The Challenge of Creative Information Retrieval

Tony Veale

Department of Computer Science,
University College Dublin, Belfield, Dublin 6, Ireland.
Tony.veale@UCD.ie

Abstract. Information retrieval (IR) is an effective mechanism for text management that has received widespread adoption in the world at large. But it is not a particularly creative mechanism, in the sense of creating new conceptual structures or reorganizing existing ones to pull in documents that describe, in novel and inventive ways, a user's information needs. Since language is a dynamic and highly creative medium of expression, the concepts that one seeks will therefore represent a moving target for IR systems. We argue that only by thinking creatively can an IR system effectively retrieve documents that express themselves creatively.

1   Introduction

Most retrieval of textual information is literal in the sense that any retrieved document will literally match the keywords of the user's initial query. The query, whether a simple bag of conjoined keywords or a complex Boolean filter, essentially specifies the indices that should be examined to find matching documents. The set of matching documents is thus circumscribed by the keywords chosen by the user, making retrieval vulnerable to the *word mismatch problem* [2] if the authors of the most relevant documents have chosen to lexicalize their ideas in a different way. Of course, statistical and knowledge-based techniques (e.g., [3, 4, 5]) can be used to expand a query with highly correlated terms to permit the retrieval of additional relevant documents that do not literally contain any of the initial query terms. However, even these techniques still operate on the literal plane of meaning, by focusing on the conventional meaning of the keywords used (e.g., by using their synonyms, hypernyms and hyponyms).

This literal mindset in information retrieval (IR) ignores the fact that language is a creatively dynamic medium, one that is always striving to find new ways to communicate the same old ideas, often with an additional connotation or a different spin [6]. So while users of IR may be relentlessly literal in their choice of search terms, the authors of the documents they are hoping to retrieve will frequently be far more creative in their choice of words. To successfully retrieve these documents, it will be necessary for IR systems to demonstrate an equal level of creativity, to predict the innovative ways in which a relevant document might speak to the information needs of the user. These predictive techniques should be creative in the sense that they are capable of reorganizing an existing conceptual worldview (modeled using a taxon-