

REPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE  
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE

UNIVERSITÉ M'HAMED BOUGARA DE BOUMERDES



FACULTE DES SCIENCES  
DÉPARTEMENT DE PHYSIQUE  
STRUCTURE INFOTRONIQUE

MEMOIRE DE FIN D'ÉTUDES  
pour l'obtention du diplôme de  
**MASTER**

Domaine: **Science et Techniques**  
Filière: **Génie Electrique**  
Spécialité: **Systèmes informatiques distribués**

Thème:

**CONCEPTION ET REALISATION D'UN  
SYSTEME D'EXTRACTION DU PROFIL  
UTILISATEUR A BASE DES TAGS**

par:

**LASBEUR Imene**

**ABDELOUAHAD Imen**

**JURY**

1. M. Riahla, Maitre Assistant A
2. M. Hamadouche, Maitre de Conférences
3. Mme Kichou, Attachée de recherche
4. Mme Benmessaoud, Maitre Assistant B

Président  
Examineur  
Encadreur  
Promotrice

## Sommaire

### Introduction generale

**Introduction** \_\_\_\_\_ **3**

**Problématique** \_\_\_\_\_ **3**

**Organisation du mémoire** \_\_\_\_\_ **4**

### Première Partie: Etat de l'Art

#### Chapitre I : les systèmes du tagging collaboratifs

**Introduction** \_\_\_\_\_ **6**

**1. Tagging collaboratif** \_\_\_\_\_ **7**

1.1 Définition \_\_\_\_\_ 7

1.2 Les tags \_\_\_\_\_ 7

1.3 Les acteurs du tagging \_\_\_\_\_ 8

1.4 Classification et fonctions des tags \_\_\_\_\_ 9

**2. Structure d'une action du Tagging collaboratif** \_\_\_\_\_ **10**

2.1 Structure tripartite de base \_\_\_\_\_ 10

2.2 Structure tripartite avec liens inter-ressources et inter\_ utilisateurs \_\_\_\_\_ 11

2.3 Structure quadripartite \_\_\_\_\_ 12

**3. Les modèles du tagging collaboratif** \_\_\_\_\_ **13**

3.1 Les modèles descriptifs \_\_\_\_\_ 13

3.2 Les modèles prédictifs \_\_\_\_\_ 14

**4. Folksonomie** \_\_\_\_\_ **14**

4.1 Les types de folksonomies : \_\_\_\_\_ 15

4.1.1 Folksonomie étroite (*narrow folksonomies*) \_\_\_\_\_ 15

4.1.2 Folksonomie générale (*broad folksonomies*) \_\_\_\_\_ 15

**5. Le nuage de tags** \_\_\_\_\_ **16**

**6. Les propriétés d'un système du tagging collaboratif** \_\_\_\_\_ **17**

**7. Les avantages des systèmes du tagging collaboratif** \_\_\_\_\_ **18**

<b>8. Les limites des systèmes du tagging</b>	<b>19</b>
<b>9. Les solutions proposées pour remédier à l’ambiguïté et les variations d’écriture</b>	<b>20</b>
<b>9.1 Clustering</b>	<b>20</b>
<b>9.2 Détection des variations orthographiques</b>	<b>20</b>
<b>9.3 Détection des tags thématiquement proches</b>	<b>21</b>
<b>9.4 Ontologies comme support des folksonomies</b>	<b>22</b>
9.4.1 Guider le tagging a l’aide d’ontologies	22
9.4.2 Exemple d’ontologies informatiques pour le tagging	23
<b>10. Conclusion</b>	<b>25</b>
<b>Chapitre II : le profil utilisateur</b>	
<b>Introduction</b>	<b>26</b>
<b>1. Définition du Profil utilisateur</b>	<b>27</b>
<b>2. Gestion du profil utilisateur</b>	<b>27</b>
<b>3. Approches de représentation du profil utilisateur</b>	<b>28</b>
<b>3.1 Représentation ensembliste ou vectorielle</b>	<b>29</b>
<b>3.2 Représentation connexionniste</b>	<b>29</b>
<b>3.3 Représentation conceptuelle et sémantique</b>	<b>29</b>
<b>3.4 Représentation hiérarchique</b>	<b>30</b>
<b>3.5 Représentation multidimensionnelle</b>	<b>30</b>
<b>4. Modèle conceptuel du profil</b>	<b>33</b>
<b>5. Modélisation du profil</b>	<b>34</b>
<b>6. Acquisition du Profil</b>	<b>34</b>
<b>7. Evolution du profil utilisateur</b>	<b>35</b>
<b>8. Recherche à base de tags et profil utilisateur dans le Web 2.0</b>	<b>36</b>
<b>9. Conclusion</b>	<b>36</b>

