement structurées issues de la communauté bases de données (données relationnelles par exemple) et les données faiblement structurées issues des communautés document numérique et RI (documents plats, images etc...).
- Le tableau suivant est un récapitulatif des principales caractéristiques des documents plats, semi structurés et structurés dans le cadre de la recherche d'information.

	Documents plats	Documents semi structurés	Documents structurés
Contenu	texte seulement	texte + structure	structure + données
Besoin d'information	général		précis
L'unité recherchée	document	élément (doxel)	(tuple)
Demande	texte seulement	contenu et/ou structure	
Requête	mots clés	mots clés et/ou langages de requête structurés	langages de requête structurés
Interprétation	vague		stricte

Tableau 1. Résumé des caractéristiques des documents plats, semi structurés et structurés

Le format XML est aujourd'hui le format de représentation le plus utilisé. Par conséquent, dans la suite nous nous focalisons sur les documents semi structurés de type XML et les technologies qui s'y rapportent.

2.1 Notions de structure

Nous présentons ici les notions de base liées à la structure, s'appliquant aussi bien à des documents SGML que des documents XML:

- Une balise (ou tag ou label) est une suite de caractères encadrés par «<>» et «>>», comme par exemple <balisename>.
- Un élément est une unité sémantique identifiée, délimitée par des balises de début < b > et de fin < /b >, comme par exemple <balisename> texte </balisename>. Les éléments peuvent être imbriqués (mais ne doivent pas se recouvrir) formant ainsi une structure hiérarchique comme le montre l'exemple suivant:

```

<Ouvrage>
  <Titre > Recherche d'Information </Titre >
  <Résumé>Devant la masse croissante d'information ...
</Résumé>
...
  <Chapitre>
    <Titre chapitre> Indexation </titre chapitre>
    <Paragraphe> L'indexation est le processus ... </

```



● ● ●
 Paragraphe>
 </Chapitre>
 </Ouvrage>

Nombreuses sont les technologies reliées au format de données XML. Parmi les plus connus et les plus susceptibles d'aider à la recherche d'information, citons les DTD (Document Type Definition), les Schémas XML et les mécanismes de base pour adresser des éléments dans des documents XML (XPath), pour traiter les documents XML (DOM et SAX), pour présenter et transformer les contenus XML (XSL), les espaces de nom, XLink et XPointer pour la gestion des liens, etc.

3. Fondements scientifiques de la recherche d'informations structurées

3.1 L'unité d'information pertinente

Si les systèmes de RI classiques considèrent les documents dans leur globalité, les systèmes de Recherche d'Informations Structurées (RIS) cherchent plutôt à identifier des parties de documents les plus pertinentes à une requête. L'introduction d'une information structurée dans le codage des documents bouleverse complètement la notion

d'unité d'information. Cette unité était jusqu'à présent le document, ce n'est pas le cas aujourd'hui. Ce changement quant à la notion d'unité d'information est à l'origine de l'apparition de nouvelles problématiques spécifiques aux documents semi structurés. Certaines de ces nouvelles problématiques sont aujourd'hui particulièrement étudiées dans le cadre de la RI sur les documents semi structurés notamment au travers de l'initiative INEX. Ainsi, le concept de «granule ou d'unité d'information» renvoyée à l'utilisateur n'est plus un document dans sa totalité. Il correspond plutôt à un nœud (sous arbre) dans la représentation arborescente du document semi structuré dont la pertinence vis-à-vis d'une requête est calculée selon deux nouvelles notions : «exhaustivité» et «spécificité». On parle alors d'unité d'information exhaustive à une requête, si elle contient toutes les informations requises par la requête, et d'unité d'information spécifique à une requête, si tout son contenu concerne la requête. De plus, la partie du document renvoyée par le système se doit être auto explicative, c'est-à-dire de ne pas dépendre du reste du document pour être comprise. Par ailleurs, les résultats renvoyés doivent se suffire à eux-mêmes. L'application stricte du principe d'auto-explicativité conduirait à une perte bien trop importante de spécificité; donc, pour satisfaire le principe d'auto-explicativité et le principe de spécificité, le système de recherche doit trouver les sous arbres de taille minimale pertinents à la requête. Par conséquent, le principe de RI dans les documents semi structurés est défini comme suit : «Un système de recherche d'information devrait toujours renvoyer l'unité d'information la plus exhaustive et spécifique répondant à une

● ● ● requête», ce qui revient à trouver les sous arbres de taille minimale pertinents à la requête.
De par leur structure, l'utilisateur interrogeant des corpus de documents XML peut formuler deux types de requêtes, selon sa connaissance du corpus:

- Les requêtes portant sur le contenu, composées de simples mots clés (la structure n'est pas spécifiée) et cela dans le cas où l'utilisateur ne connaît pas précisément la structure des documents ou bien son besoin d'information est trop vague. Dans ce cas, c'est le système qui se charge de trouver l'unité d'information la plus pertinente et qui décide du granule d'information à retourner à l'utilisateur.
- Les requêtes portant sur le contenu et la structure, où l'utilisateur peut spécifier des besoins précis sur certains éléments de structure qu'il désire voir retourner en utilisant des contraintes structurelles dans la requête. Afin de répondre à ces nouvelles notions, il y a lieu d'adapter les anciens systèmes pour pouvoir répondre aux problématiques liées à la modification de l'unité d'information recherchée que cela soit au niveau de l'indexation, l'interrogation ou même la recherche et le tri des unités d'information renvoyées.

3.2 La problématique d'indexation

La problématique dans le cadre de l'indexation se situe principalement dans l'information structurelle. Dans le cas des documents plats, le processus d'indexation consiste à collecter et à pondérer les mots clés permettant un ciblage sémantique du document. Dans le cas des documents semi structurés, la dimension structurelle s'ajoute au contenu. Donc, la vision du processus d'indexation change puisqu'elle sera guidée par la structure du document, ce qui engendre les questions suivantes :

- Que doit-on indexer de la structure des documents ?
- Comment relier cette structure au contenu même du document ?
- Comment pondérer les termes d'indexation, c'est-à-dire, comment évaluer l'importance d'un terme au sein de l'élément, du document et de la collection ?

L'objectif du processus d'indexation dans ce cas, est alors de tenir compte de la structure et du contenu pour permettre d'effectuer des recherches portant sur le contenu et/ou la structure des documents.

3.3 La problématique d'interrogation

Il s'agit de permettre à l'utilisateur d'exprimer ces besoins d'information diversifiés (concernant le contenu des documents et/ou la structure) de

- -
 -
- manière simple via un langage de requête (langage naturel, mots clés, ...), et de sélectionner les unités d'informations jugées pertinentes à ces besoins.

L'évaluation d'une information vis-à-vis d'une requête se complique et implique d'autres questions dans le cadre des documents XML, principalement en ce qui concerne la structure :

- les requêtes orientées contenu, imposent au SRI de décider la granularité appropriée de l'information à retourner.
- les requêtes orientées contenu et structure : l'utilisateur spécifie les éléments à retourner et précise la granularité de l'information pertinente, alors la dimension de la spécificité n'a plus réellement de sens. La pertinence des informations structurelles doit pouvoir être évaluée, et l'arbre de la requête et l'arbre du document doivent pouvoir être comparés de façon non stricte.

Afin de répondre à ces problématiques soulevées par la notion de structure, les systèmes de recherche d'information traditionnels doivent être adaptés ou bien de nouvelles méthodes doivent être proposées pour l'indexation et l'interrogation des documents semi structurés.

4. Techniques d'indexation des documents semi structurés

Une façon simple d'indexer des documents XML est de les considérer comme des fichiers plats, et le processus d'indexation dans ce cas-là est similaire à celui utilisé en RI traditionnelle, c'est-à-dire qu'il consiste à identifier et à extraire les termes importants du contenu textuel des documents. Cependant, aucune recherche sur la structure n'est plus possible, et l'unité d'information renvoyée à l'utilisateur est le document dans sa totalité.

Un schéma d'indexation de documents XML doit couvrir certains aspects, comme de permettre la recherche par mots clés, le traitement des expressions de chemin sur la structure XML et d'autoriser le traitement de prédicats vagues et précis sur le contenu de documents XML.

Nous allons décrire ci-dessous l'indexation de l'information de contenu (textuelle) et l'indexation de l'information structurelle tout en citant les principales techniques citées dans la littérature.

