

ИНТЕЛЛЕКТУАЛЬНАЯ ПОДДЕРЖКА ФОРМИРОВАНИЯ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ НА ОСНОВЕ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ЯЗЫКА С УЧЕТОМ ТРЕБОВАНИЙ РЫНКА ТРУДА

Д.С. Ботов

Челябинский государственный университет, г. Челябинск, Россия

Активное развитие отраслей цифровой экономики сегодня предъявляет высокие требования к адаптивности, практической направленности и качеству современных образовательных программ. Существующие подходы к интеллектуальной поддержке формирования образовательных программ на основе онтологических моделей, экспертных систем и эвристических алгоритмов не позволяют эффективно учитывать и отслеживать изменения как на рынке труда, так и в пространстве открытого образовательного контента в среде Интернет. Вместо этого предлагается использовать подходы к семантическому анализу текстов на основе известной нейросетевой модели языка word2vec, обучаемой без учителя на больших текстовых корпусах. Сложность сопоставительного семантического анализа заключается в переходе от определения меры семантической близости для отдельных коротких описаний анализируемых сущностей (тем курсов, результатов обучения, требований вакансий и т. д.) к сопоставлению больших структурируемых документов, таких как профессиональный стандарт, образовательная программа по направлению подготовки. Для учета взаимосвязей сущностей вводится графовая модель представления образовательной и профессиональной области. В работе предлагается интеллектуальный метод формирования рекомендаций по актуализации целей и содержания образовательных программ, включающий четыре этапа анализа. На первом этапе определяются актуальные требования рынка труда на основе семантического сопоставления фрагментов вакансий с содержанием профессиональных стандартов. Второй этап включает в себя семантическое сопоставление содержания учебных дисциплин с требованиями рынка труда. На третьем этапе производится семантический поиск релевантного образовательного контента среди программ дисциплин ведущих вузов и онлайн-курсов. В ходе четвертого этапа формируются итоговые рекомендации по актуализации образовательной программы. В рамках эксперимента продемонстрирована возможность применения метода для сопоставления результатов обучения и содержания дисциплин с требованиями профессиональных стандартов с оценкой качества на примере образовательной программы по направлению «Информатика и вычислительная техника».

Ключевые слова: интеллектуальный анализ данных; обработка естественного языка; нейросетевые модели языка; семантическая близость; word2vec; образовательная программа; профессиональный стандарт, рынок труда.

Введение

Активное развитие образовательных программ ведущих университетов в России и за рубежом, использование дистанционных образовательных технологий в учебном процессе, появление большого числа открытых онлайн-курсов и качественного образовательного контента в сети Интернет приводит к резкому росту конкуренции в образовательной среде, когда образовательные организации, обучающиеся и выпускники уже не привязаны к своему географическому месторасположению.

Постоянная актуализация целей, результатов обучения и содержания образовательных программ становится не просто требованием нормативных актов и образовательных стандартов, а необходимым условием обеспечения качественной подготовки выпускников, способных конкурировать на современном рынке труда.

При этом неуклонно растет скорость изменения требований работодателей, особенно в активно развивающихся отраслях цифровой экономики. Но для обеспечения качества образования становится явно недостаточным формальное соответствие компетентностным требованиям ФГОС и квалификационным требованиям профессиональных стандартов. Так, в концепции кадрового обеспечения промышленного роста в рамках программы Национальной технологической инициативы (НТИ) [1] отмечены в том числе следующие ключевые проблемы: оторванность образования от работодателя, устаревание ФГОС из-за отсутствия механизмов их обновления и соблюдения баланса «теория – практика», длительный цикл от начала разработки до внедрения профессиональных стандартов.

В связи с этим появляются концепции различных экспертных систем. Примерами концепций систем могут выступать: автоматизированная система по мониторингу и анализу соответствия кадровых потребностей на основе номенклатуры вузов [2], квалификационно-ориентированная экспертная система управления образовательным процессом вуза [3].

Данное исследование посвящено решению комплекса задач по интеллектуальной поддержке формирования образовательных программ от анализа существующих потребностей рынка труда через определение актуальных целей и результатов обучения к формированию рекомендаций по изменению содержания образовательной программы на основе интеллектуального анализа образовательных программ и отдельных дисциплин ведущих образовательных организаций и массовых открытых онлайн-курсов (МООС).

1. Обзор существующих методов и моделей интеллектуальной поддержки формирования образовательных программ

Сегодня происходит активное развитие подходов, методов и алгоритмов интеллектуальной поддержки в области управления образовательным процессом и сопоставления образовательных программ.

Концептуальные онтологические модели, дискретно-событийные модели и методы инженерии знаний используются в работах О.Н. Сметаниной, Н.И. Юсуповой [4] для создания системы поддержки принятия решений при управлении образовательным маршрутом в условиях академической мобильности. Онтологический подход применяется в работе [5] для семантического моделирования учебного плана в программах высшего образования.

В работе Е.А. Черниковой [6] предлагается для сопоставления образовательных программ использовать онтологический подход, а также анализ результатов обучения с помощью таксономии образовательных целей.

Адаптивный поиск образовательных ресурсов с помощью рассуждений по прецедентам и онтологическим моделям описан в работе А.Ю. Ужвы [7].

Стоит отметить, что большая часть работ по формированию образовательных программ и управления образовательными траекториями методологически опирается на компетентностный подход с использованием методов экспертных оценок и систем, основанных на правилах (rule-based). Примерами таких работ могут выступать: работа И.М. Харитоновой [8] по созданию формализованной модели составления учебного плана с помощью методов контент-анализа и латентно-семантического анализа (LSA), применение экспертных оценок и когнитивных карт для анализа компетенций образовательной программы в работе И.В. Сибикиной [9] в сочетании с лингвистическим классификатором на основе правил и нечетких множеств. Однако в данных работах не производится семантический анализ результатов обучения по отдельным дисциплинам, и это не позволяет дать детальные рекомендации по актуализации целей и содержания образовательных курсов.

Свое применение находят графовые модели и структурные подходы к формированию учебных планов – примерами могут выступить: работа А.С. Пирской, Л.С. Лисицыной [10] по созданию модели управления образовательными траекториями на основе план-графов, работа С.С. Котова [11] по автоматизированному составлению учебного плана на основе экспертных оценок с помощью эвристических алгоритмов и графовых моделей, описывающих структурно-логическую связь дисциплин.

Использование эвристик, систем правил, онтологических моделей и экспертных оценок не позволяет проводить эффективный анализ всего доступного пространства образовательного кон-

тента в сети Интернет, а также крайне трудоемко решается задача регулярного пересмотра образовательных программ и адаптивного реагирования на изменение потребностей рынка труда на уровне модификации отдельных результатов обучения, тем и разделов дисциплин.

