

## **Инструментальная система на базе языка шаблонов LSPL: новые средства и приложения**

Большакова Елена Игоревна, Иванов Кирилл Михайлович, Тодуа Антон Романович  
Кафедра алгоритмических языков

bolsh@cs.msu.su, ivanov.kir.m@yandex.ru, kontakt4kontakt@yandex.ru

К числу актуальных научных направлений относится разработка методов и средств автоматической обработки текстов (АОТ) на естественном языке, сфера применимости которых все более расширяется и охватывает задачи извлечения информации из специализированных текстов, реферирования и аннотирования текстов и др. Для эффективной разработки прикладных АОТ-систем применяются инструментальные средства, включающие модули анализа текста, а также формальные языки описания лингвистических свойств распознаваемых в текстах конструкций, обычно в форме специальных шаблонов и правил.

Для построения АОТ-приложений для русскоязычных текстов была создана и развивается инструментальная система [1], использующая формальный язык LSPL лексико-синтаксических шаблонов. Ядром системы является компонент поиска и выделения в тексте конструкций по их LSPL-шаблонам, с помощью шаблонов и правил этот компонент настраивается для решения конкретной прикладной задачи. На базе указанной инструментальной системы был построен ряд приложений, в том числе комплекс процедур терминологического анализа научно-технического текста. Опыт использования языка LSPL выявил его ограничения, а также слабые места программной архитектуры инструментальной системы.

В докладе рассматриваются средства языка LSPL, реализованные в последнее время и усиливающие его выразительность, а также новые программные приложения, построенные на базе рассматриваемой инструментальной системы.

Новые средства языка LSPL позволяют указывать свободный порядок элементов шаблона (за счёт введения операции перестановки), а также задавать логические условия (дизъюнкции и отрицания) на значения морфологических признаков слов-элементов, входящих в описываемую шаблоном конструкцию.

Одним из новых приложений является программное средство построения предметных указателей для русскоязычных научно-технических документов. Предметный указатель – структурированный перечень основных терминов текста с указанием страниц, на которых они употребляются. В построенном приложении используются лингвистические правила и LSPL-шаблоны грамматических образцов терминов и характерных конструкций их определений в текстах. Эти шаблоны позволяют извлекать основные термины документа, выявлять семантические связи между ними, а также определять их местоположение (страницы текста).

В докладе обсуждается также изменение программной архитектуры инструментальной системы в сторону ее гибкости, а именно: разделение монолитного ядра на несколько отдельных взаимодействующих программных компонентов, что облегчает внесение в них изменений и развитие инструментальной системы.

### **Литература**

1. Большакова Е.И., Иванов К.М., Сапин А.С., Шариков Г.Ф. Система для извлечения информации из текстов на базе лексико-синтаксических шаблонов // Пятнадцатая национальная конференция по искусственному интеллекту с междунар. участием (КИИ-2016): Труды конференции. Том. 1 – Смоленск, Универсум, 2016, с.14-22.