Les exemples illustrant l'indexation des documents semi structurés porteront sur le document XML de la figure 1 suivante :

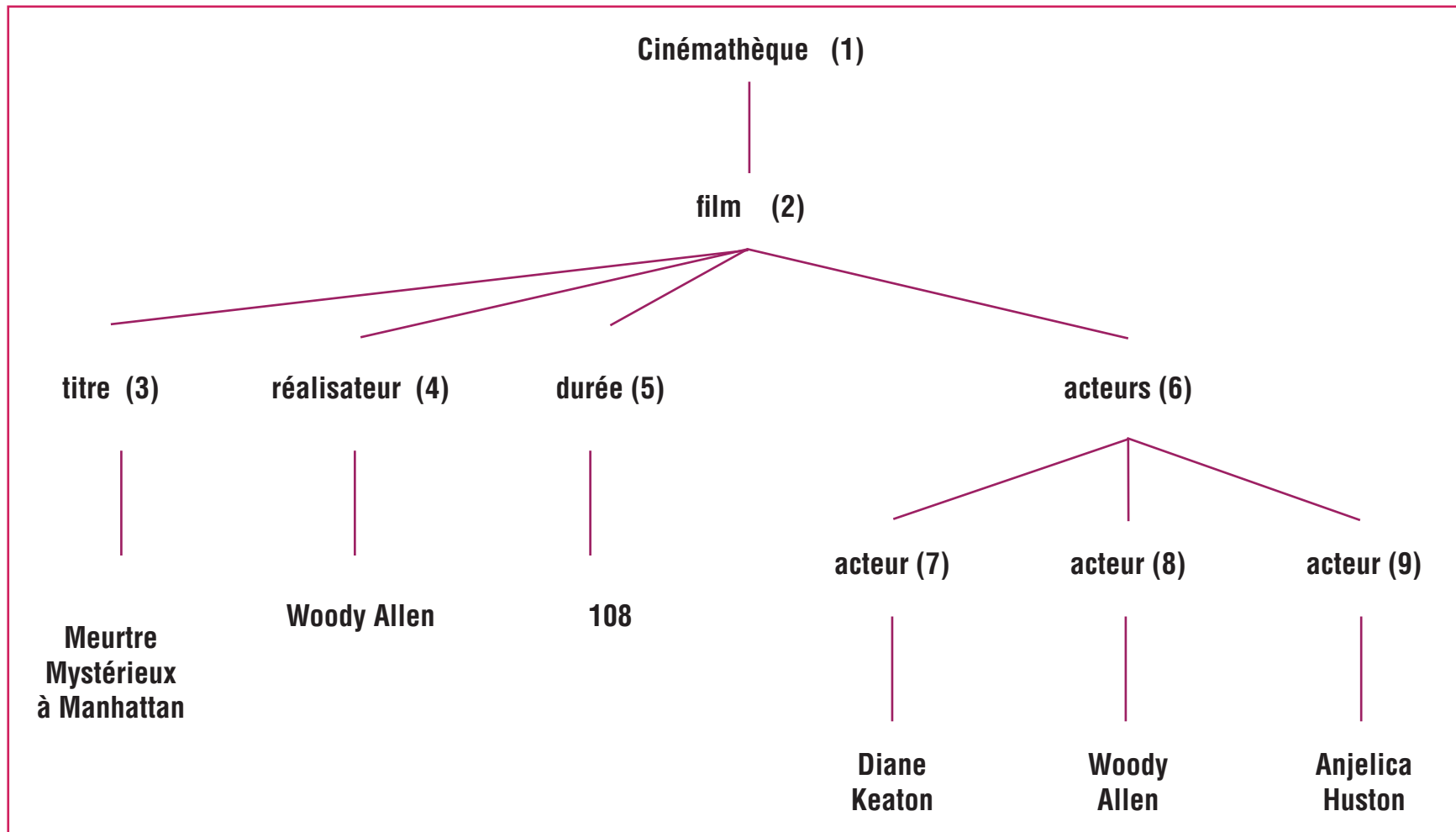


Figure 1. Exemple de document XML représenté sous forme arborescente



4.1 Indexation et pondération de l'information textuelle

4.1.1 Indexation de l'information textuelle

Le processus d'indexation, consiste comme dans le cas de la recherche d'information traditionnelle à extraire les termes importants des documents, mais pour la recherche d'informations structurées, le problème de la portée des termes d'indexation se pose et la question est: comment rattacher les termes à l'information structurelle? Doit-on chercher à agréger le contenu des nœuds ou au contraire à indexer tous les contenus des nœuds séparément? Ces deux solutions correspondent aux approches d'indexation dites des sous arbres imbriqués et des unités disjointes.

4.1.1.1 Approche des sous arbres imbriqués

Dans cette approche les termes d'un nœud feuille sont rattachés à tous ses nœuds ancêtres dans le document. Cependant, comme les documents XML possèdent une structure hiérarchique, les nœuds de l'index sont imbriqués les uns dans les autres et l'index contient de nombreuses informations redondantes.

4.1.1.2 Approche des unités disjointes

Dans cette approche le document XML est décomposé en unités disjointes, de manière que le texte de chaque nœud de l'index est l'union d'une ou plus de ces parties disjointes. Les termes des nœuds feuilles

sont uniquement reliés au nœud parent qui les contient.

Par exemple, dans l'approche des sous arbres imbriqués le terme 'Meurtre' du document XML de la figure 1 est relié aux nœuds /cinémathèque/film/titre, /cinémathèque/film et /titre. Et dans l'approche des unités disjointes, le terme 'Meurtre' est uniquement relié au nœud /cinémathèque/film/titre.

4.2 Indexation de l'information structurelle

La structure peut être indexée selon des granularités variées et donc ce n'est pas forcément toute l'information structurelle qui est utilisée dans le processus d'indexation. Parmi les approches proposées dans la littérature, nous distinguons trois types d'approches pour l'indexation de l'information structurelle: indexation basée sur des champs, indexation basée sur des chemins, et enfin indexation basée sur des arbres.

4.2.1 Indexation basée sur des champs

Dans cette approche, tout terme d'indexation est associé au nom de la balise dans laquelle il apparaît. Cette méthode a l'avantage de permettre de savoir dans quel contexte l'information textuelle est énoncée.

Par exemple, si nous appliquons cette méthode aux termes 'Meurtre', 'Allen' et '108' nous aurons les résultats figurant dans le tableau 2 qui illustre aussi la notion de propagation citée dans l'indexation de l'information textuelle:

Terme	Sans propagation	Avec propagation
Meurtre	Titre	cinémathèque, film, titre
Allen	réalisateur, acteur	cinémathèque, réalisateur, film, acteurs, acteur
108	Durée	cinémathèque, film, durée

Tableau 2. Exemple d'indexation basée sur des champs

4.2.2 Indexation basée sur des chemins

Il s'agit dans cette méthode d'indiquer le chemin d'accès logique pour tout terme d'indexation. Cette méthode permet de retrouver rapidement des documents ayant des valeurs connues pour certains éléments ou attributs et facilite la navigation de façon à résoudre efficacement des expressions XPath. Par conséquent, l'indexation produit des index de chemins, c'est-à-dire des index donnant pour chaque valeur répertoriée d'un chemin de balises (de type Xpath) la liste des documents répondants contenant un élément atteignable par ce chemin et ayant cette valeur.

Voici un exemple illustrant l'indexation basée sur des chemins :

Terme	Indexation
Meurtre	/cinémathèque/film/titre
Allen	/cinémathèque/film/réalisateur /cinémathèque/film/acteurs/acteur
108	/cinémathèque/film/durée

Tableau 3. Exemple d'indexation basée sur des chemins

4.2.3 Indexation basée sur des arbres

Un identifiant unique (UID) est associé à chaque nœud de l'arbre représentant le document XML (voir figure1) de sorte à pouvoir reconstruire la structure arborescente du document, pour la navigation et le traitement d'expression XPath. Les termes sont attribués à cet identifiant, ce qui permet de repérer précisément l'endroit où ces termes apparaissent et de retrouver les relations hiérarchiques entre les éléments.