В данном исследовании предлагается для формирования рекомендаций по актуализации образовательной программы сочетать графовые модели представления структуры документов образовательных программ, описаний онлайн-курсов, требований профессиональных стандартов и вакансий с семантическим анализом текстов на основе нейросетевой модели языка word2vec [12], обученной без учителя на больших корпусах текстов образовательных программ, отдельных дисциплин, вакансий и профстандартов. Word2vec показывает лучшее качество в решении задач семантической близости текстов, определения смысла слова в тексте, в том числе и для русского языка [13, 14].

2. Математическая модель формирования образовательной программы с учетом требований рынка труда

2.1. Обобщенная графовая модель образовательной области

Представим образовательную область как ориентированный граф $G_{edu} = (V_{edu}, E_{edu})$, где V_{edu} – множество вершин, описывающих сущности образовательной области, а E_{edu} – множество ребер (дуг), которые задают ассиметричные отношения между парами вершин графа.

Множество вершин $V_{edu} = \{C, LO\}$, включает в себя два подмножества:

- В подмножество вершин C (content) входят структурные элементы содержания образовательной программы разной степени детализации (например: структура ОП, модуль ОП, дисциплина (курс), раздел дисциплины, тема лекции / содержание практики, дидактические единицы и т. д.).

- Подмножество вершин LO (learning outcomes) включает в себя результаты обучения по образовательной программе разной степени детализации (от компетенции ФГОС до результатов обучения по отдельным дисциплинам / модулям / практикам).

Множество ребер $E_{edu} = \{Inc, Req, M\}$, включает в себя три типа отношений между вершинами графа:

- Inc (include / включает) представляет собой ассиметричное отношение $Inc \subset C \times C$ между структурными элементами ОП и описывает связь типа «Включает в себя» («часть – целое») (например: дисциплина A включает в себя раздел B , тема X является частью раздела Y)

- Req (require / result) представляет собой ассиметричное отношение $Req \subset C \times LO$ между структурными элементами ОП и результатами обучения и описывает связь типа «Требует освоения / Обеспечивает результат (выполнение)» (например: в результате освоения дисциплины A обеспечивается компетенция ПК-1, для достижения результата обучения X необходимо освоить разделы A и B).

Замечание 1. Матрица компетенций ОП является частным случаем отношения Req для дисциплин (модулей ОП) и компетенций ФГОС.

Замечание 2. Логическая последовательность изучения дисциплин в рамках ОП может быть также описана отношением Req , в этом случае множество результатов обучения $\{LO\}$ базовой дисциплины A выступают в роли входных требований (prerequisites) для следующей дисциплины B ($B \rightarrow LO \rightarrow A$).

- M (match) представляет собой ассиметричное отношение $M \subset C \times C \cup LO \times LO$, которое может задавать связи типа «Соответствует / Соотносится / Сопоставлена» между вершинами одного из подмножеств C и LO : между структурными элементами образовательной программы из подмножества C (например, онлайн-курс X соотносится с дисциплиной Y) либо между результатами обучения из подмножества LO (например, результат обучения LOa по дисциплине A соответствует результату освоения ОП (компетенции ФГОС)).

На рис. 1 представлены основные сущности и отношения графа образовательной программы, извлекаемые из рабочих программ дисциплин, учебного плана и описаний онлайн-курсов.

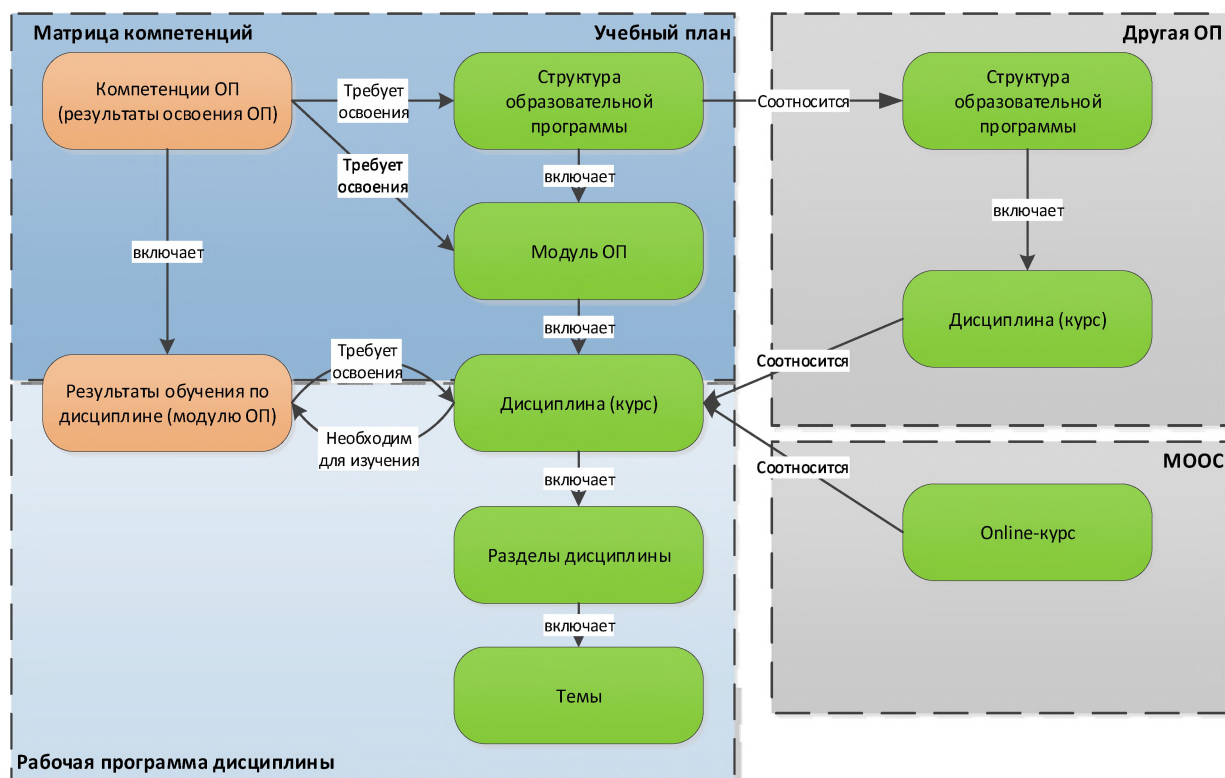


Рис. 1. Основные сущности и отношения в обобщенной графовой модели образовательной области

2.2. Обобщенная графовая модель профессиональной области

По аналогии с представлением образовательной области можно также представить профессиональную область как ориентированный граф $G_{prof} = (V_{prof}, E_{prof})$, где V_{prof} – множество вершин, описывающих сущности профессиональной области, а E_{prof} – множество ребер (дуг), которые задают ассиметричные отношения между сущностями.

