Preference-Aware Publish/Subscribe Delivery with Diversity

Marina Drosou, Kostas Stefanidis, Evaggelia Pitoura

Department of Computer Science, University of Ioannina, Greece

{mdrosou, kstef, pitoura}@cs.uoi.gr

Abstract

In publish/subscribe systems, users describe their interests and are notified whenever new interesting events become available. Typically, in such systems, all subscriptions are considered equally important. However, due to the abundance of information, users receive overwhelming amounts of events. In this paper, we propose using a ranking mechanism based on user preferences, so that only top-ranked events are delivered to each user. Since many times, top-ranked events are similar to each other, we propose increasing the diversity of delivered events. Furthermore, we examine a number of different timing policies for delivering ranked events to users. We have fully implemented our approach in SIENA, a popular publish/subscribe middleware system, and report experimental results of its deployment.

1 Introduction

With the explosion of the amount of information that is currently available online, publish/subscribe systems offer an attractive alternative to searching by providing a proactive model of information supply. In such systems, users express their interest in specific pieces of data or events via *subscriptions*. Then, they are *notified* whenever some other user generates (or *publishes*) an event that *matches* one of their subscriptions. Typically, all subscriptions are considered equally important and users are notified whenever a published event matches any of their subscriptions.

However, getting notified about all matching events may lead to overwhelming the users with huge amounts of notifications hurting the acceptability of publish/subscribe systems. To control the rate of notifications received by the subscribers, it would be useful to allow them to rank the importance or relevance of events. Then, they would only receive notifications for the most important or relevant among them. For example, take a user John that generally likes drama movies but prefers drama movies directed by T. Burton to drama movies directed by S. Spielberg. Ideally, John would like to receive notifications about S. Spielberg drama movies only if there are no, or not enough, notifications about T. Burton drama movies.

In this paper, we propose extending subscriptions to allow users express the fact that some events are more important or relevant to them than others. To indicate priorities among subscriptions, we introduce *preferential subscriptions*. In general, there are two basic approaches to specifying preferences among items: a quantitative and a qualitative approach. In the *quantitative approach* (e.g. [2, 7, 8]), users employ scoring functions that associate a numeric score with specific data items to indicate their interest in them. In the *qualitative approach* (e.g. [4, 5, 6]), preferences between two data items are specified directly, typically using binary preference relations. We show how to formulate preferences among subscriptions using each one of these approaches. Events are ranked so that an event that matches a highly preferred subscription is ranked higher than an event that matches a subscription with a lower preference.

Based on preferential subscriptions, we introduce a top-k variation of the publish/subscribe paradigm in which users receive only the matching events having the k highest ranks as opposed to all events matching their subscriptions. Since the delivery of events is continuous, we also introduce a number of timing policies that determine the range of events over which the top-k computation is performed.

However, the top-k events are often very similar to each other. Besides pure accuracy achieved by matching the criteria set by the users, diversification, i.e. recommending items that differ from each other, has been shown to increase user satisfaction [9]. For instance, our user John would probably like to receive information about different movies by T. Burton as well as a couple of S. Spielberg movies once in a while. To this end, we adjust the top-k computation to take also into account the *diversity* of the delivered events. To achieve this, we consider both the importance of each event as specified by the user preferences as well as its diversity from the other top-ranked events. We examine how the results can be diversified for each of the timing policies.

As a proof-of-concept, we have implemented a prototype, termed PrefSIENA [1]. PrefSIENA extends SIENA [3], a popular publish/subscribe middleware system, with preferential subscriptions, timing policies and diversity towards achieving top-k event delivery. We present a number of experimental results that evaluate the number of events delivered by PrefSIENA with respect to the original SIENA system, as well as the rank, freshness and diversity of such events. We also report on the overheads of supporting top-k delivery.

References

[1] *PrefSIENA*. http://www.cs.uoi.gr/∼mdrosou/PrefSIENA.

[2] R. Agrawal and E. L. Wimmers. A framework for expressing and combining preferences. *SIGMOD Rec.*, 29(2):297–306, 2000.

[3] A. Carzaniga, D. S. Rosenblum, and A. L. Wolf. Design and evaluation of a wide-area event notification service. *ACM Trans. on Computer Syst.*, 19:332–383, 2001.

[4] J. Chomicki. Preference formulas in relational queries. *ACM Trans. Database Syst.*, 28(4):427–466, 2003.

[5] P. Georgiadis, I. Kapantaidakis, V. Christophides, E. M. Nguer, and N. Spyratos. Efficient rewriting algorithms for preference queries. In *ICDE*, pages 1101–1110, 2008.

[6] W. Kießling. Foundations of preferences in database systems. In *VLDB*, pages 311–322, 2002.

[7] G. Koutrika and Y. Ioannidis. Personalized queries under a generalized preference model. In *ICDE*, pages 841–852, 2005.

[8] K. Stefanidis, E. Pitoura, and P. Vassiliadis. Adding context to preferences. In *ICDE*, pages 846–855, 2007.

[9] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *WWW*, pages 22–32, 2005.