Abstract. Traditional search techniques frequently fail the average user in their quest for online information. Recommender systems attempt to address this problem by discovering the context in which the search occurs. Though effective, these systems are often hampered by the brevity of typical user queries. In this paper we describe CASPER, an online recruitment search engine which combines similarity-based search with a client-side personalisation technique. In particular we argue that CASPER's personalisation strategy is effective in determining retrieval relevance in the face of incomplete queries.

Keywords: case-based search, personalisation, incomplete queries.

1 Introduction

As the Web expands, it’s becoming progressively more difficult for its users to find the information they want. Most use search engines to cope with the vast volumes of information currently available. These search engines often rely on traditional database or information-retrieval inspired strategies to fulfill search requests. Users often find it difficult to cope with these technologies as most have little experience in their use and may not be very knowledgeable about the domain within which they are searching. This often results in short, ambiguous and incomplete search queries that are lacking in information [19]. All of this is exacerbated by the limited screen real estate of modern mobile web-enabled devices and the reluctance of users to venture beyond the first two or three results presented to them. Recommender systems have emerged as one possible solution to this problem. By combining ideas from machine learning, user profiling and case-based reasoning, recommender systems promise more focused, intelligent and proactive search that is better adapted to the needs of the individual.

In section 2 we review some popular recommendation strategies adopted by recommender systems. In section 3 we present our own experience with
CASPER, an online recruitment search engine. CASPER comprises two sub systems: a content-free Automated Collaborative Filtering (CASPER ACF) [14, 15] system and a two-stage personalisation client that combines case-based search with a diversity aware personalisation technique (CASPER PCR). We focus on this personalisation client, in particular we argue that its personalisation strategy is effective in determining retrieval relevance in the face of incomplete queries and benefits from increased privacy compared to other approaches. Section 4 describes a live user trial of the system and the experimental analysis of the accuracy of the predictions made by CASPER PCR. Finally, we present our conclusions in section 5.

2 Background

The emergence of recommender systems has heralded an attempt to improve the quality of search results by determining the context within which the search takes place. There are two main sources of information available to a recommender system and hence two main approaches to improving the relevance of search results in the face of vague queries. Understanding the context of the query itself can reduce the ambiguity of incomplete queries and allow searches to be more focussed. Alternatively knowing more about the user can provide us with valuable preference information.

2.1 Query Context

Contextual information about the query may be obtained explicitly, through interaction with users, or implicitly, by observing their natural behaviour.

Explicit Context. One way to capture explicit context is to force users to search within a specific context. For example, category-based search engines such as Yahoo (http://www.yahoo.com) improve their search by limiting the set of documents to which a query is compared. These pre-defined categories provide context for the query but recommendations are necessarily restricted to a small percentage of the total available due to the tedious nature of authoring categories. CiteSeer [9] uses a similar approach by limiting its search to scientific papers. Its document index is built automatically by crawling the web for scientific articles in various formats and by taking advantage of the cross-referencing common to scientific documents. Categories are used in Inquirius 2 [8] not to restrict search but rather to enhance a query. Upon selection of a search category, a set of words associated with that category are added to the user’s query. For example choosing the ‘research paper’ category adds terms frequently found in research papers such as ‘abstract’ and ‘references’. A full search of the web is then performed with this enhanced query.

Implicit Context. Of course users are loathe to expend time and effort providing extra information for their searches. In addition differences in knowledge and vocabulary between users and indexers may prove frustrating. In order for these