Множество вершин $V_{prof} = \{F, S\}$, включает в себя два подмножества:

- Подмножество вершин F (function) включает в себя трудовые функции (функциональные обязанности) из профессиональных стандартов (ПС) или из описания вакансий (например: обобщенная трудовая функция ПС, трудовая функция ПС, трудовое действие ПС, обязанности из описания вакансии).
- Подмножество вершин S (skill) включает в себя требования к знаниям и умениям, предъявляемые ПС к трудовой функции либо требуемые работодателем в описании вакансии.

Множество ребер $E_{prof} = \{Inc, Req, M\}$, включает в себя три подмножества, описывающих возможные типы отношений между вершинами графа:

- Подмножество Inc (include/включает) представляет собой ассиметричное отношение $Inc \subset F \times F$ между трудовыми функциями (например: трудовое действие A включено в трудовую функцию B профстандарта, обязанность X включена в вакансию Y).

- Подмножество Req (require) представляет собой ассиметричное отношение $Req \subset F \times S$ между требуемыми знаниями / навыками / опытом и функциональными обязанностями «Требуется / Обеспечивает выполнение функции».

- Подмножество M (match) представляет собой ассиметричное отношение $M \subset F \times F \cup S \times S$, которое может задавать связи типа «Соответствует / Соотносится / Сопоставлена» между вершинами одного подмножества: между функциональными обязанностями из описания вакансий и функциями ПС из подмножества F либо между требованиями к вакансии и требованиями ПС из подмножества S .

На рис. 2 описаны основные сущности и отношения, извлекаемые из текстов профессиональных стандартов и вакансий из систем онлайн-рекрутмента.

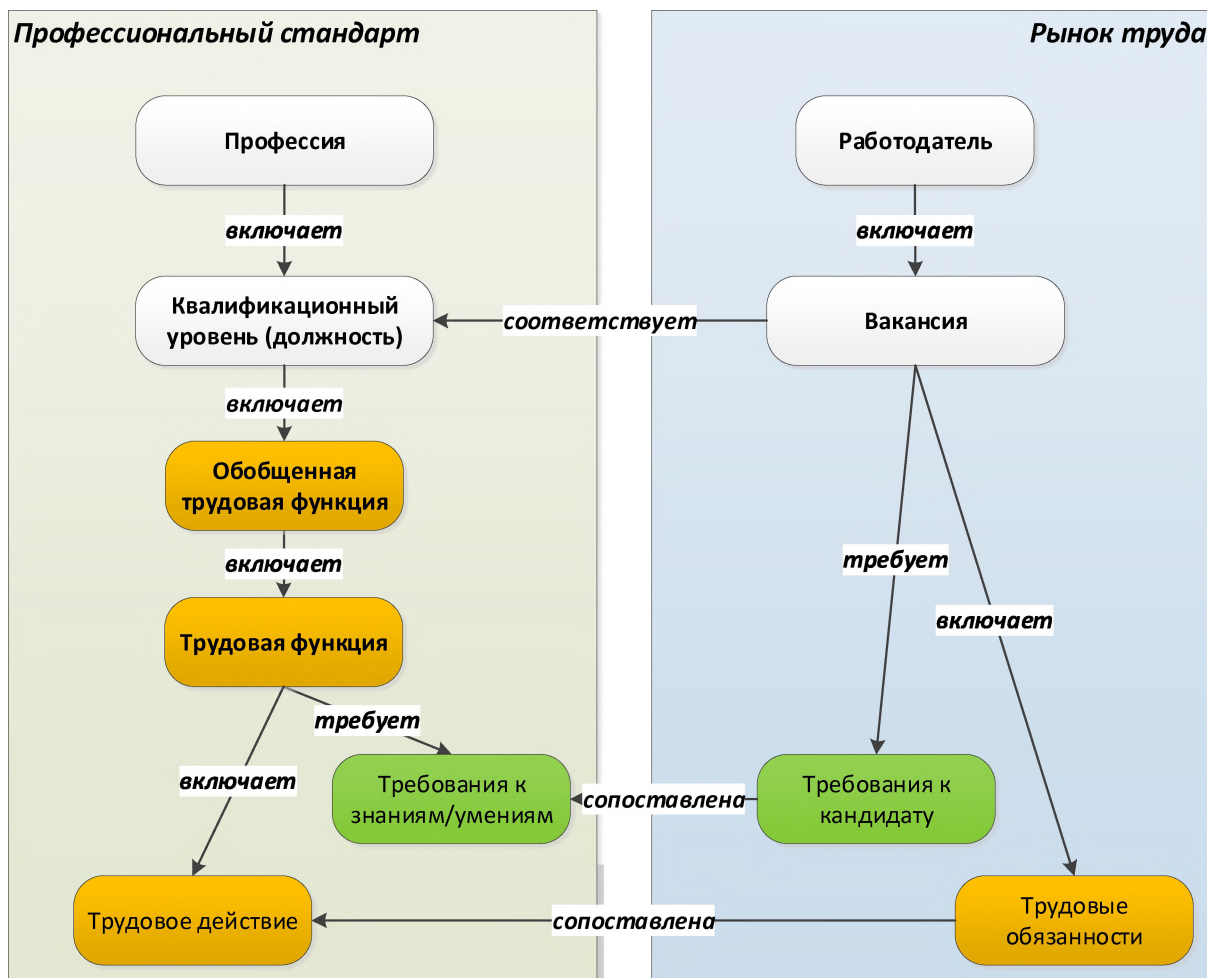


Рис. 2. Основные сущности и отношения в обобщенной графовой модели профессиональной области

3. Формирование рекомендаций на основе семантического сопоставления графовых моделей образовательных программ и требований профессиональной области

Идея метода формирования рекомендаций по актуализации образовательных программ исходит из семантического сопоставления отдельных сущностей графа образовательной программы G_{edu} и графа профессиональной области G_{prof} .

В предлагаемом методе интеллектуальной поддержки формирования образовательной программы можно выделить четыре основных этапа сопоставления графовых моделей:

Этап 1. Определение актуальных требований рынка труда.

Этап 2. Определение соответствия элементов образовательной программы (ОП) требованиям рынка труда.

Этап 3. Поиск релевантного образовательного контента: открытых онлайн-курсов (МООС) и программ учебных дисциплин ведущих вузов под требования рынка труда.

Этап 4. Актуализация ОП на основе обобщения релевантного образовательного контента.

Потоки данных в задачах семантического сопоставления графовых моделей на каждом из этапов представлены на рис. 3.

