Personalized Systems: Models and Methods from an IR and DB perspective

Yannis Ioannidis
University of Athens, Hellas
Tel: +30210-7275224
yannis@di.uoa.gr

Georgia Koutrika
University of Athens, Hellas
Tel: +30210-7275242
koutrika@di.uoa.gr

1. Introduction

In today's knowledge-driven society, information abundance and personal electronic device ubiquity have made it difficult for users to find the right information at the right time and at the right level of detail. To solve this problem, researchers have developed systems that adapt their behavior to the goals, tasks, interests, and other characteristics of their users. Based on *models* that capture important user characteristics, these *personalized systems* maintain their users' *profiles* and take them into account to customize the content generated or its presentation to the different individuals.

This tutorial will provide a comparative survey of user models and personalization methods emerging from the IR and database communities. It will also discuss open issues and challenges and will point to technology transfer opportunities between the two disciplines. The target audience includes researchers and practitioners in database and web-based systems and applications.

2. Outline

The tutorial will focus on three distinct issues:

(1) Content Personalization Methods: We will overview several forms of content personalization: (a) *Information filtering systems*: These screen out irrelevant data from incoming data streams and distribute relevant data items according to a user profile; (b) *Recommender systems*: These have automated the everyday procedure of relying on recommendations from other people whenever personal experience is not sufficient for making choices; (c) *Continuous queries*: These are issued only once and executed continuously over the database; (d) *Personalized searches*: These are based on the observation that "to enhance user searches one needs to take into account the fact that different people find different things relevant".

(2) User Modeling: User modeling refers to the representation of user characteristics and is the basis for every form of personalization support. State of the art user modeling typically captures cognitive characteristics of users, e.g., (changing) interests, skills and preferences. In this tutorial, we will mostly focus on models of user preferences, which have attracted particular attention in the research community and are categorized as follows: (a) *IR-based preference models*: These are for text data items and are based on natural language techniques. They have proved to be reasonably effective at representing user information needs; (b) *Database-based preference models*: Two approaches have been pursued: qualitative and quantitative. In the former, preferences between tuples in a query answer are specified using binary preference relations. In the later, preferences for queries are specified using scoring functions that associate a numeric score with every tuple of the query answer.

(3) User Profiling: This refers to the (implicit or explicit) process of collecting information about a user to generate an appropriate profile (based on the user model). We will discuss the following forms of user profiling methods: (a) *Relevance feedback*: Relevance feedback is typically used for query expansion during short-term modeling of a user's immediate information needs and for user profiling during long-term modeling of a user's persistent interests. Traditional methods require that users explicitly give feedback by, for example, specifying keywords, selecting and marking documents, or answering questions about their interests. (b) *Machine Learning*: Such approaches are used for learning personal profiles. Examples include neural networks, genetic algorithms, etc. (c) *Web Mining*: Web mining is more of a complete process rather than an algorithm. It uses data mining techniques to automatically discover and extract information from web documents and services. A common taxonomy of web mining defines three main research lines that will be presented: content mining, structure mining and usage mining.

Proceedings of the 31st VLDB Conference, Trondheim, Norway, 2005