Алексеевский, Даниил Андреевич Методы автоматического выделения тезаурусных отношений на основе словарных толкований

ОГЛАВЛЕНИЕ ДИССЕРТАЦИИ

кандидат наук Алексеевский, Даниил Андреевич

Оглавление

Стр.

Введение

Глава 1. Построение электронных тезаурусов: мотивация, методы,

источники

1.1 Тезаурусы и тезаурусные отношения

1.1.1 Понятие тезауруса. Основные типы тезаурусов

1.1.2 Princeton WordNet

1.1.3 Семантические сети и электронные тезаурусы

1.1.4 Семантические сети для языков мира

1.1.5 Русскоязычные электронные тезаурусы

1.1.6 Применение семантических сетей

1.1.7 Постановка задачи

1.2 Методы построения электронных тезаурусов

1.2.1 Методы, основанные на ручном труде

1.2.2 Методы, основанные на массовых опросах

1.2.3 Методы, использующие существующие структурированные ресурсы

1.2.4 Методы, использующие переводные структурированные ресурсы

1.2.5 Методы, использующие толковые словари

1.2.6 Методы, использующие текстовые корпуса

1.2.7 Методы, использующие дистрибутивные семантические модели

1.2.8 Сравнение методов

1.3 Постановка задачи: уточнение параметров

Глава 2. Толковые словари как источник полуструктурированных

данных

2.1 Типы словарей

2.1.1 Классификация толковых словарей

2.2 Организация словарной статьи

2

2.3 Типы толкований

2.4 Современные толковые словари русского языка

2.5 Машиночитаемые словари

2.6 Принципы выбора толкового словаря для исследования

Глава 3. Выделение отношений

3.1 Введение

3.2 Исходные данные

3.2.1 Большой толковый словарь русского языка

3.2.2 Электронная версия словаря

3.2.3 Восстановление структуры словарной статьи

3.2.4 Выделение толкований из словарной статьи

3.2.5 Определение частей речи

3.2.6 Результаты

3.3 Контрольный эксперимент: носитель семантического

отношения — вершина именной группы из толкования

3.3.1 Исходные допущения

3.3.2 Описание эксперимента

3.3.3 Результаты и анализ

3.4 Извлечение отношений с частичным привлечением учителя. Автоматическая кластеризация толкований

3.4.1 Машинное обучение с частичным привлечением учителя

3.4.2 Общее описание эксперимента

3.4.3 Признаки, используемые для кластеризации

3.4.4 Лексико-грамматические ^граммы

3.4.5 Материалы и методы

3.4.6 Кластеризация толкований. Эффект применения лексико-грамматических триграмм

3.4.7 Подбор параметров кластеризации

3.4.8 Правила, применяемые к кластеру толкований

3.4.9 Результаты

3.5 Обсуждение

3.6 Выводы

Глава 4. Дизамбигуация тезаурусных отношений

4.1 Задача дизамбигуации в контексте извлечения гипо-гиперонимических отношений из словаря

4.2 Подходы к задаче разрешения семантической неоднозначности

4.2.1 Алгоритм Леска и его усовершенствования

4.2.2 Методы машинного обучения в задачах разрешения неоднозначности

4.2.3 Векторные семантические модели

4.2.4 Выбор методов

4.3 Материалы исследования

4.3.1 Корпус лексических значений: БТС

4.3.2 Корпус омонимичных отношений

4.3.3 Материал для построения векторных моделей

4.3.4 Аннотация данных

4.4 Постановка эксперимента

4.5 Результаты

4.6 Анализ

4.7 Заключение

Заключение

Список литературы

Список рисунков

Список таблиц