

МЕТОД АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ИНФОРМАЦИИ В ДИСТАНЦИОННОМ ОБУЧЕНИИ

Р.В. Огородник, Л.В. Серебряная

*Белорусский государственный университет информатики и радиоэлектроники, Минск,
Беларусь, aharodnik@gmail.com, l_silver@mail.ru*

Abstract. This report reveals the possibility of using software tools for automated processing of information in distance learning. The software tool aimed at systematization of the material and help students in mastering it, and the implementation of automatic control of knowledge.

Развитие дистанционного обучения требует всё больше средств и способов упорядочения информации с помощью электронных средств. Одним из таких средств, которое может быть использовано для дистанционного обучения, является электронный помощник, представляющий собой программное средство организации обработки текстовой информации. В его основе лежит анализатор учебно-методических материалов, который позволяет структурировать данные, объединять учебные материалы в единую базу, автоматически создавать тесты по заданным темам, в режиме диалога с обучаемым, отвечать на уточняющие вопросы по материалу, представлять рефераты текста по заданным темам и ключевым словам.

Одной из основных задач при построении подобных систем является синтаксический и морфологический разбор на уровне языка учебно-методической литературы. В качестве языка обучения выбран русский язык и рассмотрена структура его морфологическо-синтаксических правил.

Для организации синтаксического разбора нужно определить грамматическую форму слова. Поэтому задачи морфологического и синтаксического разбора взаимосвязаны. Для морфологического разбора используется алгоритм Snowball[1], который хорошо себя зарекомендовал для разбора русскоязычных текстов. В качестве дополнения к разбору слов с помощью алгоритма Snowball использовалась система определения части речи и грамматической формы, а затем и части предложения на основе флексий – изменяемых частей слова, определяющих её грамматическую форму. Методом частотного анализа определялось грамматическое значение слова и соответствие части предложения. В дополнение к частотному методу анализировалась позиция слова в предложении и соседствующие рядом слова. В процессе работы программы текст конвертируется в структуру данных, понятную для компьютера, он может собирать тематическое наполнение. Весь словарь первоначально хранится в префиксном дереве[2], что обеспечивает быстрый поиск. По найденному слову определяется наличие такой словоформы и совершается быстрый поиск по индексу в базе. При поиске по ключевым словам происходит сбор данных, связанных с сущностью и её дубликатами, преобразуя заложенные данные.

Таким образом, в ходе работы был произведен морфологический анализ текста, улучшен и связан с синтаксическим анализом предложения морфологический анализ слов. Полученные результаты будут использованы в программном средстве автоматического анализа текстовой информации.

Литература

1. [Электронный ресурс] http://www.gabormelli.com/RKB/Snowball_Algorithm_Description Дата доступа - 03.11.2013
2. [Электронный ресурс] <http://habrahabr.ru/company/abbyu/blog/190158/> Дата доступа - 03.11.2013