

Chapter XV

User Profiles for Personalizing Digital Libraries

Giovanni Semeraro
University of Bari, Italy

Pierpaolo Basile
University of Bari, Italy

Marco de Gemmis
University of Bari, Italy

Pasquale Lops
University of Bari, Italy

ABSTRACT

Exploring digital collections to find information relevant to a user's interests is a challenging task. Information preferences vary greatly across users; therefore, filtering systems must be highly personalized to serve the individual interests of the user. Algorithms designed to solve this problem base their relevance computations on user profiles in which representations of the users' interests are maintained. The main focus of this chapter is the adoption of machine learning to build user profiles that capture user interests from documents. Profiles are used for intelligent document filtering in digital libraries. This work suggests the exploiting of knowledge stored in machine-readable dictionaries to obtain accurate user profiles that describe user interests by referring to concepts in those dictionaries. The main aim of the proposed approach is to show a real-world scenario in which the combination of machine learning techniques and linguistic knowledge is helpful to achieve intelligent document filtering.

INTRODUCTION

Personalization has become an important topic for digital libraries to take a more active role in dynamically tailoring their information and service offers to individuals in order to better meet their needs (Callan & Smeaton, 2003). Most of the work on *personalized information access* focuses on the use of machine learning algorithms for the automated induction of a structured model of a user's interests, referred to as user profile, from labeled text documents (Mladenic, 1999). Keyword-based user profiles suffer from problems of polysemy and synonymy. The result is that, due to synonymy, relevant information can be missed if the profile does not contain the exact keywords occurring in the documents and, due to polysemy, wrong documents could be deemed as relevant.

This work explores a possible solution for this kind of issue: the adoption of semantic user profiles that capture key concepts representing users' interests from relevant documents. Semantic profiles will contain references to concepts defined in lexicons like WordNet (Miller, 1995) or ontologies. The solution is implemented in the item recommender (ITR) system which induces semantic user profiles from documents represented by using WordNet (Degemmis, Lops, & Semeraro, 2007). An example of intelligent personalized recommendation service based on ITR will be shown.

RELATED WORK

Our research was mainly inspired by the following works.

Syskill & Webert (Pazzani & Billsus, 1997) is an agent that learns a user profile to identify interesting Web pages. The learning process is performed by first converting a hypertext markup language (HTML) source into positive and negative examples, represented as keyword vectors,

and then using learning algorithms like Bayesian classifiers, a nearest neighbor algorithm, and a decision tree learner.

Personal WebWatcher (Mladenic, 1999) is a Web browsing recommendation service that generates a user profile based on the content analysis of the requested pages. Learning is done by a naïve Bayes classifier where documents are represented as weighted keyword vectors, and classes are "interesting" and "not interesting."

Mooney and Roy (2000) adopt a text categorization method in their Libra system that performs content-based book recommendations by exploiting product descriptions obtained from the Web pages of the Amazon online digital store. Also in this case, documents are represented by using keywords, and a naïve Bayes text classifier is adopted.

The main limitation of these approaches is that they represent items by using keywords. The objective of our research is to create accurate semantic user profiles. Among the state-of-the-art systems that produce semantic user profiles, SiteIF (Magnini & Strapparava, 2001) is a personal agent for a multilingual news Web site that exploits a sense-based representation to build a user profile as a semantic network, whose nodes represent senses of the words in documents requested by the user.

The role of linguistic ontologies in knowledge-retrieval systems is explored in OntoSeek (Guarino, Masolo, & Vetere, 1999), a system designed for content-based information retrieval from online yellow pages and product catalogs. OntoSeek combines an ontology-driven content-matching mechanism based on WordNet with a moderately expressive representation formalism. The approach has shown that structured content representations coupled with linguistic ontologies can increase both recall and precision of content-based retrieval.

We adopted a content-based method able to learn user profiles from documents represented by using senses of words obtained by a word

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/user-profiles-personalizing-digital-libraries/19878

Related Content

Electronic Learning and Social Media Tools: Consecration on Alpha Trend

Geetha Manoharan (2025). *Exploring Digital Metrics in Academic Libraries* (pp. 465-492).

www.irma-international.org/chapter/electronic-learning-and-social-media-tools/367962

Assessment of the Social Media in Knowledge Sharing: A Sociological Study

P. K. Muthukumar (2020). *Challenges and Opportunities of Open Educational Resources Management* (pp. 224-236).

www.irma-international.org/chapter/assessment-of-the-social-media-in-knowledge-sharing/258136

A Survey of Digital Forensic Techniques for Digital Libraries

Yue Li (2011). *International Journal of Digital Library Systems* (pp. 49-66).

www.irma-international.org/article/survey-digital-forensic-techniques-digital/59888

Role of Public Libraries in Bridging the Digital Divide

Owajeme Justice Ofuaand Ogochukwu Thaddaeus Emiri (2011). *International Journal of Digital Library Systems* (pp. 14-22).

www.irma-international.org/article/role-public-libraries-bridging-digital/59884

Using Topic-Specific Ranks to Personalize Web Search

Sofia Stamou (2009). *Handbook of Research on Digital Libraries: Design, Development, and Impact* (pp. 188-197).

www.irma-international.org/chapter/using-topic-specific-ranks-personalize/19881