



Université des Science et de la Technologie Houari Boumediene

**Faculté de Mathématiques**

**Département de Recherche Opérationnelle**

---

Projet de Fin d'Etudes

En vue de l'obtention du diplôme  
D'Ingénieur d'Etat en Recherche Opérationnelle

**Thème**

**Développement d'une approche algorithmique  
Pour l'exploration du web**

Proposé par : M<sup>elle</sup> M.NEKRI

Présenté par : M<sup>elle</sup> Ahmed Zaid Mounia  
et M<sup>elle</sup> Chali Nadia

Encadré par : M<sup>me</sup> N.MAACHOU(USTHB)  
M<sup>elle</sup> M.NEKRI (CERIST)

Membre de jury :

Président : Mr M.MOULAI

Examineur: Mr M.SAHARI

Rapporteur: Mme N.MAACHOU

**Code mémoire :7.1.06**

## Table des Matières

INTRODUCTION.....1

### PARTIE : ÉTUDE THEORIQUE.

#### CHAPITRE I : Généralités.

##### Partie 1 : Internet et Le Web.

I.1 Introduction.....5

I.2 Présentation de l'Internet.....5

I.3 Présentation du World Wide Web.....5

I.4 La visibilité du Web.....6

I.5 Les causes du Web invisible :.....7

*I.5.1 Les documents dynamiques.....7*

*I.5.2 Le référencement.....7*

*I.5.3 L'indexation.....7*

I.6 Les différents types du Web : .....8

*I.6.1 Le Web opaque.....9*

*I.6.2 Le Web privé.....9*

*I.6.3 Le Web propriétaire.....9*

*I.6.4 Le vrai Web invisible.....9*

I.7 L'accessibilité du Web. ....9

I.8 Les causes de l'inaccessibilité du Web. ....10

Conclusion.....11

##### Partie 2 : Notions Sur Les Moteurs de Recherche.

I.9 Introduction aux Moteurs de Recherche.....13

I.10 Web crawler ou robot de Moteur de Recherche. ....13

I.11 Les politiques d'exploration : .....13

*I.11.1 La politique de sélection.....14*

*I.11.2 La politique de revisite : .....14*

*i La politique uniforme. ....15*

*ii La politique proportionnelle. ....15*

*I.11.3 La politique de politesse.....15*

<i>I.11.4 La politique de parallélisation.</i> .....	16
I.12 Exemples de Web crawlers. ....	16
I.13 Architecture d'un Web crawler. ....	18
Conclusion.....	19

## CHAPITRE II : Analyse de l'Existant.

II.1 Concepts fondamentaux de la théorie des graphes : .....	21
<i>II.1.1 Définitions</i> .....	21
<i>II.1.2 Type de représentation d'un graphe</i> .....	23
<i>II.1.3 Connexité et forte connexité</i> .....	23
II.2 Rappel en probabilité : .....	23
<i>II.2.1 Loi de probabilité</i> .....	23
<i>II.2.2 Loi de Poisson</i> .....	23
<i>II.2.3 Loi de Zipf</i> .....	24
II.3 Composition du Web : .....	24
II.3.1 Type de liens hypertextes : .....	24
i Liens internes.....	24
ii Liens externes.....	25
iii Liens emails.....	25
II.3.2 Autre classification des liens : .....	25
i Liens entrants.....	25
ii Liens sortants.....	25
II.4 La structure du Web : .....	26
II.4.1 Objectif.....	26
II.4.2 La modélisation.....	26
II.4.3 La vision macroscopique : .....	26
II.4.3.1 Structure de la vision macroscopique : .....	27
<i>i La partie centrale (noyau fortement connexe)</i> .....	27
<i>ii La partie appelée « Origine ou IN »</i> .....	27
<i>iii La partie « Extrémités ou OUT»</i> .....	27
<i>iv Les « Branches ou des TENDRILS »</i> .....	27
<i>v Les pages déconnectées</i> .....	27
II.4.3.2 Caractéristiques de la structure.....	27

II.4.4 La vision microscopique.....	28
II.5 Propriétés du graphe du Web :.....	30
II.5.1 La distribution de degrés :.....	30
II.5.2 La distance moyenne .....	34
II.5.3 Le coefficient de « clusterisation ».....	34
II.5.4 Small- world .....	34
II.6 Modèles réalistes du graphe du Web.....	34
II.7 Cartographie du Web.....	35
Conclusion.....	36

### **Chapitre III : Position du Problème.**

III.1 Introduction.....	38
III.2 Les problèmes liés à la recherche et localisation d'information sur le Web :.....	38
III.2.1 Une taille croissante des données à manipuler.....	38
III.2.2 Absence de solution optimale :.....	39
i Silence.....	39
ii Bruit :.....	42
III.2.3 Les outils actuels et leur dysfonctionnement.....	42
III.3 Problématique.....	42
III.4 Les études faites dans ce domaine :.....	43
III.4.1 Projet JAVA 2005.....	43
III.5 Les algorithmes d'exploration :.....	46
III.5.1 <i>Breadth-First</i> .....	46
III.5.2 <i>Best-First</i> .....	46
III.5.3 <i>Page Rank</i> :.....	47
III.5.4 <i>Shark-Search</i> .....	50
III.5.5 <i>L'algorithme HITS</i> .....	50
III.6 Les méthodes de tri.....	52
III.6.1 <i>La méthode de tri par pertinence</i> .....	52
III.6.2 <i>La méthode de tri par popularité</i> .....	52
Conclusion.....	52

**PARTIE : RÉALISATION.****CHAPITRE IV : Modélisation et Réalisation du Problème.**

IV.1 Introduction.....	55
IV.2 Les étapes à suivre pour la conception.....	55
IV.2.1 Les politiques d'exploration :.....	55
<i>i La politique de sélection.....</i>	55
<i>ii La politique de politesse.....</i>	55
<i>iii La politique de parallélisation.....</i>	55
<i>iv La politique de revisite.....</i>	55
IV.2.2 La méthode d'exploration.....	56
IV.2.3 La représentation graphique.....	56
IV.2.4 Structure des données.....	56
IV.3 L'algorithme.....	57
IV.4 Fonctionnement du robot d'exploration.....	59
IV.5 Etude de la complexité.....	61
IV.6 Finitudes et convergence.....	62

**CHAPITRE V : Description et Présentation du Logiciel.**

V.1 Environnement de développement.....	64
V.2 Le teste effectué.....	65
V.3 Analyse des Résultats.....	68
<b>Conclusion Générale.....</b>	<b>71</b>
<b>Glossaire.....</b>	<b>73</b>
<b>Bibliographie.....</b>	<b>76</b>