Information Systems 35 (2010) 774–790

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

SPGProfile: Speak Group Profile

Kamal Taha *, Ramez Elmasri

Department of Computer Science and Engineering, University of Texas, 701 S. Nedderman Drive, Arlington, TX 76019, United States

ARTICLE INFO

Article history:
Received 3 September 2009
Accepted 6 April 2010
Recommended by: F. Carino Jr.

Keywords:
Information retrieval
Information filtering
Recommendation system
XML

ABSTRACT

In this paper, we propose an XML-based recommender system, called SPGProfile. It is a type of collaborative information filtering system. SPGProfile uses ontology-driven social networks, where nodes represent social groups. A social group is an entity that defines a group based on demographic, ethnic, cultural, religious, age, or other characteristics. In the SPGProfile framework, query results are filtered and ranked based on the preferences of the social groups to which the user belongs. If the user belongs to social group G_x, results will be filtered based on the preferences of G_x and the preferences of *each* ancestor social group of G_x in the social network. SPGProfile can be used for various practical applications, such as Internet or other businesses that market preference-driven products. In the ontology, the preferences of a social group are identified from either: (1) the preferences of its member users or (2) from published studies about the social group. We describe and experimentally compare these two approaches. We also experimentally evaluate the search effectiveness and efficiency of SPGProfile and compare it to two existing search engines.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Traditional search engines typically return identical results for the same query, independent of the user or the context. Conventional quantitative scoring functions may not adequately reflect users' preferences, since the same document may be queried by users, whose preferences differ. By analyzing search behavior, it is possible to see that many users are not able to accurately express their needs in exact query terms [32]. In contrast to conventional search engines, a personalized search engine [22,6,43] would return different results for the same query, depending on the user and the context. Profiles can modify the representation of the user needs before the retrieval takes place. Most personalized systems lean towards being Information Filtering (IF) systems more than being general Information Retrieval (IR) systems [36].

Most existing personalized search systems do not consider group profiling. Group profiling can be an efficient retrieval mechanism, where a user profile is inferred from the profile of the social groups to which the user belongs. We propose in this paper an XML search system called SPGProfile (Speak Group Profile), which employs the concept of group profiling. SPGProfile simplifies the personalization process by pre-defining various categories of social groups and then identifying their preferences. The framework of SPGProfile categorizes social groups based on demographic, ethnic, cultural, religious, age, or other characteristics. For example, people of ethnic group E_X; people who follow religion R_Y; and people who live in neighborhood N_Y can all be considered to form various social groups. In social communities, it is commonly accepted that people who are known to share a specific background are likely to have additional connected interests [19]. SPGProfile can be used for various practical applications, such as Internet or other businesses that market

* Corresponding author.
E-mail addresses: kamal.taha@uta.edu (K. Taha), elmasri@uta.edu (R. Elmasri).

0306-4379/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.is.2010.04.001

preference-driven products. An individual user may belong to more than one social group. Therefore, SPGProfile outputs ranked lists of content items, taking into account not only the initial preferences of the user, but also the preferences of the user's various social groups. Consider for example a Mexican-American user. The user belongs to social groups Mexicans and Americans: the portion of Mexicans living in the USA. The results of a query submitted by this user will be *filtered* and *ranked* based on the union of the interests of social groups "Mexicans" and "Americans". The social groups to which a user belongs usually have *class–subclass* relationships. A subclass social group has its own properties while inheriting the properties of its superclass(s). For example, consider a user who belongs to the ethnic group "Berbers", which lives in the country of "Morocco", which is part of "North Africa". We could have the following representation of the hierarchical relationships between the three social groups: North Africans → Moroccans → Berbers. The Berbers may have their own concerns and preferences, while sharing the concerns and preferences of Moroccans, and more general concerns and preferences of North Africans. Thus, the user's query will be filtered and ranked based on the preferences of "Berbers", "Moroccans", and "North Africans".