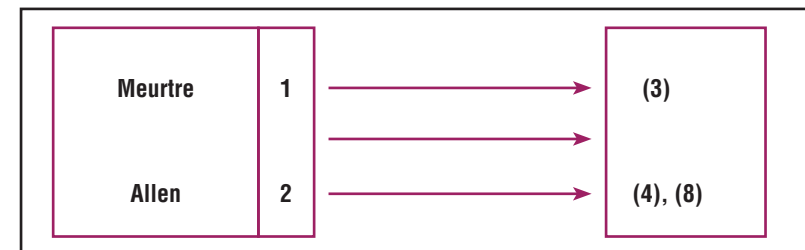


Figure 2. Exemple d'indexation basée sur des arbres



• • • 5. Langages de requêtes

Différents langages de requêtes sont proposés dans la littérature pour l'interrogation de documents XML. Ils offrent à l'utilisateur la possibilité d'exprimer des besoins diversifiés concernant le contenu et/ou la structure. Parmi les premiers langages spécifiés pour l'interrogation des documents semi structurés, nous pouvons citer LOREL, XML-QL ou XQL. Ce dernier a été développé par Microsoft et qui au lieu d'utiliser des canevas XML, il propose d'étendre les URL pour interroger des collections de documents XML avec des expressions Xpath. Le langage QUILT a été développé pour interroger des documents et des bases de données. XML-GL est un langage de requêtes graphique s'utilisant sur des graphes XML. Il présente le principal avantage d'être très ergonomique. XQuery est le langage de requêtes pour XML proposé par le W3C. XQuery tire très fortement ses constructions et caractéristiques du langage QUILT, lui-même dérivé de Xpath, XQL, XML-QL, et LOREL. XQuery peut être perçu comme

un sur ensemble du langage SQL. Les fonctionnalités de SQL sur les tables sont étendues pour supporter des opérations similaires sur des collections d'arbres. Ces extensions ont conduit à intégrer les fonctions suivantes : projection d'arbres sur des sous-arbres, sélection d'arbres et de sous-arbres en utilisant des prédicats sur les valeurs des feuilles, combinaison des arbres extraits de collection en utilisant des jointures d'arbres, réordonnancement des arbres, imbrication des requêtes, calculs d'agrégats, etc. XQuery supporte des fonctions orientées RI : en particulier, un prédicat 'contains' est intégré pour la recherche par mots-clés. Pour faciliter la recherche dans des structures mal connues, XQuery permet enfin d'exprimer des chemins indéterminés ou partiellement connus, tout comme Xpath. Nous trouverons dans le tableau suivant un exemple de requête XQuery.

```
FOR $p IN fn : distinct-values (fn :doc (« bibliotheque.xml »)// editeur)

LET $a := fn :doc (« bibliothèque.xml »)//livre[editeur =$p]/prix

WHERE fn : count ($a) >100

RETURN <grandediteur>
```

Tableau 4.
Exemple de requête XQuery : lister les noms des éditeurs qui ont publié plus de 100 livres et la moyenne des prix des livres édités pour chacun

- • • Une comparaison des langages de requêtes les plus importants, sur la base des principales caractéristiques est faite dans, elle est donnée dans le tableau suivant :

	LOREL	XML-QL	XQL	XML-GL	XQuery
Navigation	Oui	Oui	Oui	Oui	Oui
Support Xpath	Limité	Limité	Oui	/	Oui
Sélection	Oui	Oui	Oui	Oui	Oui
Jointure	Oui	Oui	Non	Oui	Oui
Tri	Oui	Oui	Non	Oui	Oui
Construction	Non	Oui	Non	Oui	Oui
Mot-clé	Non	Non	Non	Non	Oui
Fonction	Oui	Oui	Non	/	Oui
Imbrication	Oui	Oui	Oui	Oui	Oui
Agrégation	Oui	proposée	Non	/	Oui

Tableau 5. Comparaison de différents langages de requêtes pour XML

Comme nous pouvons le voir, de très nombreux langages de requêtes ont été proposés dans la littérature. La plupart sont très puissants, mais leur syntaxe, souvent dérivée de SQL, est difficilement accessible pour

les novices. Des interfaces adaptées peuvent être associées, mais leur feraient perdre de leur puissance. De plus, ces langages proposent le prédicat 'contains' pour exprimer des conditions sur le contenu des éléments. Cependant, ce type de prédicat ne permet pas de mesurer la similarité entre les unités d'informations et les conditions de contenu (requête). Pour intégrer ce dernier aspect et permettre une recherche orientée sur la pertinence (c'est-à-dire avec une pondération des résultats), des langages issues de la communauté recherche d'information ont apparu tel que le langage NEXI qui a été défini pour répondre aux besoins de la campagne d'évaluation INEX, le langage du modèle XFIRM et enfin le W3C a proposé un Working Draft, qui a pour but d'étendre les caractéristiques de recherche de XQuery à la recherche plein texte. Le langage TexQuery en est une application.

6. Modèles de recherche d'informations structurées

Les systèmes de recherche d'information traditionnels (prenant en compte les documents plats non structurés) ne prennent pas en considération l'information structurelle. Cette dernière apporte une information supplémentaire à la définition du besoin de l'utilisateur, elle est

• • • accessible directement à travers les nouveaux formats de données. Donc, il est indispensable d'adapter les systèmes de recherche traditionnels principalement pour les deux raisons suivantes :

1- Les systèmes de recherche d'information traditionnels sont conçus pour des documents linéaires plats et donc ne prennent pas en compte l'information structurelle. De plus, les corpus de documents sont composés de documents hétérogènes aussi bien dans leurs formes que dans leurs contenus. Paradoxalement, les systèmes de recherche traditionnels ont été conçus surtout pour des documents homogènes.

2- L'apparition des documents semi structurés a entraîné l'apparition de nouvelles problématiques spécifiques de recherche d'information comme l'unité d'information pertinente. La structure est une nouvelle source d'information non négligeable, dans il est même affirmé que «Ignorer la structure du document revient à ignorer sa sémantique », elle facilite la représentation des documents, ainsi que leur interprétation et leur exploitation, afin d'apporter des réponses aux nouveaux besoins en information. Par exemple, l'utilisateur peut choisir la granularité d'information à retourner en utilisant des contraintes structurelles sur les éléments à restituer.

Les différentes approches développées tendent d'une part de séparer la structure d'un document de son apparence et d'autre part de lier la structure logique et le contenu. L'accès simultané à la structure des documents et à leur contenu permet d'envisager de nouveaux modes d'exploitation de l'information textuelle. Celle-ci nécessite la création de nouveaux outils et de nouveaux modèles capables d'exploiter de telles données.

Nous présentons dans la suite les principales extensions faites sur les modèles de RI traditionnels pour les adapter à la recherche dans les documents semi structurés ainsi que d'autres approches spécifiques pour le traitement des conditions de structure.

6.1 Modèle booléen étendu

Des auteurs se sont basés sur des approches portant sur l'extension des modèles booléens aux documents structurés. Leur travail se basant essentiellement sur la définition de l'unité d'information renvoyée à l'utilisateur, ne considère que les requêtes orientées contenu. L'unité d'information est définie comme étant tout élément contenant au moins un élément feuille. Pour calculer le score d'un élément vis-à-vis d'une requête, les auteurs propagent des scores, depuis les feuilles de l'arbre

- documentaire jusqu'à l'élément considéré. Pour ce faire, chaque élément feuille e_j ($j=1, \dots, N$) est représenté par un vecteur:

$$F(e_j) = (w_{t_1}^{e_j}, w_{t_2}^{e_j}, \dots, w_{t_n}^{e_j})$$

Où n représente le nombre de termes indexés dans la collection et $w_{t_i}^{e_j}$ représente le poids d'un terme t_i dans un élément feuille qui est donné par :

$$w_{t_i}^{e_j} = \frac{f(t_i, e_j)}{\sum_{j=1}^N \sum_{k=1}^n f(t_k, e_j)} \cdot \log \frac{N_d}{df(t_i)}$$

Où $tf(t_k, e_j)$ est la fréquence du terme t_k dans l'élément e_j , N_d est le nombre de documents dans la collection et $df(t_i)$ est le nombre de documents contenant le terme t_i .

La similarité entre un élément feuille e_j (représenté par le vecteur $F(e_j)$) et une requête q (représentée par un vecteur dont les composantes sont 1 si un terme de la collection apparaît dans la requête, 0 sinon) est calculée comme suit:

$$sim(F(e_j), q) = \frac{F(e_j) \cdot q}{|F(e_j)| \cdot |q|}$$

Le score final d'un élément est calculé de manière récursive car il prend en compte le score des sous éléments:

$$RSV(q, e) = 1 - \frac{1}{\sqrt[p]{|enf(e)|}} \sum_{e' \in enf(e)} (1 - RSV(q, e'))^p$$

Où p est généralement choisi avec une valeur égale à 2, $enf(e)$ représente les sous éléments de e . Lorsqu'un élément e est une feuille :

$$RSV(q, e) = sim(F(e_j), q)$$

6.2 Modèle vectoriel étendu

Dans cette approche, les éléments sont représentés par vecteurs de termes pondérés. Généralement, les approches développées utilisant ce modèle utilisent l'indexation basée sur des arbres, elles propagent les termes des nœuds feuilles dans l'arbre du document.