### Chapitre III: Etude des approches d'extraction de profil à base des tags

<b>Introduction</b>	<b>37</b>
<b>1. Le profil utilisateur dans le web2.0</b>	<b>38</b>
<b>2. Les approches d'extraction du profil utilisateur</b>	<b>39</b>
2.1 L'approche naïve	39
2.2 L'Approche par co-occurrence	40
2.3 L'approche hybride	41
2.4 L'Approche adaptative	42
2.5 L'approche par calcul de la capacité et la force du tag	43
2.6 L'approche d'intégration de l'expertise dans le calcul du poids de tag	44
<b>3. Comparaison des différentes approches</b>	<b>45</b>
<b>4. Conclusion</b>	<b>46</b>

### Deuxième Partie Chapitre IV: Conception

<b>Introduction</b>	<b>47</b>
<b>1. Motivations</b>	<b>48</b>
<b>2. Principe général</b>	<b>49</b>
<b>3. Présentation du système</b>	<b>51</b>
3.1 Le modèle du profil	52
3.1.1 Représentation du profil	52
3.2 Pondération des tags	53
3.3 Construction du profil	57
<b>4. Conclusion</b>	<b>60</b>

### Chapitre V : Réalisation

<b>1. L'environnement de réalisation</b>	<b>62</b>
a. Oracle 10g	62
b. L'environnement SQL*Plus d'Oracle	62

c. IDE NetBeans _____	62
d. L'ontologie linguistique _____	63
<b>2. La collection de test _____</b>	<b>63</b>
2.1 Préparation de la collection _____	63
<b>3. Architecture du système _____</b>	<b>65</b>
<b>4. Architecture de la base de données _____</b>	<b>65</b>
<b>5. Description de l'application _____</b>	<b>66</b>
b. Espace utilisateur simple _____	67
c. Espace administrateur _____	69
<b>6. Résultats et discussion _____</b>	<b>73</b>
<b>7. Conclusion _____</b>	<b>73</b>
<b>Conclusion générale</b>	
<b>Synthèse _____</b>	<b>74</b>
<b>Résumé de la contribution _____</b>	<b>74</b>
<b>Perspectives _____</b>	<b>75</b>
<b>Bibliographie</b>	
<b>Les annexes</b>	

## *Résumé*

Les années 2000 ont vu l'émergence du concept du web 2.0, selon lequel l'utilisateur devient actif et collaboratif, il est impliqué dans la production de l'information où il peut enrichir le contenu des ressources en ligne. C'est les pratiques du tagging ou d'étiquetage collaboratif. Les systèmes du tagging collaboratif sont basés sur l'attribution de mots-clés librement choisis par des utilisateurs non spécialistes, qui favorise le partage et l'organisation des ressources et permet d'améliorer la recherche d'information. L'attribution des tags par les utilisateurs s'illustre tout particulièrement dans les sites de partage de photos ou vidéos (Flickr, Youtube) difficiles à indexer par les moteurs classiques. Ces tags peuvent être un bon indicateur sur les intérêts de l'utilisateur tout comme ses navigations et ses clicks. Il existe dans la littérature des approches permettant d'exploiter cette nouvelle information afin d'extraire le profil utilisateur. Pour que cette exploitation soit meilleure, les tags associés à une ressource doivent représenter le plus efficacement possible le contenu de celle-ci. La popularité est l'unique critère pris en considération dans les systèmes actuels pour juger l'efficacité du tag. Or celle-ci ne reflète pas toujours son importance et sa représentativité vis-à-vis de la ressource à laquelle il est associé. Pour cela nous avons proposé une approche d'extraction de profil, où nous procédons à la pondération des tags, favorisant ceux cités par un utilisateur expert par rapport à ceux cités par un novice, et déterminons par la suite, la force d'un tag à représenter une personne. Un vecteur des intérêts de l'utilisateur contenant les meilleurs tags est ainsi créé à partir des tops tags ayant la plus grande force.

