

УДК 004.42

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ПОИСКА СЕМАНТИЧЕСКИ ПОДОБНЫХ ЗАПРОСОВ ДЛЯ ПОИСКОВЫХ СИСТЕМ

Л.А. КОЗЛОВА^{1*}, О.Н. МАЛЫХ²

¹ магістрант кафедри САИТ, НТУ «ХПИ», Харків, УКРАЇНА

² доцент кафедри САИТ, канд. техн. наук, НТУ «ХПИ», Харків, УКРАЇНА

* email: kozlova.lesya0610@gmail.com

Многократное увеличение информационных потоков, циркулирующих между пользователями, динамичное развитие информационных ресурсов приводит к тому, что пользователь уже не в состоянии самостоятельно, без технической поддержки, находить требуемую информацию.

В настоящее время в области искусственного интеллекта разработан ряд средств представления знаний. К наиболее эффективным из них относится онтология – система понятий (концептов, сущностей, классов), отношений между ними и правил операций над ними в определенной предметной области.

Целью данной работы является реализация и сравнение алгоритмов и моделей сравнения предложений и вопросов по смыслу. Для достижения поставленной цели были рассмотрены и реализованы такие методы, как WordNet и Word2Vec. Для последующей работы с предложениями с использованием вышеуказанных методов необходимо провести их терминологический анализ, который состоит из следующих основных этапов: определение границ слов предложения, выполнение лингвистической обработки, определение частей речи слов и определение «стоп-слов».

Далее с помощью WordNet и Word2Vec методов вопросы были проанализированы на предмет смыслового подобия. WordNet является лексической базой данных, в которой все слова взаимосвязаны с помощью концептуально-семантических и лексических связей. Word2Vec – программный инструмент анализа семантики естественных языков, который представляет собой технологию, основанную на дистрибутивной семантике и векторном представлении слов. Таким образом, была выполнена проверка методов на парах вопросов в количестве 404290. В результате удалось достичь 63.8% успешных прогнозов определения сходства двух вопросов для метода WordNet, и 66.4% успешных прогнозов для метода Word2Vec.

Список литературы:

1. Михеев А.С. Когнитивная система экстрагирования концептуальных знаний из научно-технических текстов: автореф. / науч. рук. В.В. Нечаев. Москва: МИРЭА, 1990. 23 с.
2. J.J. Jiang and D.W. Conrath Semantic similarity based on corpus statistics and lexical taxonomy. In Proc. of the Int'l. Conf. on Research in Computational Linguistics, 1997. pages 19–33 available at: <https://www.cse.iitb.ac.in/~cs626-449/Papers/WordSimilarity/4.pdf> (accessed 13.02.2017)