

Modélisation dynamique et temporelle de l'utilisateur pour un filtrage personnalisé de documents textuels

Rachid AREZKI *, Abdenour MOKRANE*
Pascal PONCELET*, Gérard DRAY*
David Williams PEARSON**

*Centre LGI2P EMA, Site EERIE Parc Scientifique Georges Besse 30035 Nîmes
Cedex 1, France

{rachid.arezki, abdenour.mokrane, gerard.dray, pascal.poncelet}@ema.fr,

**IUT de Roanne, 20 Avenue de paris 42334 Roanne, France
david.pearson@univ-st-etienne.fr

Résumé. L'apprentissage efficace du profil utilisateur est un challenge car il évolue sans cesse. Dans cet article nous proposons une nouvelle approche pour l'apprentissage du profil long-terme de l'utilisateur pour le filtrage de documents textuels. Dans ce cadre les documents consultés sont classés de manière dynamique et nous analysons la répartition dans le temps des classes de documents afin de déterminer le mieux possible les classes d'intérêts de l'utilisateur. L'étude empirique confirme la pertinence de notre approche pour une meilleure personnalisation de documents.

1 Introduction

Avec le développement d'Internet et les nouveaux moyens de stockage de données, les serveurs de documents en ligne regorgent d'énormes quantités de documents de différentes thématiques. Les moteurs de recherche sont d'une grande utilité pour la recherche de documents pertinents mais ils nécessitent de formuler de nouvelles requêtes à chaque fois que l'utilisateur a besoin de nouveaux documents. Récemment, des systèmes capables de proposer des documents adaptés à l'utilisateur, sans que ce dernier ne formule de requêtes ont été développés. Cependant ils ne prennent pas en compte l'évolution dans le temps des classes de documents consultés par l'utilisateur. Dans ce papier nous proposons une nouvelle approche d'apprentissage du profil long-terme de l'utilisateur pour le filtrage de documents textuels. Cette approche est basée sur l'analyse de l'évolution dans le temps des classes de documents consultés par l'utilisateur. Dans ce cadre, les documents consultés sont classés de manière dynamique et nous analysons ensuite la répartition dans le temps de ces classes de documents. Le but de notre approche est de déterminer le mieux possible les classes d'intérêts de l'utilisateur, cela en donnant plus d'importance aux classes de documents régulièrement consultées qu'à celles concentrées sur de courtes périodes. Notre approche ne requiert donc pas que l'utilisateur fournisse de manière explicite des informations au système. Nous avons développé *LUCI*, un système permettant la personnalisation de documents en ligne via notre approche [Arezki *et al.*, 2003]. *LUCI* apprend le profil de l'utilisateur via les documents consultés par celui-ci et lui propose des documents de manière dynamique,