

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования



**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Направление подготовки/профиль 09.06.01 Информатика и вычислительная техника
Инженерная школа информационных технологий и робототехники _____
Отделение информационных технологий _____

**Научный доклад об основных результатах подготовленной
научно-квалификационной работы**

Тема научного доклада
Автоматизация построения фреймового представления смыслового содержания дефиниций русскоязычного толкового словаря интегративного типа

УДК 004.823:004.421:811.161.1'374/3

Аспирант

Группа	ФИО	Подпись	Дата
A4-36	Мишунин Олег Борисович		

Руководителя профиля подготовки

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Фадеев А.С.	К.Т.Н.		

Руководитель отделения

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Дёмин А.Ю.	К.Т.Н.		

Научный руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
	Савинов А.П.	К.Т.Н.		

Автоматизация построения фреймового представления смыслового содержания дефиниций русскоязычного толкового словаря интегративного типа

Диссертационная работа посвящена разработке машинного семантического толкового словаря, способа представления знаний в таком словаре, и системы для автоматизированного построения такого словаря на основе анализа определений из существующих толковых словарей на естественном языке. Задача понимания текстов компьютером была поставлена еще в середине 20 века, но существенно приблизиться к ее решению так и не удалось. В процессе семантического анализа текстов на естественном языке на современном этапе как правило используются лингвистические методы (морфологического и синтаксического анализа, модели управления глаголов, именных и предложно-падежных конструкций) с привлечением онтологических или тезаурусных словарей, которые позволяют реализовать семантический поиск и категоризацию документов, но не позволяют осуществлять определение их смыслового содержания и производить на его основе логический вывод. Как показал анализ, причиной создавшейся ситуации является сложность и трудоемкость построения онтологических словарей, требующая ручной работы экспертов и отсутствие описания семантических признаков понятий, без которых невозможно производить сложные операции над понятиями словаря, такие как логический вывод. Важным недостатком является и ограниченность объема онтологических словарей использованием наиболее употребительных словосочетаний и терминов определенной предметной области, в то время как в любом научно-техническом тексте наряду с профессиональными терминами применяется также и лексика, отражающая языковую картину мира, которая описана в толковых словарях. Наличие данных проблем делает актуальной задачу создания словаря нового типа, отражающего весь спектр значений, принимаемых словами русского языка, и детально описывающего их элементарные семантические признаки на основе определений данных значений в статьях толковых словарей русского языка. Объем толковых

словарей покрывает большую часть лексики русского языка, а состав определений содержит достаточные для понимания значений слов наборы их семантических признаков.

Целью работы является создание программной системы, позволяющей автоматизировать процесс построения семантических словарей полевого типа с фреймовым представлением компонентного состава дефиниций значений словарных статей, характеризуемых дифференциальными и интегральными семантическими признаками в качестве слотов построенных фреймов. Для увеличения полноты описания признаков понятий и определения полевого расположения семем и их компонентов анализ определений осуществляется из пяти толковых словарей. Чтобы достичь вышеуказанной цели, были разработаны наборы правил и алгоритмы извлечения из дефиниций признаков, соответствующих различным семантическим категориям, таким как родовидовые, партитивные, посессивные и локативные отношения, качественные и количественные признаки. Структуризация полевого расположения значений словарных статей от ядерных (основных) до периферийных (переносных) значений с дифференциальными и интегральными признаками семем осуществляется статистическим образом по частоте встречаемости их в используемых толковых русскоязычных словарях. Был разработан модуль предварительной обработки определений, приводящий их к пригодному для анализа виду. Для построения пробного варианта словаря был взят срез лексики, соответствующий лексическому минимуму по русскому языку как иностранному.

Научная новизна диссертационного исследования определяется следующими положениями:

1. Предложен и реализован новый тип организации машинного семантического словаря полевого типа, сочетающий иерархию связанных понятий с внутренним сетевым фреймовым представлением значений понятий, включающего их семантические признаки, и построен фрагмент такого словаря.

2. Предложены и реализованы алгоритмы формализации и структуризации определений значений словарных статей толковых словарей.

3. Разработана теория и алгоритм автоматизации ручного компонентного анализа, применяемого лингвистами при исследованиях смыслового содержания единиц языка.

4. Разработаны и реализованы алгоритмы представления смыслового содержания понятий в виде сети фреймов, содержащих наборы семантических дифференциальных и интегральных признаков, полученных из определений толковых словарей.

5. Предложены и реализованы алгоритмы автоматизации структуризации полевой структуры значений понятий.

6. Разработан и сформирован программный комплекс, позволяющий автоматизировать процесс создания семантического словаря полевого типа с фреймовым представлением структуры семантических признаков, характеризующих каждое значение словарной статьи.