- Il se base sur le calcul de la similarité de chaque élément à la requête, puis les éléments sont renvoyés selon leur ordre de similarité.
- La mesure suivante permet de calculer la similarité entre un nœud n et une requête $q = \{q_1, q_2, \dots, q_T\}$:

$$Sim(q, n) = \alpha(T) \cos(q, n) + \sum_{k=1}^s \frac{\cosm(q, n_k)}{\beta^{k-1}}$$

Où $\alpha(T)$ est le facteur permettant de prendre en compte le type de nœud, s le nombre d'enfants n_k de n , β le paramètre permettant d'assurer que le nombre d'enfants n'introduit pas un biais dans la formule.

La fonction $\cosm(q, n)$ est définie comme suit :

$$\cosm(q, n) = \sum_{i=1}^T \frac{w_i^q \cdot w_i^n}{|n|}$$

Où w_i^q et w_i^n les poids du terme t_i respectivement dans la requête q et dans le nœud n , $|n|$ est le nombre de termes dans le nœud n .

Le modèle vectoriel est adopté dans plusieurs approches, nous citons Zargayouna, qui définit un modèle d'indexation de documents XML implémentant le modèle vectoriel, dans lequel le document n'est pas considéré comme un vecteur, mais comme une matrice de termes et de modèles de balise.

6.3 Autres modèles

D'autres approches ont également été proposées. Nous citons par exemple celle basée sur la théorie de l'évidence de Dempster-Schafer de Lalmas. Dans cette approche, pour tout document et toute requête, il est possible de définir deux mesures : la croyance et l'incertitude. Un opérateur d'agrégation permet de combiner la croyance de ces sous structures pour calculer la pertinence d'un document. Nous citons aussi l'approche de Piwowarski qui s'appuie sur le formalisme des Réseaux Bayésiens et qui calcule des probabilités conditionnelles en fonction de la question et la recherche correspond à l'inférence dans le réseau. Dans le modèle XFIRM de Sauvagnat, le traitement des requêtes orientées contenu est effectué en deux étapes: une première étape consiste à évaluer la similarité des nœuds feuilles de l'index à la requête (on parle alors de calcul des poids des nœuds feuilles) et une seconde étape consiste à rechercher les sous arbres pertinents. La pertinence des sous arbres est évaluée en effectuant la propagation des poids des feuilles dans l'arbre du document.

Le modèle MPRIX, proposé par Bessai-Mechmache et al., est basé sur la théorie des possibilités, et plus particulièrement sur les réseaux possibilistes. La logique possibiliste offre un cadre théorique pour mieux prendre en compte l'ignorance et l'incertaine inhérente à la recherche d'information de manière générale. On trouve cette incertitude dans la notion de pertinence d'un document vis-à-vis d'une requête, dans le degré de représentativité d'un terme dans un document ou une partie

• • • de documents et dans l'identification de la partie pertinente répondant à la requête. Ce modèle permet d'interpréter la pertinence dans un cadre possibiliste. Ce cadre est plus à même de prendre en compte l'ignorance partielle qui peut affecter les informations utilisées dans les différents calculs en particulier, le poids des termes dans les éléments, la pertinence d'un élément. Plus précisément, ce modèle sépare les raisons de sélectionner un élément pertinent de celles de le rejeter, en utilisant deux mesures : la nécessité et la possibilité. La possibilité de pertinence tente d'éliminer les éléments non pertinents. La nécessité de pertinence met l'accent (le «focus») sur les éléments qui semblent très pertinents. Dans ce cadre, afin d'identifier la partie pertinente qui répond à la requête, contrairement aux approches proposées dans la littérature, le modèle MPRIX fournit le premier cadre unifié pour la recherche des éléments pertinents et leur agrégation dans un même résultat. Cette agrégation permet de réduire les efforts de l'utilisateur pour localiser les informations qu'il recherche et de renseigner l'utilisateur sur la diversité des informations du corpus en rapport avec son besoin en information.

7. Evaluation des Systèmes de Recherche d'Informations Structurées

L'évaluation d'un système de recherche se fait dans une collection de tests. Nous citons la seule compagnie, pour ce jour, qui permet de tester les systèmes de recherche d'information structurée : INEX (INitiative for the Evaluation of Xml retrieval), elle a été créée en 2002, son objectif est de permettre une évaluation et une comparaison aussi rigoureuse que possible des systèmes de recherche d'information structurée dans des collections XML orientées documents. Elle fournit un ensemble de documents, des requêtes proposées par les participants (experts dans le domaine) et des jugements de pertinence.

7.1 Collection INEX

Jusqu'en 2005, la collection INEX était composée d'articles scientifiques provenant de la IEEE Computer Society, balisés au format XML. La collection, d'environ 700 Mo, a atteint plus de 17 000 articles, publiés de 1995 à 2004, et provenant de 18 magazines ou revues différents. Un article moyen est composé d'environ 1500 éléments.

En 2006, la collection s'est renouvelée d'environ 4,5 Go, elle est composée de plus de 650 000 articles de l'encyclopédie Wikipedia.

7.2 Requêtes

Les requêtes (Topics) sont créées par les différents participants et doivent être représentatives des demandes de l'utilisateur moyen sur la collection. On distingue deux principaux types de requêtes :



■ Les CO (Content Only) : ces requêtes sont exprimées en langage naturel. Les mots-clés de la requête peuvent être groupés sous forme d'expressions et précédés par les opérateurs '+' (signifiant que le terme est obligatoire) ou '-' (signifiant que le terme ne doit pas apparaître dans les éléments renvoyés à l'utilisateur).

■ Les CAS (Content And Structure) : ces requêtes contiennent de plus des contraintes sur la structure des documents.

7.3 Tâches

La tâche principale d'INEX est la tâche de recherche ad-hoc, elle est considérée comme une simulation de l'utilisation d'une bibliothèque, où un ensemble statique de documents est interrogé avec des besoins utilisateurs, c'est à dire des requêtes. Les requêtes peuvent contenir à la fois des conditions structurelles ou de contenu, et en réponse à une requête, des éléments (et non forcément des documents) peuvent être retrouvés à partir de la bibliothèque. La tâche ad-hoc se divise en trois sous tâches:

- **Tâche CO** : La tâche CO (Content Only Task) a pour but de répondre avec des éléments/documents XML à des requêtes utilisateur orientées contenu (requête CO), c'est à dire des requêtes

contenant des mots-clés simples. Aucune indication de structure dans la requête ne peut aider les SRI à déterminer la granularité de l'information à renvoyer.

Un exemple de requête CO est présenté dans la figure suivante :

```
<inex topic topic id="98" query type="CO">
<title> "Information Exchange" +"XML" "Information Integra-
tion"</title>
<description> How to use XML to solve the information ex-
change (information integration) problem, especially in heteroge-
neous data sources ? </description>
<narrative> Relevant documents/components must talk about
techniques of using XML to solve information exchange (infor-
mation integration) among heterogeneous data sources where
the structures of participating data sources are different although
they might use the same ontologies about the same content. </
narrative>
<keywords> Information exchange, XML, information integra-
tion, heterogeneous data sources </keywords>
</inex topic>
```

Figure 3. Exemple de requête CO, issue du jeu de test 2003



- **Tâche SCAS** : La tâche SCAS (Strict Content And Structure Task) consiste à répondre avec des éléments/documents XML aux requêtes CAS de manière stricte, c'est à dire en respectant toutes les conditions sur la structure et le contenu énoncées dans les requêtes. Le champ Title des requêtes de la tâche SCAS est basé sur une syntaxe XPath. Un exemple de requête CAS pour la tâche SCAS est présenté dans la figure suivante:

```
<inex topic topic id="64" query type="CAS">
<title> //article[about(./,'hollerith')] // sec[about(./, 'DEHOMAG')] </title>
<description> In articles discussing Herman Hollerith find sections
that mention DEHOMAG </description>
<narrative> Relevant sections deal with DEHOMAG (Deutsche Hol-
lerith Maschinen Gesellschaft) in documents that discuss work or life
of Herman Hollerith </narrative>
<keywords> Hollerith, DEHOMAG, Deutsche Hollerith Maschinen
Gesellschaft </keywords>
</inex topic>
```

Figure 4. Exemple de requête CAS, issue du jeu de test 2003

- **Tâche VCAS** : La tâche VCAS (Vague Content And Structure Task), utilise elle aussi des requêtes CAS, mais pour lesquelles les participants peuvent répondre de manière vague, c'est à dire avec des éléments/documents qui satis-

font globalement les requêtes. Le champ «Title» des requêtes de la tâche SCAS est basé sur le langage de requêtes NEXI, l'extension de XPath utilisée en 2003 pour les requêtes CAS étant considérée comme trop complexe : 63% des requêtes exprimées par les participants (experts en RI) contenaient des erreurs de syntaxe. Un exemple de requête CAS pour la tâche VCAS est présenté dans la figure suivante :

```
<inex topic topic id="149" query type="CAS">
<title> //article[about(./(abs|kwd),"genetic algorithm")]// bdy//sec[about(.,
simulated annealing)] </title>
<description>Find sections about simulated annealing in article that men-
tion genetic algorithms. </description>
<narrative> I have come across the Constrained Shortest Path problem in
connection with a route planing program. I have become aware of a technique
called Simulated Annealing known from combinatorial optimization for heu-
ristic solutions to NP-hard problems that I wish to use in the route plaining pro-
gram. I have noticed a tendency for authors that mention SA in combination
with Genetic Algorithms so I expect the keyword 'genetic' to appear in relevant
articles. For the section to be relevant it has to discuss usage of Simulating
Annealing or refer to results relevant to the technique.</narrative>
<keywords> genetic, simulated annealing, optimization </keywords>
</inex topic>
```

