



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE

LA RECHERCHE SCIENTIFIQUE



UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE « HOUARI BOUMEDIENE »

FACULTE D'ELECTRONIQUE ET INFORMATIQUE

DEPARTEMENT INFORMATIQUE

*Mémoire de fin d'études pour l'obtention du diplôme
d'Ingénieur d'Etat en Informatique*

Thème :

**UN SYSTEME MULTI-AGENTS POUR
UNE EXTRACTION EFFICACE DE
L'INFORMATION WEB**

Organisme d'accueil : C.E.R.I.S.T.

Thème proposé et encadré par :

M^{me} Mellah hakima

Réalisé par :

Mr Ghezghouz bilel

Mr Mekanane Salem

Devant le jury :

M^r H.AZZOUNE Président du jury

M^{me} A.BENCHENNAF Membre du jury

M^r S.KECHID Membre du jury

PROMOTION 2004/2005

Résumé

Les outils de recherche actuels présentent des limites dans leur capacité à répondre aux besoins des utilisateurs. Ces limites sont dues au fait que ces outils ne prennent pas en considération les profils des utilisateurs.

Les projets qui font de la recherche intelligente visent à dépasser ce problème. L'un des moyens existants consiste à utiliser les systèmes multi agents.

Dans notre projet, nous avons réalisé un système intelligent qui assiste les internautes dans leurs recherches en leur permettant d'accéder directement aux pages qui correspondent à leurs profils. Cette assistance se fera avec l'étude et l'apprentissage des préférences, des centres d'intérêts et des habitudes de navigations de l'utilisateur ainsi que d'autres utilisateurs pour permettre à l'internaute de tirer profit des expériences d'autrui. Ceci nous a conduit à partager notre système que nous avons baptisé **SWA (Smart Web Assistant)** en deux processus : le processus d'apprentissage et le processus d'extraction des pages.

MOTS-CLES : profil utilisateur, système multi-agents, apprentissage, classification, extraction de pages Web, indexation, filtrage.

Abstract

The actual research tools have a restricted capacity to respond to the users needs. These restrictions are due to the fact that the tools do not take into consideration the users' profiles. The plans that realise the intelligent research aim at resolving this problem, and one of the available means is the use of the multi-agent systems.

In this sense we have realised an intelligent system that assists Net surfers in their researches and enables them to have a direct access to the pages corresponding to their profiles. This assistance is realised by the study and apprenticeship of preferences, self interests, and the navigation habits of the user and also of other users, this will on the other hand make the Net surfers benefit from others experiences. This has led us to divide the system we have baptised SWA (**S**mart **W**eb **A**ssistant) into tow processes: the apprenticeship process and the extraction process of pages.

KEY WORDS: user profile, multi-agent system, apprenticeship, classification, extraction of web pages, indexation, filtering.

Introduction.....	1
-------------------	---

PARTIE I : LES SYSTEMES DE FILTRAGE & LES METHODES DE CLASSIFICATION

Chapitre 1: Les systèmes de filtrage d'informations

1. Introduction.....	2
2. Systèmes de filtrage d'informations (SFI).....	2
2.1 Définition	2
2.2 L'indexation.....	3
2.2.1 Définition.....	3
2.2.2 Les étapes nécessaires à l'indexation.....	3
a) Reconnaissance de sections.....	3
b) Tokenisation.....	3
c) Stoplist (Anti-dictionnaire)	4
d) Lemmatisation.....	4
e) Statistiques	5
f) Pondération des termes.....	5
2.2.3 Amélioration de l'index.....	5
a) Approche basée sur la fréquence d'occurrence	5
b) Approche basée sur la valeur de discrimination	6
c) Approche basée sur $tf \cdot idf$	6
2.2.4 Comparaison entre l'indexation des documents et celle des requêtes	7
2.3 Modèles de filtrage	8
2.3.1 Modèle Matching score	8
2.3.2 Modèle Booléen	9
2.3.3 Modèle Booléen basé sur des ensembles flous	10
2.3.4 Modèle Vectoriel	10
2.3.5 Modèle Probabiliste	11
2.3.6 Modèle Connexionniste	12
3. Conclusion	14

