Chapter 18
A Genetic Fuzzy Semantic Web Search Agent Using Granular Semantic Trees for Ambiguous Queries

Yan Chen
Georgia State University, USA

Yan-Qing Zhang
Georgia State University, USA

ABSTRACT
For most Web searching applications, queries are commonly ambiguous because words or phrases have different linguistic meanings for different Web users. The conventional keyword-based search engines cannot disambiguate queries to provide relevant results matching Web users’ intents. Traditional Word Sense Disambiguation (WSD) methods use statistic models or ontology-based knowledge systems to measure associations among words. The contexts of queries are used for disambiguation in these methods. However, due to the fact that numerous combinations of words may appear in queries and documents, it is difficult to extract concepts’ relations for all possible combinations. Moreover, queries are usually short, so contexts in queries do not always provide enough information to disambiguate queries. Therefore, the traditional WSD methods are not sufficient to provide accurate search results for ambiguous queries. In this chapter, a new model, Granular Semantic Tree (GST), is introduced for more conveniently representing associations among concepts than the traditional WSD methods. Additionally, users’ preferences are used to provide personalized search results that better adapt to users’ unique intents. Fuzzy logic is used to determine the most appropriate concepts related to queries based on contexts and users’ preferences. Finally, Web pages are analyzed by the GST model. The concepts of pages for the queries are evaluated, and the pages are re-ranked according to similarities of concepts between pages and queries.

DOI: 10.4018/978-1-60566-324-1.ch018
1. INTRODUCTION

Nowadays, Web search engines play a key role in retrieving information from the Internet to provide useful Web documents in response to users’ queries. The keywords-based search engines, like GOOGLE, YAHOO Search and MSN Live Search, explore documents by matching keywords in queries with words in documents. However, some keywords have more than one meaning, and such words may be related to different concepts in different contexts, so they are potentially ambiguous. Since current search engines simply search keywords separately and do not consider the contexts of queries, word sense ambiguity may result in searching errors for Web search applications. For example, if a user searches “drawing tables in a document” by MSN Live Search, five useless results related to the furniture table will be shown in the first result page. Therefore, an exact concept of a query may be determined by the contexts. Moreover, queries are usually short and contexts in queries do not always provide enough information for disambiguating queries. Under these circumstances, users’ preferences may be helpful for determining an appropriate concept for an ambiguous word. For an example, if a biologist searches “mouse”, we can speculate that the biologist is interested in Web pages related to a rodent “mouse” instead of a computer device “mouse.” Thus, both contexts of queries and users’ preferences are useful for disambiguating queries in Web search applications.

In fact, Query Disambiguation (QD) is a special application of Word Sense Disambiguation (WSD) problems. For most WSD problems, usually a set of possible meanings for a word is known ahead of time, and stored in a lexical database. Then, the meaning for the word is assigned depending on its contexts.

In this chapter, we propose a new model called Granular Semantic Tree to represent semantic relations between concepts. Each concept is represented as one granule in the tree structure. If concept A contains concept B, then granule B is represented as a child of granule A. Thus, a granular semantic tree that contains hierarchical structures can be constructed. Then, any concepts’ relations in the granular semantic tree can be evaluated based on the hops between them. The exact concept for an ambiguous word is assigned depending on the concepts of its nearby words.

This chapter firstly discusses conventional effective methods for solving QD problems. Different from those methods, the GST model for easily expressing relations among words in contexts is proposed. Then, fuzzy logic is used for determine the most appropriate concepts related to queries based on contexts and users’ interests. Experiments and evaluations are given. Finally conclusions are described.

2. RELATED WORKS

Based on the theory of granular computing, the granules, such as subsets, classes, objects, and elements of a universe, are basic ingredients of granular computing (Yao, 2007). The general granules are constructed by grouping finer granules based on available information and knowledge, such as similarity or functionality. The term-space granulation is used in the information retrieval and WSD areas (Yao, 2003). Terms are basic granules and term hierarchy is constructed by clustering terms. Then, new terms may be assigned to clusters as labels. Usually, those labels are more general than the terms in the cluster. The notion of term-space granulation serves as an effective tool for the QD applications. Many researchers have used term clustering with its application in disambiguating queries.

One of the frequently used techniques for term clustering is Statistical Association method (Brena and Ramirez, 2006). Through measuring the co-occurrences of words in the large quantities of Web