Methodology for building extraction templates for Russian language in knowledge-based IE systems

Solovyev V., Ivanov V., Gareev R., Serebryakov S., Vassilieva N.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

In this paper we describe methodology for building information extraction (IE) rules. Rules are usually developed by experts and are widely used in knowledge-based IE systems. They consist of two parts: the left-hand side (LHS) of a rule is a template that matches a certain syntactico-semantic structure (SSS) and the right-hand side is an action that is executed when LHS template is matched against a particular text fragment. In the paper we describe the process of building a more complex LHS part (further in the paper we will refer to LHS as template). This methodology was used for developing the information extraction system that extracts business events from news articles written in Russian language. © Copyright 2012 Hewlett-Packard Development Company, L.P.

Keywords

Free-context grammars, IE, Meaning-text, Rules, Templates