

## SUJET DE THESE

### ESTIMATION DES MODELES UTILISATEURS, SOURCES ET DOCUMENTS POUR LA RECOMMANDATION DE CONTENU MULTIMEDIAS

#### Description du sujet

##### 1. Contexte

L'accès à une infinité de contenus multimédias tel que la musique, les vidéos ou les documents textuels nécessite la mise en place de systèmes de recommandation intelligents. Un utilisateur aujourd'hui qui se divertit, s'informe, étudie ou travaille est confronté continuellement à un flux d'information multimédia provenant de plusieurs sources. Le nombre de sources ainsi que le nombre de documents disponibles ne cessent de croître. L'objectif de cette thèse est d'étudier et d'expérimenter des méthodes permettant la personnalisation des flux de contenus disponibles à un utilisateur pour les rendre plus utiles.

Dans la littérature, des travaux sur la recommandation de contenu ont été effectués avec des techniques de filtrage collaboratif ou recommandation par similarité de contenu. D'autres travaux ont porté sur la modélisation des profils utilisateurs en utilisant un ensemble de données explicitement fournies ou des mots-clés explicitement et implicitement obtenus. Dans le cadre de cette thèse, nous proposons d'associer les trois composantes principales qui constituent la base de la recommandation : le document multimédia, la source du contenu et le consommateur du contenu.

##### 2. Stratégie de recherche

Nous proposons de traiter le sujet de recommandation de contenu en étudiant les problèmes suivants :

- 1- La modélisation des documents multimédias. Nous proposons la modélisation par des vecteurs de concepts. Les techniques à expérimenter ici sont des techniques de classification tel que les SVM (Support Vector Machines), les GMM (Gaussian Mixture Models), les ANN (Artificial Neural Networks) etc. Des approches basées sur LSA (Latent Semantic Analysis) ou PLSA (Probabilistic Latent Semantic Analysis) seront étudiées et comparées à une approche directe basée sur des vecteurs de concepts prédéfinis.
- 2- La modélisation des sources de documents multimédias. Plusieurs techniques peuvent être utilisées dans ce cadre tel que les SVM, GMM, ANN etc.
- 3- La modélisation du profil utilisateur consommateur. La technique proposée est la modélisation des profils par HMM (Hidden Markov Models) afin de capter l'évolution temporelle des préférences d'un utilisateur.

Corpus disponible :

- +1 Million de descriptions textuelles de documents multimédias
- Les traces d'usage de +5000 utilisateurs

### **3. Cadre de la thèse**

La thèse se déroule dans le cadre du projet ANR National VIDEOSENSE.

Le candidat sera recruté par la société Ghanni dans le cadre d'une convention CIFRE.

Le candidat aura le laboratoire LIRIS Ecole Centrale de Lyon comme laboratoire d'accueil.

### **4. Compétences requises**

Le candidat doit avoir un Master-2 ou équivalent lui permettant de s'inscrire en Ecole Doctorale en France.

Le candidat doit avoir un très bon niveau dans :

- Techniques d'Intelligence Artificielle tel que ANN (Artificial Neural Networks), SVM (Support Vector Machines)
- Techniques d'analyse et d'apprentissage statistique et probabiliste tel que GMM (Gaussian Mixture Models), EM (Expectation-Maximization), HMM (Hidden Markov Models)
- Développement C, C++, Objective C, PHP, HTML et SQL
- Plateforme XCode

Une expérience en développement JAVA est un plus.

Le candidat doit démontrer ses capacités de recherche et de développement par des publications/rapports scientifiques et des prototypes logiciels développés.

### **4. Publications**

Il sera demandé du candidat de soumettre des publications scientifiques aux journaux et conférences scientifiques du domaine.

Une liste non-exhaustive des journaux et conférences scientifiques cibles :

- User Modeling and User-Adapted Interaction Journal UMUI
- Multimedia Tools and Applications Journal MTAP
- IEEE Multimedia Journal
- IEEE Intelligent Systems
- International Conference on User Modeling Adaptation and Personalization UMAP
- IEEE International Conference on Multimedia and Expo ICME
- International Conference on Adaptive Hypermedia

### **5. Date de début**

01/10/2010 pour une durée de 3 ans

### **6. Contact : [hadi.harb@ghanni.com](mailto:hadi.harb@ghanni.com)**