

## ПРИЛОЖЕНИЕ 1. Пример оформления титульного листа статьи

УДК 004.855

СЕМАНТИЧЕСКИЙ ПОИСК УЧЕБНЫХ ДИСЦИПЛИН ПОД ТРЕБОВАНИЯ  
РЫНКА ТРУДА НА ОСНОВЕ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ЯЗЫКА  
NEURAL NETWORK-BASED SEMANTIC SEARCH OF EDUCATIONAL  
PROGRAMMES FITTING LABOR MARKET REQUIREMENTS

\*Иванов Петр Сергеевич, Челябинский ... университет, старший преподаватель  
кафедры ... , Ivanov\*\*\*@gmail.com, +7 900 0000000

Ivanov Petr Sergeevich, Chelyabinsk ... University, senior lecturer of the Department  
of ... , Ivanov\*\*\*@gmail.com, +7 900 0000000

Петров Сергей Иванович, Челябинский ... университет, аспирант кафедры ...,  
pe \*\*\*@mail.ru

Petrov Sergey Ivanovich, Chelyabinsk ... University, postgraduate student of the  
Department of ..., pe \*\*\*@mail.ru

**Введение.** В условиях роста объема открытого образовательного контента, повышения требований к профессиональному образованию со стороны рынка труда, развития концепции обучения в течение всей жизни сегодня крайне актуальной является задача регулярного обновления содержания образовательных программ и отдельных дисциплин. **Цель исследования.** Рассмотреть применение метода семантического поиска образовательного контента под заданные требования рынка труда, определяемые профессиональными стандартами. **Материалы и методы.** Традиционные подходы сопоставления и анализа содержания образовательных программ основаны на онтологических моделях и правилах. Предлагается использовать подход к распределенному представлению слов с помощью известных

нейросетевых моделей языка word2vec и fastText. В качестве исходных запросов выступают фрагменты профессиональных стандартов – конкретные требования к знаниям, умениям и описания трудовых действий и трудовых функций, а в качестве искомых документов – описания учебных дисциплин и онлайн-курсов, включающие аннотацию, результаты обучения, структуру и содержание основных тем. Для перехода от векторов слов к векторам текстов исследуются разные способы усреднения векторов слов, полученных от нейросетевых моделей, в сочетании с векторной моделью TF-IDF. **Результаты.** Приводятся данные экспертной оценки качества ранжирования по метрике NDCG (Normalized Discounted Cumulative Gain) и точности семантического поиска по метрике MAP (Mean Average Precision) на представительном корпусе программ учебных дисциплин вузов по ИТ-направлениям и массовых открытых онлайн-курсов. **Заключение.** Лучшие результаты для поиска показывают модели word2vec и fastText, обучаемые без учителя на больших специально подготовленных корпусах текстов программ учебных дисциплин и описаний онлайн-курсов.

**Introduction.** With the growth of open educational content, growing demand for professional education from the labor market, and the development of the concept of lifelong learning, the task of updating the content of educational programs today is extremely important. **Aim.** The article discusses application of the semantic search method to retrieval and ranking of educational content for the specified requirements of the labor market, determined by professional standards. **Materials and methods.** Traditional approaches of matching and analyzing the content of educational programs are based on ontological models and rules. We propose the usage of word embedding and well-known neural network language models word2vec and fastText. The initial requests are specific requirements for knowledge, skills and descriptions of labor activities and labor functions extracted from professional standards. The search results are the descriptions of academic disciplines and online courses, including goals and objectives, learning outcomes, the structure and content of the main topics. To

move from word vectors to document vectors various combinations of neural network models with the TF-IDF weighting scheme are investigated. **Results.** We include the results of the expert evaluation of the ranking quality for the semantic search by metrics NDCG (Normalized Discounted Cumulative Gain) and MAP (Mean Average Precision) on the representative corpus of IT disciplines programmes of universities and massive open online courses (MOOC). **Conclusion.** The best results for the search are shown by the word2vec and fastText models, which are trained without supervision on large specially prepared corpuses of curriculum programs and descriptions of online courses.

Ключевые слова: семантический поиск, семантическая близость, дистрибутивная семантика, word2vec, fastText, учебная дисциплина, массовые открытые онлайн-курсы, рынок труда

Keywords: semantic search, semantic similarity, distributional semantic, word2vec, fastText, academic discipline, massive open online courses, labor market

*На второй странице и далее:*

**Введение**

**Основные разделы**

**Выводы (Заключение)**

**Литература**