Figure 5. Exemple de requête CAS, issue du jeu de test 2004



● ● ● Depuis 2004, cinq nouvelles tâches ont été proposées aux participants :

- **La tâche de «relevance feedback»** : a pour but d'expérimenter l'utilisation du contenu et de la structure comme informations de base pour la formulation d'une nouvelle requête;
- **La tâche de «langage naturel»** : dans laquelle les utilisateurs formulent leurs requêtes en langage naturel, et donc sans avoir besoin d'apprendre un langage complexe;
- **La tâche «interactive»** : a pour but d'étudier le comportement des utilisateurs face à des corpus XML et donc de cerner au mieux leurs besoins;
- **La tâche « hétérogène »** : propose aux participants de nouvelles collections, afin de développer des approches indépendantes de la structure du document XML.
- **La tâche multimédia**, dédiée pour le moment à la recherche de contenus images au sein de documents XML.

7.4 Jugements de pertinence

L'évaluation de la pertinence des SRI passe par une première phase de validation des documents renvoyés par les SRI. Chaque élément/document est jugé à la main (par les participants) pour chaque requête, en utilisant le système de jugement en ligne. Les participants surlignent dans les documents les parties qui ne contiennent que de l'information pertinente (indépendamment de la structure des documents). Le juge-

ment de pertinence (niveaux d'exhaustivité et de spécificité) de chaque élément est ensuite déduit par le système.

Le jugement de pertinence des éléments, restitués par un système de recherche, vis-à-vis des requêtes est établi par les deux dimensions suivantes : **l'exhaustivité et la spécificité.**

■ **L'exhaustivité** : décrit à quel point l'élément traite le sujet de la requête, elle est évaluée par un juge humain sur une échelle à quatre niveaux:

- Non exhaustif: aucune correspondance entre la requête et l'élément.
- Marginalement exhaustif : uniquement certains aspects du sujet de la requête sont traités par l'élément.
- Assez exhaustif : les aspects du sujet de la requête sont en grande partie traités par l'élément.
- Très exhaustif : tous les aspects du sujet de la requête sont traités par l'élément.

■ **La spécificité** : décrit à quel point l'élément se focalise sur le sujet de la requête, elle est considérée depuis la campagne INEX 2005 comme une dimension continue représentée par une valeur réelle dans l'intervalle [0,1]. Elle peut être calculée comme le rapport entre la longueur du texte pertinent et la longueur totale du texte de l'élément restitué en réponse à la requête.



- Ces deux dimensions de la pertinence reflètent la pertinence d'un élément par rapport à ses descendants. En effet, un élément peut être plus exhaustif que chacun de ses descendants pris séparément. De même, des éléments peuvent être plus spécifiques que leurs parents.

Le degré de pertinence d'un élément/document renvoyé par les systèmes de RI entrés en campagne pour chaque requête, est jugé (à la main) par les participants en utilisant un système de jugement en ligne. Le degré de pertinence est donné par la paire $(e, s)_{e \in E, s \in S}$, avec $E.S = \{(0, 0), (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\}$ est l'ensemble des 10 valeurs possibles sur l'échelle combinant les deux dimensions d'exhaustivité et de spécificité (les autres combinaisons ne sont pas possibles, par exemple: un élément qui n'est pas exhaustif, ne peut être spécifique, et inversement).

7.5 Mesures d'évaluation

L'évaluation de la performance des systèmes de recherche d'information structurée a commencé par l'utilisation des méthodes basées sur les mesures de rappel et précision, en cherchant à prendre en compte la structure des documents XML. Ces mesures présentent cependant un inconvénient majeur: elles ne prennent pas en compte l'imbrication (overlap) des éléments, inhérente aux documents XML, et évaluent le retour d'un élément pertinent sans prendre en compte le fait qu'il soit peut-être déjà vu entièrement ou en partie par l'utilisateur. Par exemple, un système A renvoyant un élément pertinent et aussi un de ses élé-

ments fils pertinents obtient les mêmes performances qu'un système B renvoyant deux éléments pertinents non imbriqués.

Dans INEX 2005, la mesure XCG (XML Cumulated Gain) a été considérée comme la mesure d'évaluation officielle. La motivation derrière XCG est de pallier au problème des éléments imbriqués mais elle est loin d'être adoptée de manière définitive, une réflexion continue ayant lieu sur les mesures utilisées et la notion de pertinence.

8. Conclusion

Les systèmes de recherche d'information structurée sont principalement conçus pour la prise en compte des documents semi-structurés, ils doivent permettre de considérer les termes présents dans un document en fonction de leurs contextes d'apparition, c'est-à-dire l'endroit où ils apparaissent dans la structure du document.

La présence de la structure entraîne deux nouvelles exigences, d'une part, celle de retourner des unités d'information plus petites au lieu du document entier, et d'autre part, celle de permettre la prise en compte de la structure dans la requête pour des réponses plus précises. Par conséquent, la structure offre un apport sémantique non négligeable, puisque l'utilisateur peut préciser l'unité d'information qu'il recherche en incluant des contraintes structurelles dans sa requête.



La plupart des modèles de recherche d'information structurée proposés dans la littérature sont des adaptations des modèles traditionnels. Des extensions sont alors apportées aux modèles classiques de recherche d'information afin de prendre en compte l'impact de la présence de la structure dans les documents, et pour remédier aux différentes problématiques soulevées par cette dernière, comme l'unité d'information adéquate à retourner à une requête et la prise en compte de l'information structurelle au niveau de l'indexation d'un document. Par exemple, en ce qui concerne la pondération des termes d'indexation, les lois utilisées pour la recherche d'information classique ne sont pas forcément valides dans la recherche d'information structurée, comme la loi de Zipf. Alors de nouvelles approches sont développées, comme la loi de pondération **tf-idf (Term Frequency – Inverse Tag and Document Frequency)** qui permet de calculer la force discriminatoire d'un terme pour une balise relative à un document donné. L'information structurelle subit rarement un traitement spécifique par ces modèles adaptés, généralement ces derniers utilisent un filtre

afin de vérifier les contraintes structurelles des requêtes. Et l'interrogation des corpus hétérogènes (c'est-à-dire comportant plusieurs structures différentes) pose un problème ouvert jusqu'à présent, les approches proposées se base principalement sur une correspondance syntaxique entre les arbres des requêtes et ceux des documents, or dans la réalité, il faut qu'elle soit basée sur une correspondance sémantique, et ce problème bien sûr, augmente le silence lors de l'interrogation.

Enfin, la recherche d'information structurée est un sujet d'actualité, nous assistons aujourd'hui aux déploiements de nouvelles méthodes, comme par exemple, l'intégration des outils de traitement automatique de langue pour l'interrogation, dont l'objectif est de traduire du mieux possible les requêtes exprimées en langage naturel vers un langage formel adapté au système de recherche d'information et au corpus. En ce qui concerne la problématique d'interrogation des corpus hétérogènes, une des solutions proposées est de transformer l'ensemble des documents du corpus dans un format de médiation connu par l'utilisateur formulant la requête.

Liste des références à consulter

[1] F.Z. BESSAI, M. BOUGHANEM, Z. ALIMAZIGHI, "Vers un modèle possibiliste pour la recherche d'information dans des documents structurés", Document numérique, Hermès Science, Lavoisier, volume10, n°1/2007, p.109-130, ISBN 978-2-7462-1969-4.

[2] F. Z. BESSAI-MECHMACHE and Z. ALIMAZIGHI, "Possibilistic Model for Aggregated Search in XML Documents", International Journal of Intelligent Information and Database Systems, IJIDS, Inderscience Publishers, Vol 6, No 4, p.381-404, ISSN (Online)1751-5866, ISSN (Print)1751-5858, 2012.

[3] G. KAZAI, "Choice of Parameter Values for the INEX Evaluation Metrics: Sensitivity Analysis", In Proceedings of INEX 2006 Workshop, 2006.

