ABSTRACT

Most search engines do their text query and retrieval using keywords. However, vendors cannot anticipate all possible ways in which shoppers search for their products. In fact, many times, there may be no direct keyword match between a search phrase and descriptions of products that are perfect hits for the search. A highly automated solution to the problem of bridging the semantic gap between product descriptions and search phrases used by Web shoppers is developed. By using scalable information extraction techniques from Web sources and a frequent itemset mining algorithm, our system can learn how meanings can be ascribed to popular search phrases with dynamic connotations. By annotating the product databases based on the meanings of search phrases mined by our system, catalog owners can boost the findability of their products.

Keywords: Web mining; phrase mining; frequent itemset mining; labeling; information extraction

INTRODUCTION

The World Wide Web has made a dramatic transition from its early beginnings as a distributed repository of browsable information into a dominant medium for conducting e-commerce. In particular, it has become a mainstream shopping medium for retail goods. Making purchases has emerged as a pervasive online activity. Keyword searching is the most common form of product search on the Web. Most search engines do their text query and retrieval using keywords. The average keyword query length is less than three words (2.2 words) (Cutting & Douglas, 1997). Recent research (Andrews, 2003) found that 40% of e-commerce companies rate their search tools as not very useful or only somewhat useful. Further, a review of 89 sites (Andrews, 2003) found that 75% have keyword search engines that fail to retrieve important information and to put...
results in order of relevance; 92% fail to provide guided search interfaces to help offset keyword deficiencies (Andrews, 2003), and seven out of 10 Web shoppers were unable to find products using the search engine, even when the items were stocked and available.

The Problem Definition

However, vendors cannot anticipate all possible ways in which shoppers search for their products. In fact, many times, there may be no direct keyword match between a search phrase and descriptions of products that are perfect hits for the search. For example, if the shopper uses motorcycle jacket, then unless the publisher knows that every leather jacket is a motorcycle jacket, it can not produce all matches for the user’s search. Thus, for certain phrases, there is a semantic gap between the search phrase used and the way the corresponding matching products are described. A serious consequence of this gap is that it results in unsatisfied customers. Thus there is a critical need to boost product findability by bridging the semantic gap that exists between search phrases and product descriptions. Closing this gap has a great potential to translate Web traffic into higher conversion rates and more satisfied customers.

Issues in Bridging the Semantic Gap

We denote a search phrase to be a target search phrase if it does not directly match certain relevant product descriptions. The semantics of products matching such target search phrases is implicit in their descriptions. For phrases with fixed meanings (i.e., their connotations do not change, such as in animal print comforter), it is possible to close the gap by extracting their meanings with a thesaurus and relating it to product descriptions, such as zebra print comforter, leopard print bedding, and so forth. They pose a more interesting challenge when their meanings are subjective or driven by perceptions, and hence, their connotations change over time, as in the case of fashionable handbag and luxury bedding. The concept of a fashionable handbag is based on trends that change over time, and correspondingly, the attribute values characterizing such a bag also change. Similarly, the concept of luxury bedding depends on the brands and designs available on the market that are considered luxury and their attributes. Bridging the semantic gap, therefore, is, in essence, the problem of inferring the meaning of search phrases in all their nuances.

Our Approach

In this article, we present an algorithm that (1) finds products matching a target search phrase from Web documents by utilizing search engines, (2) extracts and structures product attributes and product descriptions for these products, and (3) uses an optimized frequent itemset mining algorithm to learn the target phrase definitions.

Our Contributions

We present a novel Web-mining approach for inferring the meaning of search phrases from keywords matching product information. The mined rules can be used by a publisher or a search engine in order to boost the item findability. We present algorithms for the following:

- Automated item information extraction from heterogeneous Web sources
An Intelligent Particle Swarm Optimization for Fuzzy Based Heterogeneous Radio Access Technology (RAT) Selection


[www.igi-global.com/article/intelligent-particle-swarm-optimization-fuzzy/74828?camid=4v1a](www.igi-global.com/article/intelligent-particle-swarm-optimization-fuzzy/74828?camid=4v1a)