Web Application for User Profiling

Taous Iggui, LIMED Laboratory, Department of Computer Science, University of A. Mira, Bejaia, Algeria
Hassina Nacer, Department of Computer Science, University of Science and Technology Houari Boumediene (USTHB) Algiers, Algeria
Youcef Sklab, LIMED Laboratory, Departement of Computer Science, University of A. Mira, Bejaia, Algeria
Taklit Ait Radi, Department of Computer Science, University of A. Mira, Bejaia, Algeria

ABSTRACT

User’s profiles play an important role when information systems try to meet their needs. This work presents a novel approach to build user profiles. It is based on information extraction techniques and proceeds by iterative steps. The use of different statistic metrics, Natural Language Processing (NLP) techniques and semantic descriptions (ontologies) in the authors’ approach, has provided it with a good precision degree when extracting information from texts. This has been demonstrated by an application prototype which is an automatic user profile constructor, using the texts of emails job applications (E recruitment field).

KEYWORDS

Information Extraction, Natural Language Processing, Ontologies, Semantic Web, Statistics, User Profile

1. INTRODUCTION

Information systems are, increasingly, accessible via Internet or Intranet, they allow users to reach an enormous mass of information from several and different sources. The most important challenge in current researches deals on how to optimize the performance and the accuracy of the results returned by applications and computer systems, in order to better satisfy the user’s needs (Ryan & Finn, 2005; Weib et al, 2008). In order to meet user’s needs (Harrathi, 2009), the most important challenge in current researches deals on how to optimize the performance and the accuracy of the results returned by applications and computer systems. The integration and consideration of the semantic aspect in information processing is extremely important to better understand user’s requests. To do this, a set of constraints is considered, and user’s profile (Schiaffino & Amandi, 2009) is a decisive one (Lee, 2009). It is a set of information concerning the user (personal data, education, professional information, preferences, interests …etc.).

The user profile is a crucial concept for various fields, such: e-recruitment, e-commerce, criminal domain, information retrieval, education, e-learning, security (Kaczmarek et al, 2010; Jeong & Choi, 2012; Satler et al, 2010) and many others. It’s related to several disciplines such as: Artificial Intelligence (AI), Language Processing, Text-mining and Semantic Web.

In this context and given the great impact of using the user profile by several systems and applications, the authors, in this paper, propose an information extraction system to construct user’s profile, from electronic e-mails of job applications (in the e-recruitment domain), by combining between several statistical metrics, inspired by works of (Bossard, 2010; Boudin & Torres, 2009) semantic Web technologies, as well as NLP (Natural language Processing) techniques and methods.
So, the system can take into account all the relevant information about user’s preferences and activities. The remainder of the document is organized as follows:

In section 2, the authors synthesize the most important works in the field of Information Retrieval (IR), Information Extraction (IE) and building a user profile. In Section 3, they propose their system architecture, entitled PBS (Profile Builder System), an automatic system to construct user’s profile and they detail the various component modules. Section 4 illustrates the evaluation of the proposed approach, by presenting results and performance measures of this one. Finally, conclusion and future works are described in Section 5.

2. RELATED WORKS

Information Extraction (IE) aims at exploring and exploiting various formats of data. It uses a set of techniques and methods to extract relevant information that can be conveyed by an information source. Several research studies have emerged, in this section we present a classification of the most relevant works according to their application field:

2.1. Automatic Summarization

An automatic summarization system aims at returning a condensed representation of an original text while keeping its semantic. In other words, it aims at extracting relevant information. Several summarization techniques have been developed and allow the construction of several summarization systems such (Torres & Rodriguez, 2010; Bossard, 2010; Boudin, 2008; Erkan & Radev, 2004), YACHS2, CORTEX and ROUGE2.

They use a set graph theory techniques and statistical metrics. Once assembled, they allow to assign a score to each sentence of the document, then, sentences having the highest scores are selected to be a part of the summary.

2.2. Information Retrieval

The arising problem in the field of information retrieval is about how to get relevant information to meet user’s needs according his preferences and interests.

Authors like (Sieg et al; 2007) have used user profiles for a personalized information search. A user profile is used in the results ranking step. In this area, generally, the user profile is represented by a set of key words, extracted from consulted documents by using the tf.idf measure (Tamine et al, 2007), ontologies (Daoud et al, 2008), and other knowledge bases like Wikipedia in (Nicoletti et al, 2012).

2.3. Information Extraction (IE)

Actually the intersection of methods and techniques, offered by IE discipline and the user profile concept, gives rise to the topic of building this user profile. Techniques of IE allow us to construct the user profile, from a set of relevant key words extracted from applications titles (Seguela et al, 2010), using a set of named entities (Omrane et al, 2011), from its comments in Web forums (El manar el bouanani & Kassou, 2012) by using a set of clustering techniques (K-means, EM), methods of stylometry as well as ontologies of speech.

In the recruitment area, (Kaczmarek et al, 2010) treats documents in Polish language, by exploiting extraction rules and an XML representation.
Reliability Analysis of Service Composition with Service Pools and Optimal Configuration of Service Pool Size
www.igi-global.com/chapter/reliability-analysis-service-composition-service/74237?camid=4v1a

Trust Management Tools
www.igi-global.com/chapter/trust-management-tools/30458?camid=4v1a