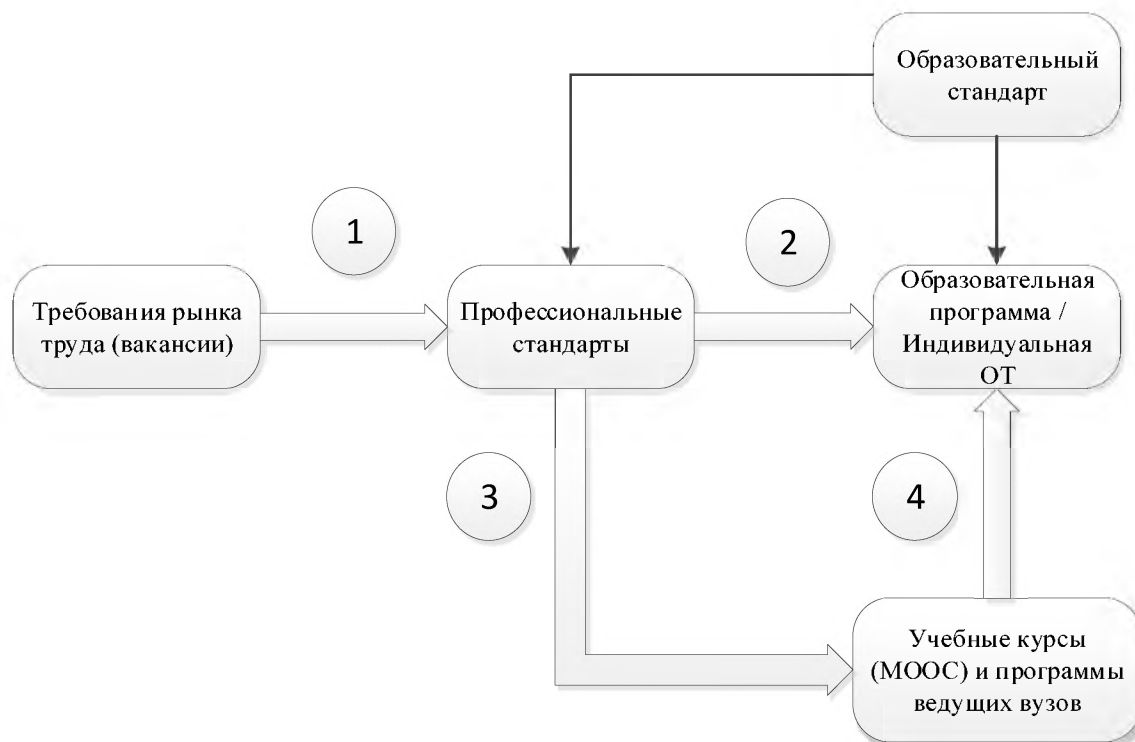


Рис. 3. Потoki данных при семантическом сопоставлении графовых моделей образовательной и профессиональной областей для формирования рекомендаций по актуализации ОП

Для формирования набора рекомендаций в рамках каждого из этапов выполняется схожая последовательность шагов:

Шаг 1. Определение (сужение) пространства поиска для решения проблемы лексической многозначности при сопоставлении отдельных сущностей. Определить, какие подграфы (подмножества сущностей) стоит использовать в сопоставлении, можно на основе предварительной классификации или кластеризации текстов документов, а также используя явно заданные связи между профессиональными и образовательными стандартами.

Шаг 2. Семантическое сопоставление сущностей в графе с расширением контекста (к описанию дочерней сущности («часть») добавляется описание связанной с ней отношением *include* родительской сущности («целое»)), вычисляется значение семантической близости между сущностями с учетом контекста, для близких сущностей (отбирается top-N ближайших) формируются новые отношения сопоставления *match* в графе.

Значение семантической близости (sim) для векторов A и B из текстовых описаний сопоставляемых сущностей (вектора описания генерируются с помощью нейросетевой модели word2vec путем усреднения векторов слов, входящих в описание сущности (average word2vec) [15]) вычисляется как косинусная близость между векторами описания сущностей:

$$sim = \frac{A \cdot B}{AB}$$

Шаг 3. На основе веса связанных сущностей определяются веса сопоставленных сущностей путем агрегации полученных значений семантической близости и дополнительных оценок, которые могут быть заданы экспертами. Возможны варианты агрегации (agg): сумма, среднее значение или максимум семантической близости сопоставленных (*match*) вершин.

Шаг 4. На основе ранжирования весов сопоставленных сущностей производится формирование рекомендаций по модификации исходных документов (профстандарта, образовательной программы, программы учебной дисциплины).

Детализация шагов и сущностей на каждом из этапов сопоставления описана в табл. 1.

Таблица 1

Детализация шагов к формированию рекомендаций в рамках этапов сопоставления графовых моделей образовательной и профессиональной областей

Шаги	Этап 1	Этап 2	Этап 3	Этап 4
Шаг 1. Определение пространства поиска	Вакансии ↓ Профессии	Профессии (ур. квалификации) ↓ Образовательная программа	Кластера дисциплин ↓ Профессии	Релевантные учебные курсы ↓ Кластера дисциплин
Шаг 2. Сопоставление сущностей графов с учетом контекста, формирование связей <i>match</i>	Требования вакансий ↓ <i>sim()</i> Трудовые функции ПС + Действия / Требования к умениям / знаниям	Требования рынка труда ↓ <i>sim()</i> Результаты обучения / Раздел + Тема дисциплины	Содержание и цели курса ↓ <i>sim()</i> Требования рынка труда	Элементы учебных курсов ↓ <i>sim()</i> Элементы дисциплин ОП
Шаг 3. Определение веса (важности) сущностей графа по агрегации связей	<i>w</i> (Трудовая функция) ↑ <i>agg(sim)</i> <i>w</i> (Действия / умения / знания) ↑ <i>agg(sim)</i> Требования / обязанности вакансий	<i>w</i> (Дисциплины) ↑ <i>agg(sim)</i> <i>w</i> (Результатов обучения) <i>w</i> (Раздел / Темы) ↑ <i>agg(sim)</i> <i>w</i> (Требования рынка труда)	<i>w</i> (Курса) ↑ <i>agg(sim)</i> <i>w</i> (Элемент курса) ↑ <i>agg(sim)</i> <i>w</i> (Требования рынка труда)	<i>w</i> (Дисциплины ОП) ↑ <i>agg(sim)</i> <i>w</i> (Элемента дисциплины) ↑ <i>agg(sim)</i> <i>w</i> (Элемента учебного курса)
Шаг 4. Формирование рекомендаций	Ранжирование по важности требований и функций ПС Формирование требований рынка труда путем расширения требований ПС ключевыми словами из требований вакансий	Удаление элементов ОП, не соответствующих требованиям рынка труда Дополнение результатов обучения актуальными требованиями рынка труда	Ранжирование курсов по релевантности требованиям рынка труда	Удаление нерелевантных элементов ОП Добавление новых дисциплин и отдельных элементов курсов (разделов / тем)

Примеры возможных вариантов модификации ОП на основе сформированных рекомендаций:

- Добавить актуальный результат обучения, отражающий требования рынка труда, и связать его с компетенцией ФГОС.
- Добавить новый образовательный контент (дисциплину, раздел, тему) для обеспечения результата обучения.
- Определить к образовательному контенту входные требования (*prerequisites*) и базовый образовательный контент, который эти требования должен обеспечить (рассматриваем входные требования как результаты обучения другого курса, базового по отношению к заданному).
- Удалить неактуальный результат обучения.
- Удалить или сократить неактуальный образовательный контент (дисциплину, раздел, тему).
- Расширить список рекомендуемых источников (учебных материалов, онлайн-курсов) для дисциплины.