### *Mots-clés*

Web2.0, Tagging collaboratif, Profil utilisateur, Extraction du profil.

## *Abstract*

The 2000s saw the emergence of the concept of Web 2.0, in which the user becomes active and collaborative, he is involved in the information production where he can enrich the content of online resources. It is the practice of collaborative tagging or labeling. Collaborative tagging systems are based on assigning keywords freely chosen by non-expert users, which promotes the sharing and organization of resources and improves the information retrieval. The tags allocation by users is illustrated particularly in sites sharing photos or videos (Flickr, YouTube) difficult to index by conventional engines.

These tags can be a good indicator of the user's interests as its navigations and clicks. There are approaches in the literature to exploit this new information to extract user profile. To have a better exploitation, tags resource must represent well this content. Popularity is the only criterion considered in existing systems to judge the effectiveness of the tag. But it does not always reflect its importance and representativeness to the resource. For this, we proposed an approach for extracting profile, we proceed to the tags weighting , favoring those cited by an expert user in relation to those cited by a novice, and then determine the strength of a tag to represent a person. A vector of the interests of the user containing the best tags is created from the tops tags with the greatest strength.

### Keywords

Web2.0, Tagging collaborative, user profile, profile extraction.

## ملخص

شهدت سنوات 2000 ظهور الويب، 2.0 الذي أصبح المستخدم فيه نشطا و متعاوننا في انتاج المعلومات حيث يمكنه إثراء محتوى الموارد في الانترنت. تلك هي ممارسات التوسيم التعاوني، أنظمة التوسيم التعاونية تعتمد على إضافة كلمات دالة مفتاحيه تلقائية من طرف مجموعة المستخدمين الغير الخبراء، الذين يساهمون في تقسيم وتنظيم الموارد وتحسين البحث عن المعلومات . إضافة الكلمات المفتاحية من طرف المستخدمين تتضح خاصة في مواقع مشاركة الصور أو مقاطع الفيديو(فليكر- يوتيوب)، التي من الصعب على المحركات التقليدية تصنيفها، هذه الكلمات المفتاحية يمكن أن تكون مؤثرا جيدا لاهتمامات المستخدم تماما مثل تصفحاته ونقراته، يوجد في الميدان مناهج تمكننا من استغلال هذه المعلومات الجديدة لاسترجاع اهتمامات المستخدمين ليكون هذا الاستثمار هو الأفضل، الكلمات المفتاحية المضافة لمورد ما، يجب أن تمثل محتواه بطريقة فعالة، الشعبية هي النظام الوحيد المعمول به في النظم الحالية للحكم على فعالية العلامة لكنه لا يعكس دائما أهمية وتمثيل الموارد المضافة إليها لهذا لقد اقترحنا منهجا لاستخراج اهتمامات المستخدم ، التي نستعملها في ترجيح هذه العلامات أو الكلمات المفتاحية ، يفضل هذا المنهاج الكلمات المفتاحية التي تم إدراجها من طرف مستخدم خبير عوضا عن تلك المستعملة من طرف مستخدم مبتدى، ومن ثم تحديد قوة العلامة لتمثيل الشخص، النتيجة هي مجموعة من اهتمامات الشخص التي تحتوي على أفضل الكلمات المفتاحية التي تم إنشاؤها من طرف أفضل الكلمات المفتاحية التي عندها اكبر القوى .

### الكلمات المفتاحية

الملاحظات، التوسيم التعاوني، مواصفات المستخدم، استرجاع المعلومات.