AL-FARABI KAZAKH NATIONAL UNIVERSITY

Aliya Nugumanova, Madina Mansurova

AUTOMATIC RECOGNITION OF TERMS IN TEXTS IN NATURAL LANGUAGE

Monograph

Almaty

"Kazakh University"

2020

Mansurova Madina, Nugumanova Aliy,

Automatic recognition of terms in natural language texts: monograph / Mansurova Madina, Nugumanova Aliya. - Almaty: Kazakh University, 2020. – 96 p.

The automatic extraction of terms from subject matter texts is a task that has many applications. Automatically extracted terms can be used as classification features for categorizing documents, as semantic concepts for generating thesauri and ontologies, as basic concepts for media content analysis. Almost all automatic word processing tasks, such as annotation, indexing, classification, machine translation, knowledge extraction, etc., require terminology extraction. To solve this problem, a large number of effective methods have been developed that allow automating the extraction of terms from the texts of the subject area. This monograph is intended to characterize and systematize these methods in order to determine the prospects and problems of the development of natural language processing technologies. The significance of the identified problems is due to the phenomenon of the information explosion that modern society is experiencing. For researchers, specialists in the field of natural language processing, university teachers, doctoral students, undergraduates and students.

Introduction

Automatic term recognition based on natural language texts is both a complex and popular scientific task that has many applications in information retrieval and knowledge engineering. The solution to this problem is directed to the automatic formation of the terminological lexicon of any subject area. This problem statement emerged as an alternative to the traditional manual term extraction, in which a terminology specialist initially compiles a list of candidate terms, and then consults with a subject matter expert to validate the final vocabulary.

The relevance of the task of automatic recognition of terms is determined by the fact that in a rapidly changing world, where new technological industries are constantly appearing, new concepts and terms appear, and the volume of technical vocabulary increases exponentially, manual construction and description of terminology is a labor-intensive enterprise. Therefore, the methods of highlighting terms using software tools are of great theoretical and practical value.

The lists of terms extracted by software methods are the most simple machine-readable structures of knowledge, but despite their simplicity, their role in information retrieval and other practical applications can hardly be overestimated. With the help of machine-readable lists of terms, you can annotate and index documents without involving experts, set their thematic focus (categorize) and support machine translation. Also