[4] A. KOPLIKU, F. DAMAK, K. PINEL-SAUVAGNAT and M. BOUGHANEM, "A user study to evaluate aggregated search", In IEEE/WIC/ACM International Conference on Web Intelligence (WIC 2011), France, 2011.

[5] K. SAUVAGNAT, M. BOUGHANEM, C. CHRISMENT, "Answering content-and-structure-based queries on XML documents using relevance propagation", Information Systems, Special Issue SPIRE 2004, vol. 31, p. 621-635, Elsevier, 2006.



LES CONSEILS DE DZ-CERT

Bonnes pratiques pour le déploiement sécurisé du navigateur Google Chrome

Aujourd'hui, Google Chrome est devenu l'un des navigateurs les plus utilisés par les internautes. Ce navigateur possède plusieurs fonctionnalités telles qu'un système de « bac à sable », des possibilités de déploiement et de configuration centralisés, et son système de mise à jour automatique réactif. Nous présentons dans ce qui suit quelques recommandations pour un déploiement sécurisé de ce navigateur web.

1. Protéger ses mots de passe enregistrés

Si vous laissez Chrome enregistrer vos mots de passe, toute personne qui utilise votre PC peut facilement y accéder à travers le panneau réglages. Il est donc recommandé de ne

permettre qu'aux personnes de confiance d'utiliser votre compte utilisateur. Vous pouvez créer un nouveau compte standard (non administratif) pour les autres utilisateurs ou les laissez utiliser le compte Invité. Vous pouvez aussi utiliser des extensions comme ChromePW, Browser Lock, ou un profil chrom sécurisé par mot de passe. Une autre option consiste à stocker vos données sensibles en utilisant un gestionnaire de mot de passe tiers.

2. Sécuriser vos données synchronisées

Si vous avez activé la synchronisation, Chrome conserve vos renseignements (préférences, mots de passe, etc) en utilisant vos identifiants de compte Google. La synchronisation devient cependant une vulnérabilité si votre compte Google venait à être piraté. Vous pouvez ajouter un niveau de protection en définissant un mot de passe de synchronisation, qui sera toujours requis pour accéder à vos données. Pensez aussi à renforcer la sécurité de votre compte Google en utilisant la validation en deux étapes par exemple.



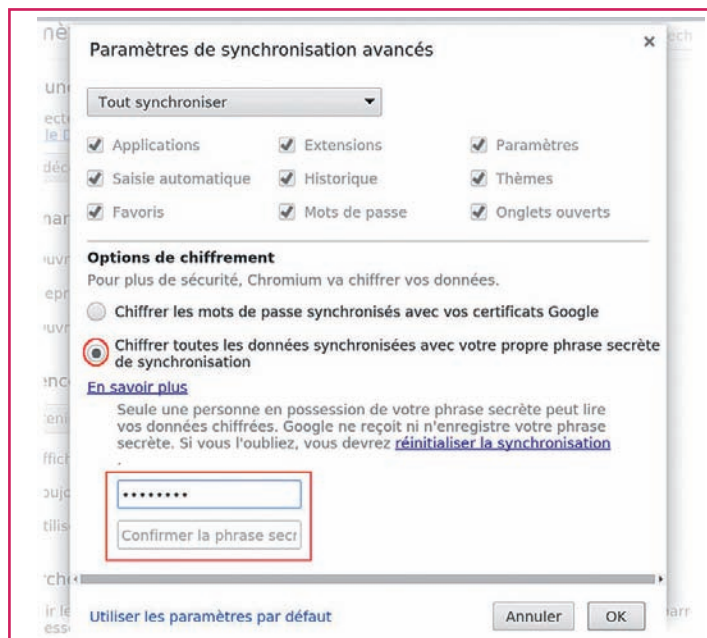


Figure 1 : Mot de passe de synchronisation

3. Safe browsing

Safe browsing est une fonctionnalité de sécurité qui vise à protéger l'internaute contre les sites de phishing et les téléchargements malveillants. Il fonctionne en envoyant les adresses de sites Web visités (sous forme de hash) à Google. Dans le cas d'une

page dangereuse, l'utilisateur sera averti par message. La navigation sécurisée peut être entièrement désactivée si le compromis entre la vie privée (envoi de données de navigation à google) et la sécurité (liens et pages dangereuses) n'est pas acceptable.

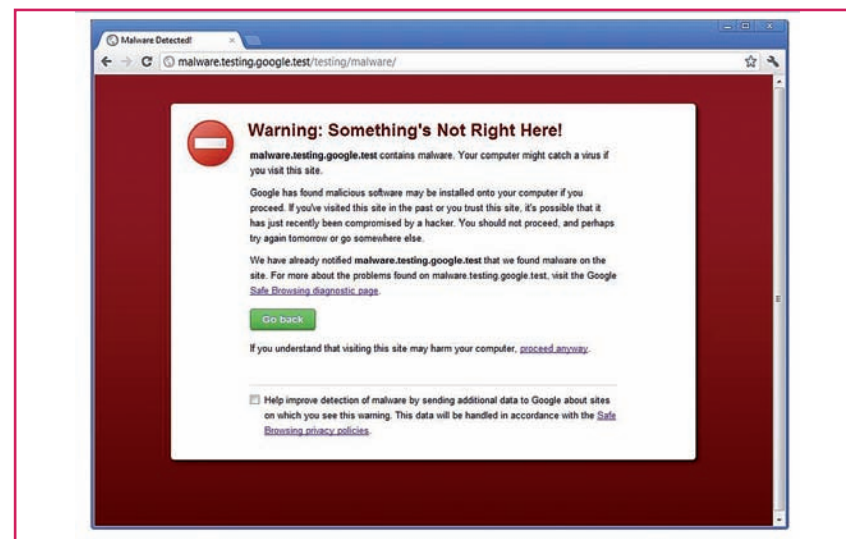


Figure 2 : Fonctionnalité Safe browsing

4. Sécurité des plugins/extensions

L'utilisateur peut installer des plugins et des extensions de la boutique en ligne Chrome.



Les plugins Chrome peuvent être développés en utilisant une des deux interfaces de programmation:

- NPAPI qui permet l'exécution des plugins avec le niveau de privilège de l'utilisateur. Une vulnérabilité affectant un plugin permet alors de compromettre la session ou le système.
- PPAPI: dont l'objectif est de rendre les plugins portables et sécurisés, et permet leur exécution en mode « bac à sable » en séparant les processus du navigateur et du moteur de rendu. L'exploitation de vulnérabilités du plugin reste alors confinée au processus.

Il est recommandé de n'autoriser que les plugins développés via PPAPI à l'inverse de NPAPI qui resté, malheureusement, l'outil le plus utilisé pour le développement de plugins. De même, il est recommandé de n'autoriser aucune extension sauf celles qui se justifient par un réel besoin métier comme par exemple les extensions de développement Web.

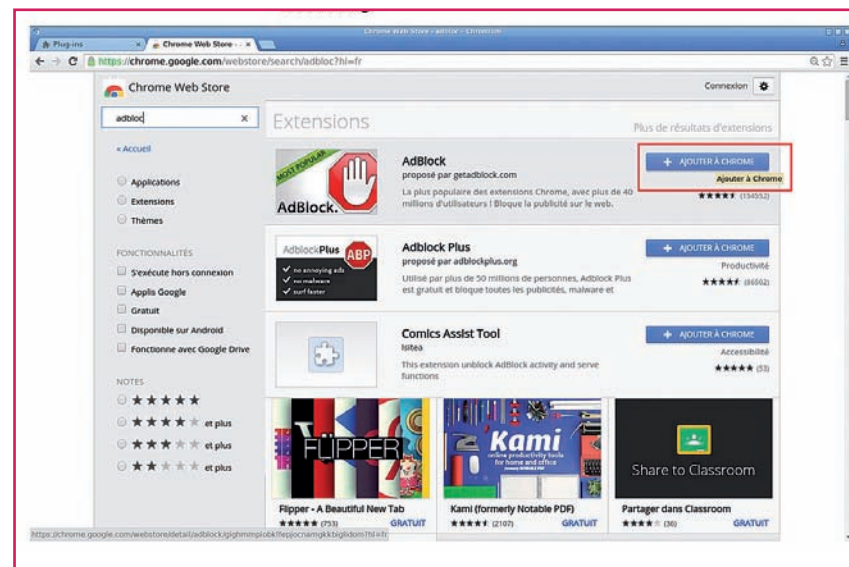


Figure 3 : Ajout des extensions dans Google Chrome

5. Installer des extensions pour plus de sécurité

Diverses extensions permettent d'ajouter encore plus de fonctions de sécurité. Par exemple, Adblock permet de supprimer les publicités qui peuvent conduire à des logiciels malveillants ou à des sites de Phishing.

