МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Поддубных Мария Сергеевна

«Автоматизированное выявление параметров концепта формы в лексическом описании для тезауруса RussNet»

Профиль:  «Компьютерная лингвистика и интеллектуальные технологии»

Программа: «Прикладная, экспериментальная и математическая лингвистика»

Научный руководитель: к.ф.н., доц. Азарова И. В.

Данная работа проводится в рамках проекта RussNet, компьютернего тезауруса wordnet-типа, который может использоваться как электронный лексикон для текстовых анализаторов русского языка.

Цель данного исследования: описать структуру семантического дерева «Форма» для RussNet и выяснить, можно ли осуществлять дальнейшее пополнение этого дерева в полуавтоматическом режиме.

Методы исследования включали существующую стандартную методику пополнения RussNet, предполагающую дефиниционный, контекстный и статистический анализ данных по корпусу текстов, а также методику полуавтоматического пополнения посредством программных инструментов.

Материалом для исследования послужили 16 частотных лексем и 11 низкочастотных лексем с заданным значением. 19 значений этих лексем были представлены в виде ядра иерархической структуры семантического дерева «Форма». Контекстные маркеры рассмотренных слов были использованы при задании лексико-грамматических шаблонов для программы автоматического пополнения данной лексико-семантической группы. Реализованный на языке С# основанный на правилах алгоритм показал полноту 66,67% и точность 91,25%. Программа была проверена на 275 контекстах (5015 словоупотреблений).

Актуальность и новизна исследования определяется необходимостью введения в тезаурус RussNet данных семантического дерева «Форма» и исследования возможности дальнейшего пополнения дерева в полуавтоматическом режиме.

Работа состоит из введения, трех глав, заключения, списка литературы и приложений. Список использованной литературы содержит 39 источников, из них 20 — на иностранном языке.