Studies in Computational Intelligence 598

Dariusz Barbucha
Ngoc Thanh Nguyen
John Batubara *Editors*

New Trends in Intelligent Information and Database Systems

Springer

Part II: Intelligent Computational Methods in Information Systems

Part III: Semantic Web, Social Networks and Recommendation Systems

Part IV: Cloud Computing and Intelligent Internet Systems

Part V: Knowledge and Language Processing

Finite-State Transducers with Multivalued Mappings for Processing of Rich Inflectional Languages

Ualsher Tukeyev[1], Marek Miłosz[2], and Zhandos Zhumanov[1]

[1] Al-Farabi Kazakh National University, Almaty, Kazakhstan
{ualsher.tukeyev,z.zhake}@gmail.com
[2] Lublin University of Technology, Lublin, Poland
marekm@cs.pollub.pl

Abstract. This paper proposes a processing for rich inflectional languages such as Russian and Kazakh, based on finite state transducers with multivalued mappings. We propose to simplify grammar of inflectional languages and use multivalued mappings. An advantage of finite state transducers with multivalued mappings proposed in this paper is that it automatically generates possible alternatives of words' grammatical characteristics, while in existing rule-based technologies alternatives are written by hand. Ambiguity of grammatical characteristics is solved by comparing alternative grammatical characteristics between adjacent words in the source sentence and matching grammatical characteristics are selected. Here an advantage of proposed method should be noted, it does not require explicit description of matching agreements for grammatical characteristics of adjacent words in a sentence, as it is done in existing rule-based methods.

Keywords: Finite-state transducers, multivalued mappings, computational intelligence, machine translation, natural language processing.

1 Introduction

In the field of languages processing (machine translation in particular) the problem of quality still remains a key topical issue.

In this paper, in line with the development of efficient technologies for languages processing, we propose using of the multivalued mappings theory [1]. The theory of multivalued mappings has been actively developed in the past 30 years, especially in game theory, theory of extreme problems, mathematical economics [2]. In the field of language processing there are examples of using the method of mappings in machine translation [3, 4], but they were not multivalued mappings. Proposed method is closed to finite-state transducers (FST) [5, 6] and especially close to a kind of FST transducers named *p*-**subsequential** transducers which allow ambiguity of final output strings to be associated with each final state [7]. Also there are many papers using FST for machine translation, in particular, for morphology of agglutinative languages [8, 9, 10, 11]. Approach proposed in this paper is based on simplifying morphology of investigated languages' words for improving of computational processing of machine translation by using of finite state transducers with multivalued mappings.

D. Barbucha et al. (eds.), *New Trends in Intelligent Information and Database Systems*,
Studies in Computational Intelligence 598, DOI: 10.1007/978-3-319-16211-9_28