4. Извлечение информации из текстов образовательных программ, онлайн-курсов, вакансий и профессиональных стандартов

Для построения графовых моделей образовательной и профессиональной областей предлагается реализовать алгоритмы извлечения информации из структурированных текстов образовательных программ, онлайн-курсов, вакансий и профессиональных стандартов.

При анализе документов извлекаются основные сущности, описанные в графовых моделях (см. п. 2), и определяются отношения типа *include* (включение) и *require* (требования). На рис. 4 представлен алгоритм извлечения сущностей и отношений между ними.



Рис. 4. Алгоритм извлечения информации из текстовых документов

С учетом большой вариативности структуры и содержания текстов рабочих программ дисциплин, образовательных программ и текстов вакансий к правилам извлечения сущностей и отношений предъявляются требования по независимости от входного формата документа (для этого вводится шаг преобразования формата), а также требования по независимости от различий структуры документов в разных образовательных организациях и MOOC-платформах, а также от структуры описания вакансий различными компаниями-работодателями (для этого вводится шаг анализа структуры документа).

Токенизация и обработка текста представляет собой традиционный шаг для задач обработки естественного языка. Правила анализа структуры документа и извлечения сущностей и отношений из предобработанного текста реализованы с помощью открытого инструмента Yargy (yargy.readthedocs.io), позволяющего описывать правила с использованием контекстно-свободных грамматик и словарей ключевых слов.

Для построения графовых моделей и обучения нейросетевых моделей языка в рамках предметной области подготовки ИТ-специалистов были собраны текстовые корпуса, содержащие следующие данные:

- 40 профессиональных стандартов из реестра «06. Связь, информационные и коммуникационные технологии»;
- 25 ФГОС ВО (3+) по направлениям подготовки бакалавриата и магистратуры, связанным с ИТ;
- 600 тыс. вакансий с онлайн-ресурсов;
- 2000 русскоязычных курсов из 4 MOOC-платформ;
- 1000 программ учебных дисциплин с сайтов 10 университетов.

5. Эксперимент по семантическому сопоставлению графовых моделей образовательной программы и профессионального стандарта

Проведенный в рамках исследования эксперимент иллюстрирует этап 2 по определению соответствия элементов образовательной программы (ОП) требованиям профессионального стандарта.

В рамках эксперимента была построена графовая модель образовательной программы на основе анализа 20 рабочих программ дисциплин из образовательной программы одного из вузов по направлению подготовки 09.03.01 «Информатика и вычислительная техника». В анализ отбирались программы профессиональных дисциплин, читаемых на старших курсах. Также были построены графовые модели профессиональных стандартов по профессиям «Программист», «Специалист по информационным системам», «Системный администратор информационно-коммуникационных систем». Данные профстандарты были выбраны исходя из требований нового ФГОС ВО (3++) по направлению «Информатика и вычислительная техника».

С помощью предобученной модели word2vec было выполнено семантическое сопоставление сущностей профессиональных стандартов: трудовых действий, требований к знаниям и требований к умениям со следующими сущностями учебных дисциплин из графа образовательной программы с учетом расширения контекста (+): Тема + Раздел дисциплины, Результат обучения (знать / уметь / владеть) + Компетенция.

Экспертами была размечена релевантность сопоставленных с каждым требованием профстандарта по 10 ближайших по значению косинусной семантической близости сущностей образовательной программы. Дана оценка качества сопоставления сущностей образовательной программы с требованиями профстандарта по традиционной для информационного поиска метрике Mean Average Precision (MAP). Результаты представлены в табл. 2.

Таблица 2

Оценка качества семантического сопоставления отдельных сущностей ОП с требованиями профстандарта

Сопоставляемые сущности	MAP@1	MAP@3	MAP@5	MAP@10
Требования ПС ↓ sim() Раздел + тема дисциплины	0,867	0,833	0,816	0,765
Требования ПС ↓ sim() Компетенция + результат обучения	0,905	0,883	0,875	0,813

По результатам оценки качества можно отметить, что после 5 ближайших значений релевантность сопоставленных элементов начинает падать. В дальнейших расчетах в рамках эксперимента использовалось 5 ближайших сопоставлений для каждого требования профстандартов.

Визуализация результатов семантического сопоставления требований профстандарта «Программист» с содержанием дисциплины «Проектирование и разработка распределенных программных систем» представлена на рис. 5. Для наглядности визуального представления выполнена агрегация до уровня разделов дисциплины. С каждым разделом связаны 5 самых близких по значению косинусной семантической близости требований профстандарта.