In the framework of SPGProfile, the preferences of a social group could be identified from either: (1) the preferences of its member users or (2) from published studies about the social group (the availability of such data has significant boost with the emergence of World Wide Web). We describe and experimentally compare the two approaches in this study. SPGProfile helps searchers find what they need faster by trying to put the most relevant results at the top of the result list, a process known as relevance ranking. To the best of our knowledge, this is the first work advocating filtering and ranking results based on the preferences of social groups defined based on characteristics such as ethnicity, culture, religion, or the like. We make the following contributions in the paper:

- We experimentally evaluate the search effectiveness and efficiency of SPGProfile and compare it to two existing search engines. A demo of the SPGProfile system, running on 1000 MB grocery data and 1000 MB car data, is available at: http://dbse1.uta.edu/~kamal/?action=home.
- We propose ontology-driven social networks for modeling social groups.
- We propose *rule-chaining* (recursive querying) techniques for *filtering* query results based on the preferences of users' social groups.

2. Concepts used in the paper

In this section, we define key notations and basic concepts used in the paper. We use the term "domain" throughout the paper to mean an area of activity, belief, culture, ethnicity, demography, pursuit, or the like. A *Single-Domain Group* (*SDG*) is a group of people sharing common domain interests. For example, people of ethnic group E_X represents a *SDG*. We now formalize the concept of *SDG*.

Definition 1. Single-Domain Group (SDG):

A *SDG* is an aggregation G of individual users, where for each x, $y \in G$ $(x \neq y)$: x and y share a common and distinctive culture, ethnicity, religion, demography, language, or the like. That is, x and y share the same interests of *only one* domain group.

We use the term "system administrator" throughout the paper to mean a person employed to maintain and operate a system. The system administrator predefines SDGs, which are usually defined in publications such as: (1) published government's census and statistical studies (e.g., [42,30]) and (2) published studies conducted by specialized centers belonging to universities and organizations (e.g., [10,27,28,25]).

The smaller a social group is, the more granular and specific its interests are. Therefore, we introduce another class of social groups called a *Multi-Domain Group (MDG)*, whose size is usually smaller than a SDG. A *MDG* is formed from an aggregation of people sharing common *multi-domain* interests. Thus, a MDG is formed from the *intersection* of two or more SDGs. For example, the *portion* of ethnic group E_X who follow religion R_Y and live in neighborhood N_Y forms a MDG: the intersection of $E_X \cap R_Y \cap N_Y$. The interests of a MDG are the *union* of the interests of the SDGs forming it. Thus, the interests of a MDG are more specific than the interests of each of the SDGs forming it. To fine-grain a user's query results, SPGProfile outputs filtered and ranked list of items taking into account the preferences of the user's MDG. We now formalize the concept of *MDG*.

Definition 2. Multi-Domain Group (MDG):

Let S be the set of all *SDGs* that exist. A *MDG* is an aggregation G of individual users, where: $\forall x \in G$: x shares the same interests of $\exists s \in S$: $|s| \geq 2$. That is, G is formed from the intersection $\cap_{SDG_i \in s,\, s \subseteq S} SDG_i$. The interests of G are the union $\cup_{SDG_i \in s,\, s \subseteq S} interest(SDG_i)$.

SDGs forming a MDG usually have *class–subclass* relationships, where a subclass has its own properties while inheriting the properties of its superclass(s).

The preferences (interests) of a SDG G_x are stored in SPGProfile's database in the form of a *trigger rule*, called TrigRule(G_x). In response to a user's query, SPGProfile triggers the trigger rules of the SDGs forming the user's MDG. The trigger rules filter the XML tuples, retaining only those satisfying the preferences of the SDGs. The construct of a trigger rule is formed from the "WHERE" clause of XQuery [7]. That is, a trigger rule contains predicate Boolean conditions and these conditions are the preferences of a SDG. Fig. 1 shows a form of a trigger rule. The symbol Δ denotes either an XQuery's child operator '/' or a descendant operator '//'. The symbol $\square$ denotes an XQuery comparison operator. The letter P denotes a *preference* of a SDG corresponding to the *value* contained in an element labeled L in the XML document.

Download English Version:

https://daneshyari.com/en/article/396735

Download Persian Version:

https://daneshyari.com/article/396735

Daneshyari.com