

**Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique**

**Institut National de formation en Informatique (I.N.I)  
Oued-Smar Alger**

**Mémoire de fin d'études**

Pour l'obtention du diplôme d'ingénieur  
d'état en informatique

**Option : Systèmes Informatiques**

**Thème**

**Recherche Intelligente  
sur Documents HTML**

**Réalisé par :**

Mlle KHIAT Nassima  
Mlle MOUHEB Djedjiga

**Proposé par :**

M<sup>r</sup> MAABOUT Sofiane

**Dirigé par :**

M<sup>me</sup> BOUMELLIL Lamia-Sabéha  
M<sup>me</sup> KEDDARI Djalila

**Organisme d'accueil :** Centre de Recherche sur l'Information Scientifique et Technique  
(CERIST)

**Promotion:** 2002/2003

## **RESUME**

Les moteurs de recherche actuels permettent d'effectuer une recherche sur le WEB en utilisant des mots clés ou, plus précisément, des combinaisons de mots clés, et retournent l'ensemble des documents contenus dans le Web, associés à ces mots. De plus, ils trient les résultats obtenus dans une liste, par ordre décroissant de pertinence.

Cependant, dans certains cas, la présentation des résultats sous forme d'une liste n'est pas suffisante pour retrouver rapidement les documents qui répondent à la requête de l'utilisateur, surtout lorsque le mot recherché a plusieurs sens. Ce qui serait intéressant dans ce cas, c'est de pouvoir associer chaque document à un thème particulier, représentant un sens du mot recherché. L'ensemble des thèmes formera une ontologie. Le résultat de la recherche apparaîtra ainsi sous forme de graphe, les nœuds correspondant à des thèmes et les arcs représentant des liens entre ces thèmes. Ceci représente la tendance actuelle avec l'avènement annoncé du « WEB sémantique ».

Notre travail s'inscrit dans le cadre des recherches sur le Web sémantique. Il s'agit de la réalisation d'un moteur de recherche permettant d'effectuer une recherche sémantique sur le site Web d'un journal (résultats classés d'après les thèmes abordés par le journal).

### **Mots clés**

Web sémantique, ontologie, indexation, moteur de recherche.

## **ABSTRACT**

Today's Web search engines allow retrieval information on the Web by using a combination of keywords, and return a set of Web documents that contains those words. More than this, they classify the results in a list by increasing order of pertinence.

However, in certain cases, the result presentation in a list form is not sufficient to find rapidly documents that answer to the user request, especially when the word being searched has several meanings. In this case, it would be more interesting to associate each document to a specific topic representing a sense of the searched word. The set of topics will form an "ontology". Thus, the search results appear in a graph form, where nodes correspond to topics, and arcs represent links between those topics. This is the current tendency with the announced advent of "the Semantic Web".

Our work is registered in the context of semantic Web research. It consists on a realization of a Web search engine that makes possible a semantic search on a newspaper Web site (the result are classified by topics approached by the newspaper).

### **Keywords**

Semantic Web, Ontology, indexing, Web search engine.

# SOMMAIRE

## Introduction générale

## Chapitre I : Du Web traditionnel au Web sémantique

Introduction.....	1
I.1 – Préhistoire du Web.....	1
I.2 – Le Web traditionnel.....	2
I.2.1 – HTML.....	2
I.2.2 – Hypertexte.....	2
I.2.3 – Modèle Client/Serveur.....	3
I.2.4 – HTTP.....	5
I.3 – Le Web structuré.....	5
I.3.1 – XML.....	5
I.3.2 – XML Schéma.....	7
I.4 – Le Web sémantique.....	7
I.4.1 – RDF.....	8
I.4.2 – RDF Schéma.....	10
Conclusion.....	12

## Chapitre II : Ontologie

Introduction.....	13
II.1 – Historique.....	13
II.2 – Qu’est-ce qu’une ontologie ?.....	13
II.3 – Pourquoi utiliser une ontologie ?.....	14
II.4 – Construction d’une ontologie.....	15
II.4.1 – Une méthodologie simple pour la construction d’une ontologie.....	15
II.4.2 – Définition des classes et de leur hiérarchie.....	19
II.4.3 – Implémentation d’une ontologie.....	23
II.4.3.a – OIL.....	24
II.4.3.b – DAML et DAML+OIL.....	24

II.4.3.c – OWL.....	24
II.4.4 – Documentation d’une ontologie.....	25
II.4.5 – Evaluation d’une ontologie.....	25
II.5 – Quelques ontologies existantes.....	25
II.5.1 – WordNet.....	25
II.5.2 – AAT.....	26
II.5.3 – EngMath.....	26
Conclusion.....	26

### Chapitre III : Systèmes de recherche et d’indexation

Introduction.....	27
III.1 – Définitions d’un SRI.....	27
III.2 – But d’un SRI.....	28
III.3 – Composants d’un SRI.....	28
III.3.1 – Indexation	
III.3.1.a – Définition.....	29
III.3.1.b – But de l’indexation.....	30
III.3.1.c – Composants de l’indexation.....	30
III.3.1.d – Indexation manuelle, mixte et automatique.....	34
III.3.1.e – Techniques d’indexation.....	35
III.3.1.f – Avantages et inconvénients de chaque technique.....	44
III.3.1.g – Critères d’une bonne indexation.....	46
III.3.2 – Recherche d’informations	
III.3.2.a – Définition.....	47
III.3.2.b – Techniques de recherche.....	47
III.3.2.c – Comment exprimer une requête.....	48
III.4 – Présentation des résultats de la recherche.....	51
III.5 – Evaluation d’un SRI.....	54
III.5.1 – Bruit et silence.....	54
III.5.2 – Rappel et précision.....	55
Conclusion.....	56

## **Chapitre IV : Etude Conceptuelle**

Introduction.....	57
IV.1 – Schéma conceptuel.....	58
IV.1.1 – Construction de l’ontologie.....	59
IV.1.2 – Implémentation de l’ontologie.....	62
IV.1.3 – Indexation.....	64
IV.1.3.a – Elimination des balises.....	66
IV.1.3.b – Elimination des mots vides.....	66
IV.1.3.c – Normalisation.....	67
IV.1.3.d – Mise à jour de l’index.....	69
IV.1.4 – Recherche et présentation des résultats.....	71
IV.1.4.a – Reconnaissance de la requête.....	73
IV.1.4.b – Normalisation des mots de la requête.....	75
IV.1.4.c – Formulation de la requête.....	75
IV.1.4.d – Affectation des articles aux thèmes.....	76
IV.1.4.e – Tri des résultats.....	78
IV.1.4.f – Affichage des résultats.....	78

## **Chapitre V : Réalisation**

Introduction.....	79
V.1 –Fonctionnalités du système réalisé.....	79
V.2 – Choix des logiciels .....	79
V.2.1 – Choix du langage de programmation.....	79
V.2.2 – Choix du SGBD.....	80
V.3 – Choix du journal.....	80
V.4 – Implémentation du système .....	81
V.4.1 – Module de saisie de l’ontologie d’un journal.....	81
V.4.2 – Recherche et affichage des résultats.....	87

## **Conclusion générale**

**Annexe : Outils JAVA**

## **Bibliographie**