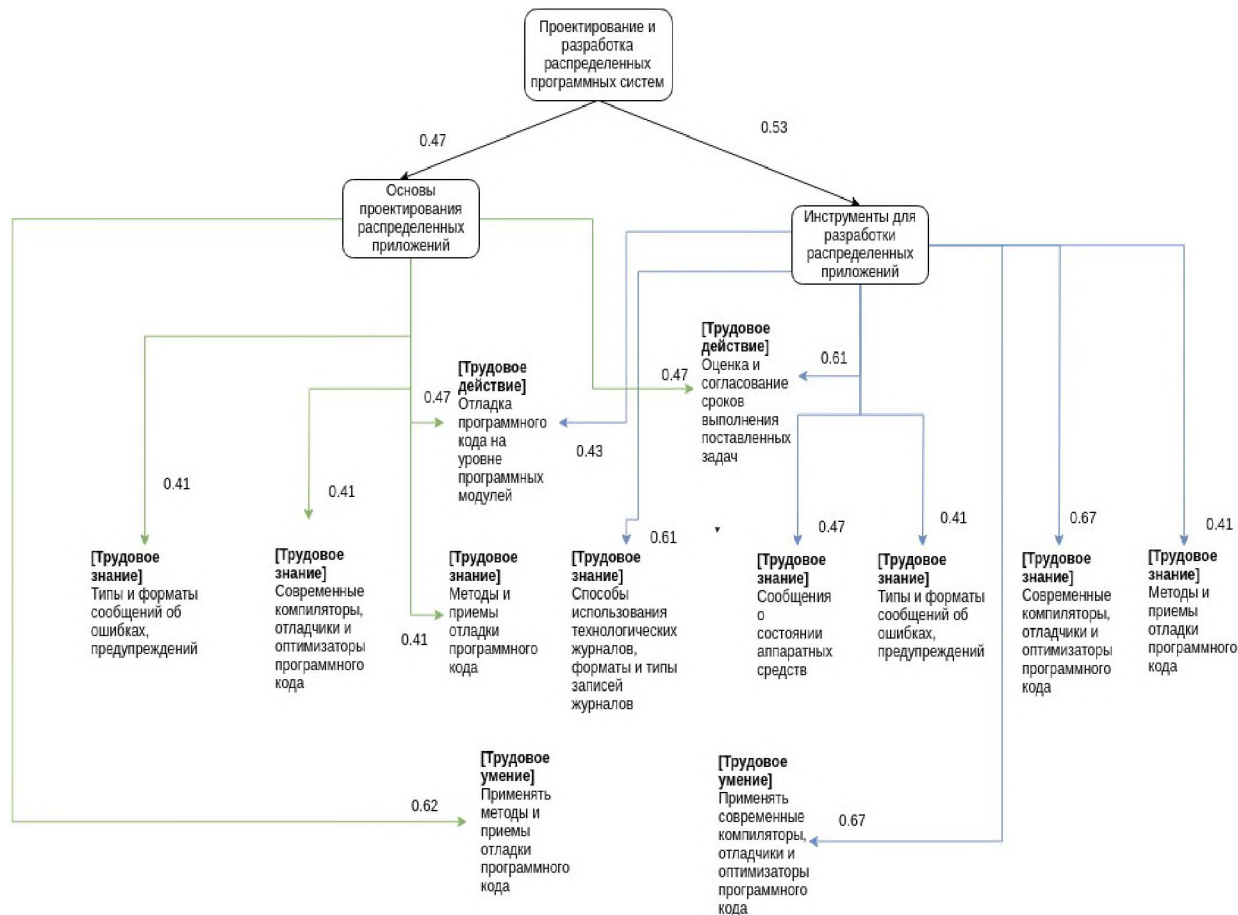


Рис. 5. Пример сопоставления требований профстандарта «Программист» с содержанием дисциплины «Проектирование и разработка распределенных программных систем»

На рис. 6 представлен фрагмент визуализации графовой модели образовательной программы с агрегацией до уровня дисциплин и сопоставленных им требований профстандартов с агрегацией до трудовых функций (с указанием профессии). Агрегация выполняется путем усреднения семантической близости сопоставленных сущностей. По представленным на иллюстрации для примера дисциплинам из образовательной программы можно отметить, что дисциплина «Базы данных» имеет самый высокий вес по результатам агрегации сопоставления, что говорит о важности данной дисциплины для всех трех профессий «Программист», «Специалист по информационным системам», «Системный администратор». Также можно отметить, что ни один элемент дисциплины «Теория систем и системный анализ» не попал в топ-5 ближайших элементов по сопоставлениям с требованиями указанных профстандартов. Можно предложить перенести изучение данной дисциплины на следующий уровень образования – в магистерские программы, где были бы учтены требования по подготовке научных кадров.

Для улучшения качества сопоставления помимо ограничения числа ближайших элементов для формирования связей *match* могут быть выставлены абсолютные и относительные пороги по семантической близости.

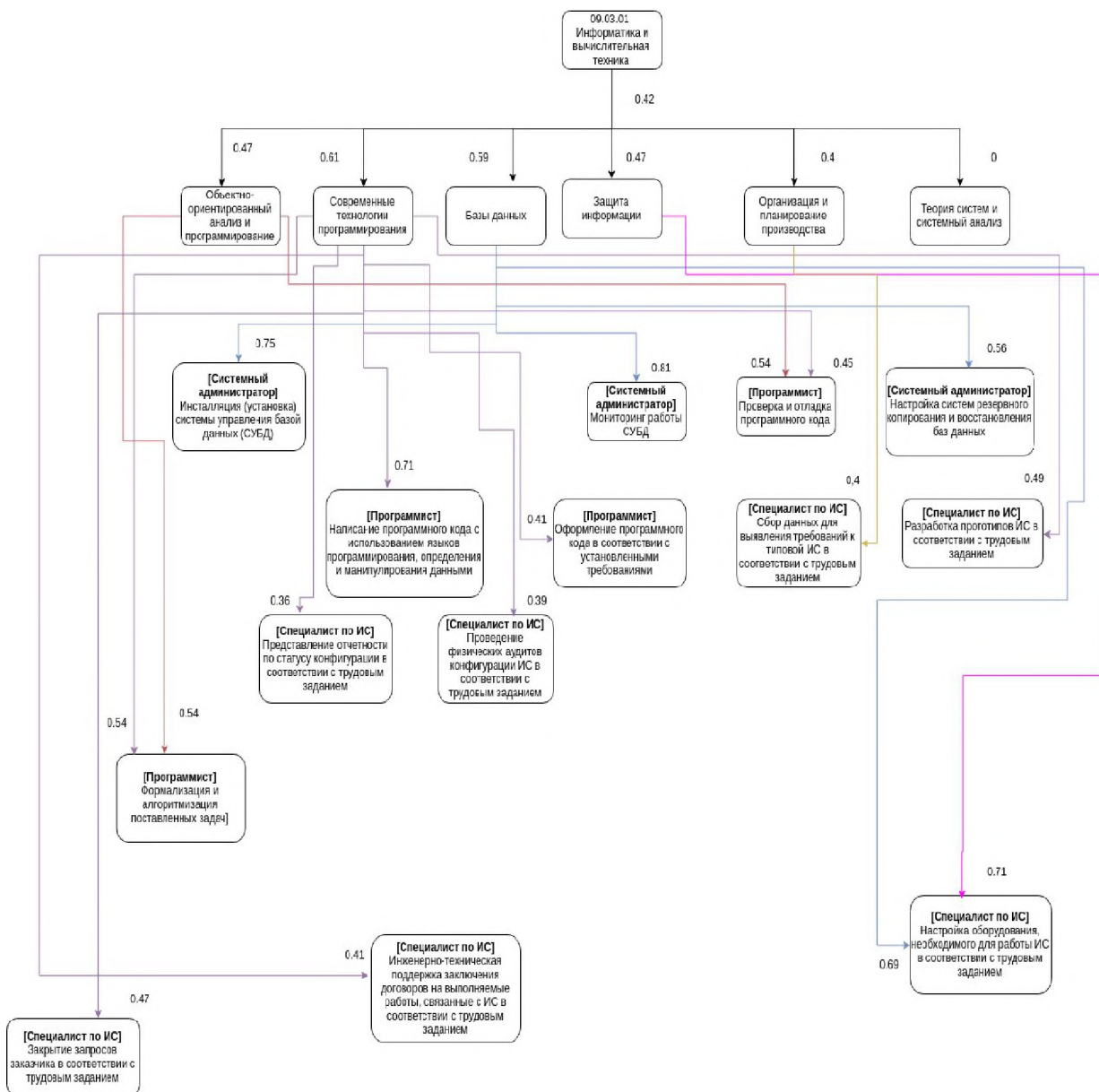


Рис. 6. Пример агрегации сопоставления дисциплин ОП по направлению 09.03.01 «Информатика и вычислительная техника» с требованиями ПС «Программист», «Специалист по информационным системам», «Системный администратор»

Заключение

В работе предложен интеллектуальный метод поддержки принятия решений при формировании образовательных программ с учетом потребностей рынка труда. В рамках метода предлагается использовать сочетание применения нейросетевых моделей языка word2vec, обученных на больших текстовых корпусах образовательной и профессиональной области, для оценки семантической близости сущностей графовых моделей документов образовательных программ, профессиональных стандартов и вакансий. Определяются шаги формирования рекомендаций по актуализации образовательных программ на основе анализа потребностей рынка труда, требований профстандартов, а также результатов обучения и содержания учебных дисциплин ведущих университетов и открытых онлайн-курсов.

