Chapter 4.15
Dynamic Taxonomies

Giovanni M. Sacco
Università di Torino, Italy

INTRODUCTION

End-user interactive access to complex information is one of the key functionalities of knowledge management systems. Traditionally, access paradigms have focused on retrieval of data on the basis of precise specifications: examples of this approach include queries on structured database systems, and information retrieval. However, most search tasks, and notably those typical of a knowledge worker, are exploratory and imprecise in essence: the user needs to explore the information base, find relationships among concepts, and thin alternatives out in a guided way.

Examples of this type of access include the selection of the "right" product to buy, of a candidate for a job, but also finding the likely cause of a malfunction, and so forth. Indeed, exploratory access applies to an extremely wide range of practical situations. Traditional access methods are not helpful in this context, and new access paradigms are needed. Effective end-user access requires a holistic approach, in which modeling, interface, and interaction issues are considered together.

BACKGROUND

Since the vast majority of knowledge is textual and unstructured in nature, information retrieval (IR) techniques (van Rijsbergen, 1979) have been extensively used in the past. IR techniques require almost no editorial work or manual preprocessing of information. However, their limitations have been known for some time: a study on a legal environment reported that only 20% of relevant documents were actually retrieved (Blair & Maron, 1985). Such a significant loss of information is due to the extremely wide semantic gap between the user model (concepts) and the model used by commercial retrieval systems (words). IR systems are also poor from the point of view of user interaction because the user has to formulate his query with no or very little assistance. Finally, results are presented as a flat list with no systematic organization, so that browsing/exploring the knowledge base is impossible.

Hypermedia (see Groenbaek & Trigg, 1994) addresses the problem of browsing/exploration, but it has a number of serious drawbacks: there is no systematic picture of relationships among knowledge base components; exploration is performed one document at a time, which is quite time consuming; and building and maintaining complex hypermedia networks is very expensive.

Traditional taxonomies are used by many systems, such as Yahoo. Here, a hierarchy of concepts can be used to select areas of interest and restrict the portion of the infobase to be retrieved. Taxonomies support abstraction and are easily understood by end-users. However, they are not scalable for large knowledge bases (Sacco, 2002) because they can be used for discrimination just down to terminal concepts, which are no further specialized. As the knowledge base grows, the average number of documents associated to a terminal concept becomes too large for manual inspection.

Solutions based on semantic networks were proposed in the past (e.g., Schmeltz Pedersen, 1993) and are being reconsidered in the current effort on ontologies and Semantic Web. Although more powerful and expressive than plain taxonomies, general semantic schemata are difficult to understand and manipulate by the casual user. They are better suited to programmatic access, and user interaction must be mediated by specialized agents. This increases costs, time to market, and decreases generality and flexibility of user access.

MAIN FOCUS OF THE ARTICLE

Dynamic taxonomies (Sacco, 1987, 2000; also known as faceted classification systems) are a general knowledge management model based on a multidimensional classification of heterogeneous data items and are used to explore/browse complex knowledge bases in a guided yet unconstrained way through a visual interface.

The intension of a dynamic taxonomy is a taxonomy designed by an expert. This taxonomy is a concept hierarchy going from the most general to the most specific concepts. Directed acyclic graph taxonomies modeling multiple inheritance are supported but rarely required. A dynamic taxonomy does not require any other relationships in addition to subsumptions (e.g., IS-A and PART-OF relationships).

In the extension, items can be freely classified under n (n>1) topics at any level of abstraction (i.e., at any level in the conceptual tree). This multidimensional classification is a departure from the monodimensional classification scheme used in conventional taxonomies. Besides being a generalization of a monodimensional classification, a multidimensional classification models common real-life situations. First, items are very often about different concepts: for example a news item on September 11, 2001, can be classified under "terrorism," "airlines," "USA," and so forth. Second, items to be classified usually have different features, "perspectives," or facets (e.g., Time, Location, etc.), each of which can be described by an independent taxonomy.

By taking a "nominalistic" approach—that is, concepts are defined by instances rather than by properties—a concept C is just a label that identifies all the items classified under C. Because of the subsumption relationship between a concept and its descendants, the items classified under C (items (C)) are all those items in the deep extension (Straube & Ozsu, 1990) of C; that is, the set of items identified by C includes the shallow extension of C (i.e., all the items directly classified under C) union the deep extension of C's sons. By construction, the shallow and the deep extension for a terminal concept are the same.

There are two important consequences of this approach. First, since concepts identify sets of items, logical operations on concepts can be performed by the corresponding set operations

Related Content