Ngoc Thanh Nguyen · Richard Chbeir ·
Ernesto Exposito · Philippe Aniorté ·
Bogdan Trawiński (Eds.)

LNAI 11684

Computational Collective Intelligence

11th International Conference, ICCCI 2019
Hendaye, France, September 4–6, 2019
Proceedings, Part II

2 Part II

Springer

Lecture Notes in Artificial Intelligence **11684**

Subseries of Lecture Notes in Computer Science

Ngoc Thanh Nguyen · Richard Chbeir ·
Ernesto Exposito · Philippe Aniorté ·
Bogdan Trawiński (Eds.)

Computational Collective Intelligence

11th International Conference, ICCCI 2019
Hendaye, France, September 4–6, 2019
Proceedings, Part II

Springer

Editors
Ngoc Thanh Nguyen
Ton Duc Thang University
Ho Chi Minh City, Vietnam

Wrocław University of Science
and Technology
Wrocław, Poland

Ernesto Exposito
University of Pau and Pays de l'Adour
Pau, France

Bogdan Trawiński
Wrocław University of Science
and Technology
Wrocław, Poland

Richard Chbeir
University of Pau and Pays de l'Adour
Pau, France

Philippe Aniorté
University of Pau and Pays de l'Adour
Pau, France

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-030-28373-5 ISBN 978-3-030-28374-2 (eBook)
https://doi.org/10.1007/978-3-030-28374-2

LNCS Sublibrary: SL7 – Artificial Intelligence

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Problems of Semantics of Words of the Kazakh Language in the Information Retrieval

Rakhimova Diana[1(✉)] and Shormakova Assem[2(✉)]

[1] Al-Farabi KazNU, Street Al-Farabi 71, Almaty, Kazakhstan
di.diva@mail.ru

[2] Institute of Information and Computational Technologies, Street Pushkin 125, Almaty, Kazakhstan
shormakovaassem@gmail.com

Abstract. The theme of this research is intelligent search engines that can search and extract new information from text data in Kazakh language and education. The significance of the research topic due to the growing amount of data represented in digital form, which provide the ability to access various sources of electronic documents. The use of intelligent search engines will allow you to meet the information needs of users. In this regard, the development of information-analytical search engines that allows you to work with data in Kazakh language is relevant. The goal of this research is to develop efficient algorithms and models for intelligent search systems, based on modern technologies in the field of information retrieval and natural language processing teaching them.

Keywords: Kazakh languages · Semantics · Text · Information retrieval

1 Overview Modern Text Information Retrieval

Kazakh language belongs to the languages of agglutinative type. Agglutinative languages (from the Latin agglutinatio - gluing) are languages with a system in which agglutination is the dominant type of inflection ("gluing" of different formants (suffixes or prefixes)), each of which carries only one meaning. Kazakh language is very rich in various words forms and its semantics.

The problem of finding a document that meets certain criteria, occurs in any data warehouse that contains more than one document. It is obvious, that the solution of this problem is, somehow, confined to those which are used in the design of storage systems. One can specify two basic ways: (1) using a hierarchical model; (2) the use of hypertext models.

The use of a hierarchical multilevel model involves the categorization of information resources. To select the path to the desired document uses the description drawn up by the support of this system. Hypertext model allows to link document's links, which are located directly in the text. These two models have obvious drawbacks. As multi-level categorization, and the placement of links is performed by highly qualified specialists. And the volume treated in this document may not be very large. For this reason, suffers the relevance of the description in the array of documents. In addition,

N. T. Nguyen et al. (Eds.): ICCCI 2019, LNAI 11684, pp. 70–81, 2019.
https://doi.org/10.1007/978-3-030-28374-2_7