Эксперимент по сопоставлению требований профстандартов с результатами обучения и содержанием дисциплин по программе «Информатика и вычислительная техника» иллюстрирует возможности применения данного подхода для актуализации содержания образовательных программ и повышения качества подготовки выпускников.

Помимо word2vec в качестве альтернативных способов определения семантической близости в других работах автора были также рассмотрены популярные модели fastText, paragraph2vec. Также в случае сопоставления сущностей из одной предметной области с одной и той же лексикой (например, сопоставление элементов различных программ учебных дисциплин) может быть более эффективной и традиционная векторная модель TF-IDF.

Исследование выполняется при поддержке Российского фонда фундаментальных исследований в рамках проекта № 18-47-860013 р_а «Интеллектуальная система формирования образовательных программ на основе нейросетевых моделей естественного языка с учетом потребностей цифровой экономики Ханты-Мансийского автономного округа – Югры» (договор № 18-47-860013/18).

Литература

1. Национальная технологическая инициатива «Кадровое обеспечение промышленного роста» (2018). – <https://asi.ru/staffing/> (дата обращения: 05.12.2018).
2. Автоматизированная система мониторинга и анализа кадровых потребностей по номенклатуре специальностей вуза / П.В. Зрелов, В.В. Кореньков, Н.А. Кутовский и др. // Федерализм. – 2016. – № 4 (84). – С. 63–76.
3. Стаин, Д.А. Квалификационно-ориентированная экспертная система управления образовательным процессом вуза в современных процессах непрерывного квалификационного развития кадров в России / Д.А. Стаин, Н.О. Вербицкая, Т.Г. Калугина // Вестник Южно-Уральского государственного университета. Серия «Образование. Педагогические науки». – 2018. – Т. 10, № 1. – С. 27–36. DOI: 10.14529/ped180104
4. Юсупова, Н.И. Организация информационной поддержки принятия решений при управлении образовательным маршрутом на основе онтологии / Н.И. Юсупова, О.Н. Сметанина, А.В. Климова // Труды Четвертой Международной научной конференции «Информационные технологии и системы» (Банное, 25 февраля – 01 марта 2015 г.). – Челябинск, 2015. – С. 109–111.
5. Chung, H. An Ontological Approach for Semantic Modeling of Curriculum and Syllabus in Higher Education / H. Chung, J. Kim // International Journal of Information and Education Technology. – 2016. – Vol. 6 (5). – P. 365–369. DOI: 10.7763/IJET.2016.V6.715
6. Chernikova, E. The Similarity Measure and Algorithm for Comparison of the Learning Outcomes / E. Chernikova, P. Nikolaev // Proceedings of the Fourth International Conference on Internet Technologies and Applications. – 2011. – P. 465–473.
7. Ужва, А.Ю. Автоматизированная разработка онтологической модели предметной области для поиска образовательных ресурсов с использованием анализа текстов рабочих программ / А.Ю. Ужва // Современные проблемы науки и образования. – 2013. – № 1. – <http://science-education.ru/ru/article/view?id=8324> (дата обращения: 05.12.2018).
8. Харитонов, И.М. Алгоритм формирования учебного плана с применением методики формализованного представления учебной дисциплины (на примере дисциплины «моделирование систем») / И.М. Харитонов // Вестник АГТУ. Серия «Управление, вычислительная техника и информатика». – 2011. – № 2. – С. 178–185.
9. Сибикина, И.В. Построение лингвистических шкал в целях выявления важных дисциплин, формирующих компетенцию / И.В. Сибикина, Н.Ю. Квятковская // Вестник Астраханского государственного технического университета. Серия «Управление, вычислительная техника и информатика». – 2012. – № 2. – С. 182–186.
10. Лисицына, Л.С. Автоматизация управления образовательными траекториями для разработки модульных компетентностно-ориентированных образовательных программ вуза / Л.С. Лисицына, А.С. Пирская // Сборник трудов Всероссийской научно-практической конференции с международным участием «Информационные технологии в обеспечении нового качества высшего образования». – М., 2010. – С. 75–86.
11. Котов, С.С. Управление структурой образовательных программ компетентностного содержания с учетом нечетких социальных предпочтений / С.С. Котов, В.Ю. Столбов // Системы управления и информационные технологии. – 2009. – № 1.3. – С. 411–416.
12. Distributed Representations of Words and Phrases and their Compositionality / T. Mikolov, I. Sutskever, K. Chen et al. // Advances in neural information processing systems. – 2013. – P. 3111–3119.

13. *RUSSE: The First Workshop on Russian Semantic Similarity* / A. Panchenko, N.V. Loukachevitch, D. Ustalov et al. // *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue 2015 (Moscow, RGGU)*. – Moscow, 2015. – Vol. 2. – P. 89–105.

14. *RUSSE'2018: A Shared Task on Word Sense Induction for the Russian Language* / A. Panchenko, A. Lopukhina, D. Ustalov et al. // *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue 2018 (Moscow, RGGU)*. – Moscow, 2018. – P. 547–564.

15. Lilleberg, J. *Support vector machines and word2vec for text classification with semantic features.* / J. Lilleberg, Y. Zhu, Y. Zhang // *Proceedings of International Conference on Cognitive Informatics & Cognitive Computing (Beijing, China)*. – IEEE 14th, 2015. – P. 136–140. DOI: 10.1109/ICCI-CC.2015.7259377

Ботов Дмитрий Сергеевич, старший преподаватель кафедры информационных технологий и экономической информатики, заведующий лабораторией машинного обучения и интеллектуального анализа данных, Челябинский государственный университет, г. Челябинск; dmbotov@gmail.com.

Поступила в редакцию 25 декабря 2018 г.