6. Chrome, active directory et politique de sécurité

Dans ses dernières versions, Active Directory de Microsoft permet de définir des règles de sécurité dans les GPO (Group Policy Object), permettant ainsi une configuration centralisée du navigateur.



- • • En entreprise, il est recommandé d'utiliser cette fonctionnalité pour inclure des règles de configuration importantes pour la sécurité de la navigation.

7. Installer un logiciel antivirus à jour

Il est indispensable d'installer des programmes de sécurité tel qu'un antivirus et de faire en sorte qu'il soit régulièrement mis à jour afin de se protéger contre les nouvelles menaces. Certains de ces programmes permettent d'empêcher le téléchargement des logiciels dangereux sur votre ordinateur pendant la navigation sur Internet. De plus, ils permettent d'identifier les sites Web suspects au niveau des résultats de recherche vous informant du risque que présentent ces sites avant d'y accéder.



Google chrome

Zoom Sur le projet

**Reinforcement of ICT Regulations and
ICTs for tackling Societal Challenges links
in Europe and Mediterranean countries
(ClusMED)**

Mme Hadjira Boudier

M. Rafik Adour

Division Recherche et Développement en Science de l'Information -CERIST-



Présentation du projet

Le projet **ClusMed**¹- dont l'intitulé est « **Reinforcement of ICT Regulations and ICTs for tackling Societal Challenges links in Europe and Mediterranean countries** » - rendu en français par « Renforcement de la réglementation des TIC pour aborder et étudier les défis sociétaux en Europe et dans les pays Méditerranéens (MED) »- est un projet de coopération euro-méditerranéen financé entièrement par l'Union Européenne.

Le projet ClusMed constitue une action de soutien (AS) au volet coopération internationale entre l'UE et les pays méditerranéens (MED), dans le domaine des réglementations des TIC comme support aux défis sociétaux.

Lancé en janvier 2014 et prévu pour une durée de deux (2) ans, ce projet s'inscrit dans le cadre du programme européen FP7. Il regroupe onze (11) partenaires des deux rives de la méditerranée : **APRE** (Italie), **A-CING** (Espagne), **Université Sidi Mohammed Ben Abdellah** et **Maroc Numeric Cluster** (Maroc), **Digital Europe** (Belgique), **Ecole Supérieure des Communications de Tunis** (Tunisie), **Ministère de l'Enseignement Supérieur et de la Recherche Scientifique** (Algérie), **Egyptian Information, Telecommunication, Electronics and Software Alliance** (Egypte), **Arab Open University** et **The**

1. <http://www.clusmed.eu/>

Arab ICT Association (Liban) et **Federation of Hellenic Information Technology and Communication Enterprises** (Grèce).

Afin de maximiser l'impact et la diffusion des résultats de ce projet, des synergies ont été mises en place avec d'autres importantes initiatives entrant dans le cadre de la coopération internationale entre l'UE et les pays MED, notamment avec MOSAIC et MED Dialogue, les deux autres projets résultant de l'appel « FP7-ICT-2011.10.3 International partnership building and support to dialogues », mais aussi avec EMERG : l'initiative euro-méditerranéenne du groupe de régulateurs. ClusMed établira également un échange d'information approprié avec d'importantes plateformes technologiques européennes relatives aux TIC.

La Direction Générale de la Recherche Scientifique et du Développement Technologique (DGRSDT), en tant que partenaire du projet ClusMed représentant l'Algérie, a officiellement saisi le CERIST et la Division (DRDSI) a été choisie pour la conduite de ce projet.

Objectifs du projet

ClusMed s'est donné pour objectif principal l'implémentation de la recherche coopérative dans le secteur des TIC à travers l'identification, l'analyse et le benchmarking des réglementations nationales en matière de TIC, tenant compte de cinq pays cibles de la zone MED (Algérie, Egypte, Liban, Maroc, Tunisie).



- ● ● Il vise à mettre l'accent sur le rôle des régulations en tant que facilitateurs du développement du secteur des TIC : à cette fin, il évaluera et comparera les cadres juridiques relatifs aux TIC en vigueur au sein des pays MED, dans le but d'identifier et de partager les meilleures pratiques. Les régulations en matière de TIC des pays MED feront également l'objet d'un benchmarking avec les normes et expériences européennes de régulation, afin de surmonter les obstacles empêchant l'exploitation complète des TIC comme appui à des défis sociétaux. Cet objectif sera en grande partie accompli à travers l'implémentation de workshops constitués de plusieurs points focaux (stakeholders), qui contribueront également à instaurer un dialogue permanent entre l'Europe et les pays méditerranéens sur les questions relatives aux régulations dans le domaine des TIC, comme un des facteurs clés de la croissance durable et inclusive. L'objectif principal, cité ci-dessus, peut se décliner en objectifs secondaires, à savoir :

- Accroître l'intérêt et l'importance de la réglementation en tant qu'outil de développement des TIC.
- Créer et renforcer des partenariats stratégiques inhérents à la réglementation et la régulation.
- Identifier et partager les meilleures pratiques en matière de réglementations et régulations des TIC en tant que support aux défis sociétaux.

Impacts socio-économiques du projet

Ce projet devrait avoir pour impacts ce qui suit :

- Instaurer un environnement de sécurité et de confiance chez les utilisateurs des TIC (entreprises, institutions, ménages,...), à travers notamment l'application du cadre juridique actualisé,
- Désenclaver les régions rurales,
- Répondre aux attentes du grand public à travers des services en ligne et de qualité,
- Réduction des coûts et temps d'exécution.

Action à effectuer

En vue d'atteindre les objectifs précédemment cités, une feuille de route a été mise en place. Cette dernière constitue un processus à quatre étapes appelées Work packages (WP par abréviation),

- Le WP1 (réservé à la Gestion du projet) : il vise à mettre en place une méthodologie pour l'accomplissement des tâches et activités prévues, ainsi qu'une politique de coordination entre pays partenaires, et entre ces derniers et leurs stakeholders respectifs. Il



visé aussi à proposer un modèle pour chaque rapport d'activité et prévoit une politique afin de gérer les potentiels risques du projet.

■ Le WP2 est dédié au « mapping », en l'occurrence la mise en place d'une cartographie du cadre juridique des TIC propre à chacun des pays de la rive sud de la méditerranée. Il comprend les étapes suivantes :

■ Identification des secteurs ayant introduit les TIC dans leurs activités respectives, et des points focaux au sein de ces secteurs (stakeholders),

■ Identification du cadre législatif et réglementaire afférents aux projets TIC initiés par ces secteurs, toujours par le biais des stakeholders,

■ Elaboration de la cartographie du cadre juridique des TIC, sur la base des données collectées dans les deux étapes précédentes via les stakeholders.

■ Ces derniers jouent un rôle capital, car considérés comme des points focaux. Ils sont informés des projets et initiatives lancés non seulement au sein de leurs institutions, mais également dans tout le secteur. Ils sont appelés à renseigner un questionnaire retraçant l'historique du cadre juridique mis en place par leurs institutions au profit du secteur dans lequel elles activent.

Les résultats de ce Work Package sont capitalisés dans une base de données accessible via le site web du projet ClusMed.

■ Le WP3 s'appuie quant à lui sur les données collectées dans le cadre du WP2. Ces données font l'objet d'une analyse au niveau national (notamment les partenaires de la rive sud de la méditerranée), suivi d'une comparaison des rapports d'analyse élaborés par les partenaires sud méditerranéens (comparaison régionale). Les résultats de ces deux étapes seront à leur tour comparés aux dispositifs juridiques adoptés par l'Union Européenne dans le domaine des TIC, en vue d'aboutir à un guide des meilleures pratiques.

■ Enfin le WP4 comprend la dissémination et le réseautage (du projet et de ses résultats), à travers le lancement d'activités promotionnelles et de communication, l'organisation et la gestion du site web (du projet), la mise en place d'un Groupe de Multi-stakeholders et la tenue des ateliers, l'organisation d'une Conférence Finale et le Feedback et l'interface avec d'autres initiatives.

Résultats du projet

Les résultats prévus dans le cadre de projet peuvent être résumés dans les points suivants :



- Cartographie de la réglementation relative aux TIC dans les différents secteurs d'activités.
- Identification des praticiens et professionnels ayant pour mission la mise en place de la réglementation TIC.
- Identification des points positifs et des points négatifs de la réglementation TIC en vigueur.
- Synthèse des expériences étrangères en la matière.
- Mise en place des stratégies d'avenir en matière de réglementation des TIC dans la zone Euro-MED.

