ABSTRACT

In this paper, we present an original approach for text summarization using conceptual data classification. We show how a given text can be summarized without losing meaningful knowledge and without using any semantic or grammatical concepts. In fact, concept date classification is used to extract the most interacting sentences from the main text and ignoring the other meaningless sentences in order to generate the text summary. The approach is tested on Arabic and English texts with different sizes and different topics and the obtained results are satisfactory. The system may be incorporated with the indexers of search engines over the Internet in order to find key words and other pertinent information of the new deployed Web pages that would be stored in databases for quick search.
INTRODUCTION

Most of intelligent information retrieval methods need to realize data classification and make minimization of their representation in the memory. The word classification means pattern generation and recognition. Formal concept analysis offers simple, original, uniform, and mathematically well-defined methods for data clustering. In fact, we can easily associate a pattern with a formal concept (i.e., a set of objects sharing the maximum of properties). In this article, we are interested in text summarization using conceptual data classification. This new approach allows us to extract the most significant and interacting sentences in a given text in order to generate its summary. It can be applied on any text of any language—Arabic, English, French, etc.—as it doesn’t use any semantic or grammatical concepts. The idea consists of minimizing the information representation by selecting only “optimal concepts.” We assume that data may be converted into a binary relation as a subset of the product of a set of objects and a set of properties. Due to the huge amount of data contained in most existing documents and databases, it becomes important to find a priority order for concept selections to enable the users to find pertinent information from a given document. A search engine over the Internet can be extended with such a system in order to select the most appropriate documents and Web pages that match with the user’s queries. In fact, the system can be used as an indexer of the search engine in order to select the most appropriate documents and Web pages that match with the user’s queries. We mention that we have used this approach in our previous works to supervise learning and we obtained defendable results in terms of error rate comparing with other well known methods (for more details you can see Al-Rashdi, Al-Muraikhi, Al-Subaiey, Al-Ghanim, & Al-Misaifri, 2001; Ben Yahia, Arour, Slimani, & Jaoua, 2000; Ben Yahia & Jaoua, 2001; Jaoua & Elloumi, 2002; Maddouri, Elloumi, & Jaoua, 1998; Mineau & Godin, 1995).

The article is organized as follows: in the second section, we present the mathematical foundation of concept analysis. In the third section, we give a polynomial approximate algorithm to the NP-complete problem of the coverage of a binary context with a minimal number of optimal concepts. In the fourth section, we explain how to apply the idea of optimal concept (called also optimal rectangle) on text summarization for English and Arabic texts. In the fifth section, we show the obtained results of our experiments, while in the sixth section we conclude the article.

MATHEMATICAL FOUNDATIONS

Among the mathematical theories found recently with important applications in computer science, lattice theory has a specific place for data organization, information engineering, data mining, and reasoning. It may be considered as the mathematical tool that unifies data and knowledge, as well as information retrieval and reasoning (Davey & Priestley, 1990; Ganter & Wille, 1999; Jaoua, Bsais, Consmtini, 1999; Jaoua, Boudriga, Durieux, & Mili, 1991; Schmidt & Ströhlein, 1989). We will define the binary context, the formal concept, and the lattice of concepts associated with a binary context.

Definition 1 (Binary Context): A binary context (or binary relation) is a subset of the product of two sets $O$ (the set of objects) and $P$ (the set of properties).

Example: Let $O = \{\text{Leech, Bream, Frog, Dog, Spike-weed, Reed, Bean, Maize}\}$, and let $P = \{a, b, c, d, e, f, g, h, i\}$.

where $O$ is a set of some animals, and $P$ is a set of properties displayed in Table 1.
Related Content

A Case Study for External Users
www.igi-global.com/chapter/case-study-external-users/8171?camid=4v1a

A Unified Approach to Uncertainty-Aware Ubiquitous Localisation of Mobile Users
www.igi-global.com/article/unified-approach-uncertainty-aware-ubiquitous/65067?camid=4v1a

Toward Mobile Web 2.0-Based Business Methods: Collaborative QoS-Information Sharing for Mobile Service Users
www.igi-global.com/chapter/toward-mobile-web-based-business/37701?camid=4v1a

eDAR Algorithm for Continuous KNN Queries Based on Pine
www.igi-global.com/article/edar-algorithm-continuous-knn-queries/2615?camid=4v1a