DOI: 10.14529/ctcr190101

INTELLIGENT SUPPORT DEVELOPMENT OF EDUCATIONAL PROGRAMS BASED ON THE NEURAL LANGUAGE MODELS TAKING INTO ACCOUNT OF THE LABOR MARKET REQUIREMENTS

D.S. Botov, dmbotov@gmail.com

Chelyabinsk State University, Chelyabinsk, Russian Federation

The active development of the digital economy today imposes high requirements on the adaptability, practical orientation and quality of educational content. Existing approaches to the intelligent decision support of the formation of educational programs based on ontological models, expert systems and heuristic algorithms do not allow effectively taking into account and tracking changes both in the labor market and in the open educational content space in the Massive Open Online Courses (MOOC). Instead, it is proposed to use approaches to the semantic analysis based on the well-known neural network language model word2vec, which is trained without supervision on large text corpora. The complexity of semantic analysis is to move the definition of semantic similarity measures for short texts of the extracted entities (course topics, learning outcomes, job requirements, etc.) to matching of large structured documents, such as professional standard, an educational program. To take into account the interrelations of entities, a graph model is introduced for representing the educational and professional domain. The paper proposes an artificial intelligent method of forming recommendations for the actualization of the learning outcomes and content of educational programs. At the first stage, the actual requirements of the labor market are determined based on a semantic matching of job requirements with the content of professional standards. The second stage includes a semantic matching of the content of academic disciplines with the requirements of the labor market. At the third stage, a semantic search of relevant educational content is carried out among the programs of disciplines of leading universities and massive open online courses (MOOC). During the fourth stage, final recommendations on updating the educational program are formed. The experiment demonstrated the possibility of applying the method for matching learning outcomes and content of disciplines with the requirements of professional standards and evaluation using the example of the educational program (bachelor degree) of computer science and engineering.

Keywords: data mining, natural language processing, neural language models, semantic similarity, word2vec, fastText, educational program, professional standard, labor market.

References

1. *Skills Development for Industrial Growth (2018)*. Available at: <https://asi.ru/eng/staffing/> (accessed 5 December 2018).
2. Zrellov P.V., Korenkov V.V., Kutovskiy N.A., Petrosyan A.S., Rumyantsev B.D., Semenov R.N., Filozova I.A. [Automated System for Monitoring and Analysis of Compliance of the Labour Resources Needs According the Specialties' Nomenclature of Higher Educational Institution]. *Federalism*, 2016, no.4 (84), pp. 63–76. (in Russ.)
3. Stain D.A., Verbitskaya N.O., Kalugina T.G. [Qualification-Oriented Expert System of Management of Educational Process of Higher Education in Modern Processes of Continuing Qualification Development of Personnel in Russia]. *Bulletin of the South Ural State University. Ser. Education. Educational Sciences*. 2018, vol. 10, no. 1, pp. 27–36. (in Russ.) DOI: 10.14529/ped180104
4. Yusupova N.I., Smetanina O.N., Klimova A.V. [Organization of Information Decision Support in the Management of Educational Routes Based on Ontology]. *Trudy Chetvertoy Mezhdunarodnoy nauchnoy konferentsii "Informatsionny'e tekhnologii i sistemy"* [Proc. 4th Int. Conf. "Information Technology and Systems"], Chelyabinsk, 2015, pp. 109–111. (in Russ.)
5. Chung H., Kim J. An Ontological Approach for Semantic Modeling of Curriculum and Syllabus in Higher Education. *International Journal of Information and Education Technology*, 2016, vol. 6 (5), pp. 365–369. DOI: 10.7763/IJiet.2016.V6.715
6. Chernikova E., Nikolaev P. The Similarity Measure and Algorithm for Comparison of the Learning Outcomes. *Proc. of the Fourth International Conference on Internet Technologies and Applications*, 2011, pp. 465–473.
7. Uzhva A.Yu. [Automatic Development of Ontology Model for Case-Based Reasoning in Search of Educational Resources Using Analysis of Education Programs]. *Modern Problems of Science and Education*, 2013, no. 1. Available at: <http://science-education.ru/ru/article/view?id=8324> (accessed 5 December 2018). (in Russ.)
8. Kharitonov I.M. [The Study Plan Forming Algorithm Based on the Study Discipline Formalized Presentation Procedure (by the Example of "System Simulation" Discipline)]. *Bulletin of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*, 2011, no. 2, pp. 178–185. (in Russ.)
9. Sibikina I.V., Kvyatkovskaya I.Y. [Construction of Linguistic Scales with the Purpose of Revelation of Important Disciplines Developing the Competence]. *Bulletin of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*, 2012, no. 2, pp. 182–186. (in Russ.)
10. Lisitsyna L.S., Pirskaya A.S. [Automation of Management of Educational Trajectories for the Development of Modular Competence-Oriented Educational Programs of the University]. *Sbornik trudov Vserossiyskoj nauchno-prakticheskoy konferentsii s mezhdunarodnym uchastiem. Informatsionnye tekhnologii v obespechenii novogo kachestva vysshego obrazovaniya*. [In Proceedings of the All-Russian Scientific-Practical Conference with International Participation "Information Technology in Providing a New Quality of Higher Education"]. Moscow, 2010. pp. 75–86. (in Russ.)
11. Kotov S.S., Stolbov V.Yu. [Management of the Structure of Educational Programs of Competence Content Taking into Account of Fuzzy Social Preferences]. *Management Systems and Information Technology*, 2009, no. 1.3, pp. 411–416. (in Russ.)
12. Mikolov T., Sutskever I., Chen K., Corrado G.S., Dean J. Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 2013, pp. 3111–3119.
13. Panchenko A., Loukachevitch N.V., Ustalov D., Paperno D., Meyer C.M., Konstantinova N. RUSSE: The First Workshop on Russian Semantic Similarity. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue 2015*, RGGU, Moscow, 2015, vol. 2, pp. 89–105.
14. Panchenko A., Lopukhina A., Ustalov D., Lopukhin K., Arefyev N., Leontyev A., Loukachevitch N. RUSSE'2018: A Shared Task on Word Sense Induction for the Russian Language. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue 2018*, RGGU, Moscow, 2018, pp. 547–564.

15. Lilleberg J., Zhu Y., Zhang Y. Support Vector Machines and Word2vec for Text Classification with Semantic Features. *Proceedings of International Conference on Cognitive Informatics & Cognitive Computing*, Beijing, China, IEEE 14th, 2015, pp. 136–140. DOI: 10.1109/ICCI-CC.2015.7259377

Received 25 December 2018

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Ботов, Д.С. Интеллектуальная поддержка формирования образовательных программ на основе нейросетевых моделей языка с учетом требований рынка труда / Д.С. Ботов // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». – 2019. – Т. 19, № 1. – С. 5–19. DOI: 10.14529/ctcr190101

FOR CITATION

Botov D.S. Intelligent Support Development of Educational Programs Based on the Neural Language Models Taking into Account of the Labor Market Requirements. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2019, vol. 19, no. 1, pp. 5–19. (in Russ.) DOI: 10.14529/ctcr190101