Conclusion

Le projet ClusMed constitue une véritable opportunité pour l'Algérie dans sa mission d'adaptation de son dispositif légal au développement des TIC, en vue de promouvoir leur utilisation et de profiter au mieux de leurs avantages pour appuyer les défis sociétaux majeurs tels que : la santé, l'agriculture, l'éducation, la justice, ... C'est aussi une chance à saisir pour s'imprégner des expériences étrangères dans le domaine, pour mettre en place un cadre juridique national mieux adapté aux TIC, à travers le guide de bonnes pratiques notamment.

Par ailleurs, ces expériences- aussi encourageantes et efficaces soient elles- peuvent ne pas suffire sans l'implication des différents acteurs (décideurs, universitaires et experts), ou encore sans la prise en considération des particularités locales du pays. D'où la nécessité d'impliquer davantage le réseau des stakeholders dans les différentes étapes du projet et de veiller à son maintien même après sa clôture.

The screenshot shows the ClusMED website interface. At the top, there is a logo for ClusMED with the text 'Renforcement des réglementations et des lois régissant les TIC afin d'aborder les défis sociétaux en Europe et dans les pays méditerranéens.' Below the logo is a navigation bar with links for 'Accueil', 'Base de données d'information', 'à propos', 'Téléchargements publiques', 'Communauté', and 'Espace privé'. The main content area features a section titled 'Méthodologie du programme' with a sub-heading 'La méthodologie du programme peut se résumer en quatre étapes:'. Below this, there are four icons representing the steps: 1. Mapping of National ICT Regulations, 2. Analysis of the ICT regulatory landscape, 3. Benchmarking with European countries experience, and 4. Identification and dissemination of best practices and lessons learnt. Below the icons, there is a list of countries: 'Algeria, Egypt, Lebanon, Morocco, Tunisia and European Union'. To the right of the methodology section, there is a 'Nous avons l'énorme plaisir de vous présenter le projet ClusMED!' section with a sub-heading 'Événements'. This section lists three events: 'Le Séminaire "Le cadre juridique et institutionnel des TICs au Maroc et accès la citoyenneté numérique"', 'Ateliers ClusMED - Septembre / Novembre 2014', and 'Conférence finale de ClusMED - Octobre 2015'. Below the events section, there is a 'Nouvelles' section and a 'Bulletin d'information électronique' section with a form for 'Nom *' and 'Email *'.

<http://www.clusmed.eu>



FORMATION

Plusieurs formations assurées par des ingénieurs du centre ainsi que des enseignants de l'USTHB ont eu lieu durant la période de janvier à mars 2015. Deux sessions de formation, l'une sur la conception de site web niveau 1 et l'autre sur le niveau avancé, ont eu lieu au profit de l'Ecole Nationale Supérieure de Sciences Politiques (ENSSP) et de l'Ecole Nationale Supérieure de Journalisme (ENSJ). Les cadres de la Caisse Nationale du Logement (CNL) ont, eux, bénéficié d'une formation sur le langage Javascript durant le mois de février 2015. Au total, trente neuf cadres ont été formés.

RAPPORTS DE RECHERCHE INTERNES ([http : // www.cerist.dz/publications](http://www.cerist.dz/publications))

Abdi Radia, Cloud Computing. Alger: CERIST: 2015. ISRN CERIST-DRDSI/RR-15-00000008-dz
<http://dl.cerist.dz/handle/CERIST/720>

Djenouri Djamel, Bagaa Miloud, Energy Harvesting Aware Relay Node Addition for Power-Efficient Coverage in Wireless Sensor Networks. Alger: CERIST: janvier 2015. ISRN CERIST-DTISI/RR--15-00000001—DZ
<http://dl.cerist.dz/handle/CERIST/702>

- • • Bessai Fatma-Zohra, Dynamic Structure Generator for Heterogeneous XML Documents. Alger: CERIST: février 2015. ISRN CERIST-DTISI/RR--15-00000002—DZ
<http://dl.cerist.dz/handle/CERIST/707>
- Khider Hadjer, Benna Amel, Social Business Process Management Approaches: A Comparative Study. Alger: CERIST: février 2015. ISRN CERIST-DSISM/RR--15-00000006—DZ
<http://dl.cerist.dz/handle/CERIST/713>
- Bedjou Khaled, Berkani Lamia, Aouaou Faïçal, Une approche sémantique pour la recommandation de services dans le cadre d'une formation en ligne. Alger: CERIST: mars 2015. ISRN CERIST-DTISI/RR--15-00000007—dz
<http://dl.cerist.dz/handle/CERIST/714>
- Djenouri Youcef, Bendjoudi Ahcène, Mehdi Malika, Habbas Zineb, Nouali-Taboudjemat Nadia, Data reordering for minimizing threads divergence in GPU-based evaluating association rules. Alger: CERIST: 26 mars 2015. ISRN CERIST-DTISI/RR--15-00000009—dz
<http://dl.cerist.dz/handle/CERIST/722>
- Djenouri Djamel ; Bagaa Miloud, A Variant of Connected Dominating Set for Application in Communication Networks. Alger: CERIST: 30 mars 2015. ISRN CERIST-DTISI/RR--15-00000010—DZ
<http://dl.cerist.dz/handle/CERIST/723>
- Djenouri Djamel ; Karbab, ElMouatezbillah ; Boulkaboul, Sahar ; Bagula, Antoine, Car Park Management with Networked Wireless Sensors and Active RFID. Alger: CERIST: 30 mars 2015. ISRN CERIST-DTISI/RR--15-00000011--DZ
<http://dl.cerist.dz/handle/CERIST/724>

CERIST

Bases de données documentaires

Accessibles sur : www.cerist.dz

CERISTNEWS



Le CERIST permet l'accès à une documentation électronique nationale et internationale couvrant tous les domaines scientifiques et techniques grâce au Système National de la Documentation en Ligne (SNDL). Ce système concerne les chercheurs, les enseignants chercheurs et les étudiants.

De plus amples informations sont disponibles sur le site www.sndl.cerist.dz

SndL SYSTÈME NATIONAL DE DOCUMENTATION EN LIGNE

cerist

A PROPOS ACTUALITES BASES DE DONNEES PORTAILS FORMATIONS CONTACTS Connexion

SCIENTES & TECHNIQUES Plus

SCIENTES DE LA VIE & DE LA TERRE Plus

SCIENTES HUMAINES & SOCIALES Plus

PLURIDISCIPLINAIRES Plus

Pour effectuer une recherche, CLIQUEZ ICI

→ A Propos Du SNDL ?

Votre portail d'accès aux ressources électroniques nationales et internationales en ligne

Le SNDL vous permet l'accès à la documentation

→ Charte SNDL

Le SNDL comprend plusieurs catégories de ressources électroniques :

- ✓ Les ressources acquises via des abonnements chez des fournisseurs habilités : Elles sont classées en quatre grands domaines : Sciences de la vie et de la terre, Sciences et techniques, Sciences humaines et sociales, Multidisciplinaires. Ces ressources sont de plusieurs types : e_journals, bases de données scientométriques, e_books, etc.
- ✓ Les ressources libres disponibles sur le Net

→ Actualités et Nouveautés

- NEWS... WEB OF KNOWLEDGE: Nouvelles Séances de Formation: Le Web of Science et la bibliométrie, Le Journal Citation Reports

Directeur de publication

Pr. BADACHE Nadjib

Dossier : RECHERCHE D'INFORMATIONS STRUCTURÉES

Réalisé Par : **Dr F.Z. Bessai-Mechmache**

Maître de Recherche B,

Équipe Recherche d'Information,
Division Théorie et Ingénierie des Systèmes informatiques - CERIST -

Rubrique : Les Conseils de DZ - CERT

L'ÉQUIPE DZ-CERT

Rubrique : Zoom sur un Projet

CERIST

Comité de communication et de rédaction

BEBBOUCHI Dalila

BENNADJI Khedidja

DJETTEN Fatiha

Photographies

BENBOURECHE Amina

Réalisation graphique

BOUKEZOULA Mohamed Amine

Publié par le CERIST

5, rue des 3 Frères Aissou. Ben Aknoun. BP 143, 16030 - Alger

Tél : +213 (21) 91 62 05 – 08 / Fax : +213 (21) 91 21 26

E - mail : vr@mail.cerist.dz

www.cerist.dz

Impression

ANEP

ISSN : 2170-0656 / DÉPÔT LÉGAL : 2690-201



Le Bulletin CERISTNEWS

CENTRE DE RECHERCHE SUR L'INFORMATION SCIENTIFIQUE ET TECHNIQUE - CERIST

5, Rue des Trois Frères Aissou, Ben - Aknoun - BP 143. 16030 - Alger

Tél : +213 (21) 91 62 05 - 08 / Fax : +213 (21) 91 21 26

www.cerist.dz