Information Alert in Distributed Digital Libraries: The Models, Languages, and Architecture of DIAS*

M. Koubarakis, T. Koutris, C. Tryfonopoulos, and P. Raftopoulou

Dept. of Electronic and Computer Engineering
Technical University of Crete
73100 Chania, Crete, Greece
{manolis,koutris,trifon,rautop}@intelligence.tuc.gr
www.intelligence.tuc.gr/~manolis

Abstract. This paper presents DIAS, a distributed alert service for digital libraries, currently under development in project DIET. We first discuss the models and languages for expressing user profiles and notifications. Then we present the data structures, algorithms and protocols that underly the peer-to-peer agent architecture of DIAS.

1 Introduction

Users of modern digital libraries can keep themselves up-to-date by searching and browsing their favourite collections, or more conveniently by resorting to an alert service. Recently, the participants of project HERMES [15] have argued very convincingly that an alert service which integrates information from a wide variety of information providers can be indispensable to users. It relieves them from the tedious and cumbersome task of searching and browsing, or even from subscribing to individual alert services such as Springer Link Alert[1] or Elsevier Contents Direct2.

In this paper, we discuss the models, languages and architecture of DIAS, a Distributed Information Alert System currently under development in the context of the European project DIET [24][27][18]. DIAS adopts the basic ideas of project HERMES [15] and extends them in various interesting ways.

Our main technical contributions can be summarized as follows. First, we introduce the peer-to-peer (P2P) agent architecture of DIAS which has been inspired by the event dissemination system SIENA [6]. We also discuss the requirements imposed by this architecture on the data models and languages to

* This work was carried out as part of the DIET (Decentralised Information Ecosystems Technologies) project (IST-1999-10088), within the Universal Information Ecosystems initiative of the Information Society Technology Programme of the European Union.

1 http://link.springer.de/alert
2 http://www.elsevier.nl
be used for specifying user profiles/queries and notifications. Then we develop formally the data models $WP$, $AWP$ and $AWPS$, and their corresponding languages for specifying queries and notifications. Data model $WP$ is only briefly presented and more details can be found in [19,20]. $WP$ is based on free text and its query language is based on the boolean model with proximity operators. The concepts of $WP$ extend the traditional concept of proximity in IR [24,20] in a significant way, and utilize it in a query language targeted at information alert for distributed digital libraries. Data model $AWP$ is based on attributes or fields with finite-length strings as values. Its query language is an extension of the query language of data model $WP$. Our work on $AWP$ complements recent proposals for querying textual information in distributed event-based systems [6] by using linguistically motivated concepts such as word and not arbitrary strings. This makes $AWP$ very appropriate in information alert systems for digital libraries. Finally, the model $AWPS$ extends $AWP$ by introducing a “similarity” operator in the style of modern IR, based on the vector space model [2]. The novelty of our work in this area is the move to query languages much more expressive than the one used in the information dissemination system SIFT [24] where documents and queries are represented by free text. The similarity concept of $AWPS$ is essentially the similarity concept pioneered by IR systems [2], database systems with IR influences (e.g., WHIRL [11]) and more recently by the XML query language ELIXIR [10]. We note that both WHIRL and ELIXIR target information retrieval and integration applications, and pay no attention to information dissemination and the concepts/functionality needed in such applications. The first presentation of model $AWPS$ is the one given in this paper and [21].

In the second part of our paper, we present the detailed protocols and algorithms of our architecture and discuss its implementation in the context of project DIET [24,27,18]. Contrary to project HERMES [15], we develop a distributed information alert service from scratch based on the ideas of SIENA [6] and do not rely on any pre-existing message-oriented middleware.

The rest of the paper is organised as follows. Section 2 introduces the DIAS architecture, and discusses the requirements for data models and languages to be used in this context. Section 3 presents data model $WP$ and its semantics. Then Sections 4 and 5 build on this foundation and develops the same machinery for data models $AWP$ and $AWPS$. Section 6 discusses some interesting details of the DIAS architecture and implementation. Finally, Section 7 gives our conclusions and discusses future work.

---

3 In this paper, the terms query and profile will be used interchangeably. We are in an information alert setting where a profile is simply a long-standing query.

4 Also, for other commercial systems where similar models are supported already for retrieval.