Chapitre 2 : Les méthodes de classification

1. Généralités sur la classification automatique.....	15
2. Représentation des données.....	15
3. Quelques méthodes de classification.....	16
3.1 Agrégation autour des centres mobiles.....	16
3.1.1 Algorithme d'agrégation autour des centres mobiles.....	16
3.1.2 Techniques connexes.....	18
3.2 Classification hiérarchique ascendante (CAH).....	19
3.2.1 Distance entre éléments et entre groupes.....	19
3.2.2 Algorithme de la CAH.....	20
3.3 Classification descendante hiérarchique.....	21
3.3.1 La méthode DIVAF.....	21
3.3.2 La méthode DIVOP.....	22
4. Conclusion.....	22

PARTIE II : ETAT DE L'ART SUR LES SYSTEMES

MULTI AGENTS (SMA)

1. Introduction.....	23
2. Qu'est ce qu'un Agent logiciel ?.....	23
3. Historique.....	23
4. Qu'est ce que l'intelligence d'un agent ?.....	24
5. Définition d'Agents intelligents.....	24
6. Caractéristiques d'un agent intelligent.....	25
7. Composants d'un Agent.....	26
8. Typologies des agents.....	27
8.1 Les agents réactifs.....	28
8.2 Les agents cognitifs.....	28
9. Architectures pour les agents intelligents.....	28
9.1 Architectures logiques.....	29
9.2 Architectures réactives.....	30
9.3 Architecture BDI.....	31
9.4 Architectures multi niveaux.....	34
10. Définition de système multi-agents.....	37
11. Propriétés des SMA.....	37
12. Les avantages des SMA.....	38
13. Problématiques des SMA.....	38
14. Les modèles de communication dans les SMA.....	39
15. Environnement dans un SMA.....	39
16. Interaction dans les SMA.....	40
16.1 Définition.....	40
16.2 Modélisation des interactions.....	40
16.3 Langage de Communication inter-Agents.....	41
16.3.1 Actes de langage.....	41
16.3.2 Communication utilisant KQML.....	42
16.3.3 Communication utilisant ACL de FIPA.....	42
16.3.4 Ontologie.....	43
16.4 Les protocoles d'interactions entre agents.....	44
16.4.1 Protocole de coordination entre agents.....	44
16.4.2 Protocole de coopération entre agents.....	44
a) Allocation par réseau d'accointances.....	45
b) Le Contract Net.....	46
17. Conclusion.....	46

PARTIE III : CONCEPTION

1. Introduction.....	47
2. Problématique.....	48
3. Solution proposée.....	48
4. Modélisation du système.....	49
4.1 Processus d'apprentissage.....	50
4.2 Processus d'extraction des pages.....	51
4.3 Identification des agents.....	51
4.3.1 Description des agents du processus d'apprentissage.....	51
4.3.2 Description des agents du processus d'extraction des pages.....	61
4.4 Identification des scénarios des processus.....	63

PARTIE IV : IMPLEMENTATION

1. Introduction.....	71
2. Environnement de développement.....	71
2.1 Choix du constructeur.....	71
2.2 Choix du langage de programmation.....	71
3. Les caractéristiques de SWA	72
4. Implémentation du système SWA	72
4.1 Architecture générale.....	72
4.2 Implémentation d'un agent avec Dmas.....	75
4.3 Implémentation d'une tâche.....	75
4.4 Fichiers et base de données.....	76
4.5 Les interfaces	78
5. Communication entre agents distants.....	80
6. Conclusion	81

Conclusion et Perspectives.....	82
---------------------------------	----

Annexe A : Models de communication

Annexe B : Web mining et fichiers Log

Annexe C : Dmas builder

Bibliographie