

Исследование зависимости качества распознавания онтологических объектов от глубины гипонимии.*

Кузнецов М. Д., Дочкина В., Резяпкин В., Русскин А., Ярмошик Д.

kuznetsov.md@phystech.edu

Данная работа посвящена изучению онтологических объектов и их гипонимий. Проведен сбор датасета, состоящий из текстов на английском языке как с уже сделанной NER разметкой, так и абсолютно её не имеющих, анализ и разметка данных текстов по нескольким уровням вложенности. Также (будет) проведена серия экспериментов для определения зависимости качества распознавания объектов от уровня вложенности. Обученные модели, полученные в результате данных экспериментов, могут использоваться в получении дополнительных признаков для решения различных задач NLP. Исследованная методика разметки онтологических объектов в свою очередь может быть использована для автоматической разметки других текстов.

Ключевые слова: гипонимия, онтология, онтологические объекты.

Введение

Именованное распознавание объектов (NER - Named-entity recognition) является важной задачей NLP (Neuro-Linguistic Programming — нейролингвистическое программирование), которая заключается в автоматическом обнаружении объектов в тексте и классификации их в заранее определенные типы сущностей, такие как люди, организации, геополитические субъекты, места, события и т. д. NER является фундаментальным компонентом многих методов извлечения информации, включая извлечение связей, связывание объектов, “ответов на вопросы” и др. Один из аспектов “связывания” объектов есть отношения обобщений, или гипонимии. Каждая сущность может быть обобщена до более высокого уровня вложенности понятий. Данные связи можно активно использовать для повышения качества выполнения задач NLP, а также разработки новых возможностей для уже существующих моделей. Для реализации исследования использованы данные из WordNet.

Литература

- [1] Воронцов К. В. \LaTeX в примерах. 2006. <http://www.ccas.ru/voron/latex.html>.
- [2] Львовский С. М. Набор и верстка в пакете \LaTeX . 3-е издание. Москва: МЦНМО, 2003. 448 с.
- [3] Jingbo Shang, Liyuan Liu, Xiang Ren, Xiaotao Gu, Teng Ren, Jiawei Han Learning Named Entity Tagger using Domain-Specific Dictionary
- [4] Emily Sheng, Prem Natarajan. A Byte-sized Approach to Named Entity Recognition
- [5] Bill Yuchen Lin, Wei Lu Neural Adaptation Layers for Cross-domain Named Entity Recognition
- [6] Dilek Küçük Joint Named Entity Recognition and Stance Detection in Tweets
- [7] Jian Ni, Radu Florian Improving Multilingual Named Entity Recognition with Wikipedia Entity Type Mapping