

О. И. Федяев, Н. В. Мелешченко
ФГБОУ ВО «Донецкий национальный технический университет»
283001, РФ, ДНР, г. Донецк, ул. Артёма, 58

РОЛЕВЫЕ МОДЕЛИ АГЕНТОВ СИСТЕМЫ МОДЕЛИРОВАНИЯ ПРОЦЕССА ОБНОВЛЕНИЯ УЧЕБНЫХ ДИСЦИПЛИН С УЧЁТОМ ТРЕБОВАНИЙ ПРЕДПРИЯТИЙ

O. I. Fedyaev, N. V. Meleshchenko
Donetsk National Technical University
Russian Federation, DPR,
283001, Donetsk, Artema st., 58

ROLE MODELS OF AGENTS OF THE SYSTEM FOR MODELING THE PROCESS OF UPDATING ACADEMIC DISCIPLINES IN ACCORDANCE WITH THE REQUIREMENTS OF ENTERPRISES

Формализован процесс извлечения новых компетенций из текстов рекомендаций предприятий к выпускникам университета. Это позволит своевременно обновлять учебные программы дисциплин выпускающей кафедры университета с учётом требований рынка труда. Поставленная задача решена путём компьютерной обработки текстов рекомендаций на естественном языке методами машинного обучения. Алгоритм её решения реализует специальный программный агент с BDI-архитектурой во взаимодействии с другими агентами, имитирующими роли предприятий и преподавателей на основе принципа ограниченной рациональности. Проведены экспериментальные исследования разработанных алгоритмов и программ.

Ключевые слова: кафедра университета, требования предприятий, учебные программы дисциплин, извлечение знаний из текста, машинное обучение, программные агенты.

The process of extracting new competencies from the texts of recommendations of enterprises to university graduates has been formalized. This will make it possible to update the curricula of the disciplines of the graduating department of the university in a timely manner, taking into account the requirements of the labor market. The task was solved by computer processing of the texts of recommendations in natural language using machine learning methods. The algorithm for its solution implements a special software agent with a BDI architecture in interaction with other agents that simulate the roles of enterprises and teachers based on the principle of limited rationality. Experimental studies of the developed algorithms and programs have been carried out.

Key words: university department, requirements of enterprises, academic programs of disciplines, knowledge extraction from text, machine learning, software agents.

Введение

Одной из важных проблем системы высшего образования является несоответствие между компетенциями, получаемыми выпускниками в учебных заведениях, и необходимыми профессиональными умениями для трудоустройства. Таким образом, возникает необходимость в регулярной и профессионально-ориентированной кооперации учебных заведений и предприятий для решения данной проблемы.

Как известно, уровень профессиональной подготовки выпускников во многом определяется рабочими программами дисциплин. Их содержание инспектирует учебно-методическая комиссия кафедры на предмет их соответствия государственному стандарту по направлению подготовки. Важной функцией учебно-методической комиссии является своевременная актуализация содержания рабочих программ дисциплин, которая должна способствовать подготовке настоящих профессионалов, востребованных на рынке труда. Рабочие программы при актуализации должны периодически обновляться в соответствии с требованиями современной цифровой экономики и согласовываться с работодателями. Учебно-методическая комиссия кафедры анализирует требования работодателей и даёт рекомендации соответствующим лекторам на корректировку рабочих программ, которых они затрагивают. Участники этого процесса (предприятия, методическая комиссия кафедры, лекторы) образуют распределённую систему, для которых характерна территориальная удалённость, автономность и функциональная неоднородность (рис. 1). Взаимодействие участников в этой системе осуществляется на уровне смыслового анализа текстовых документов (рекомендации предприятий, рабочие программы дисциплин) и синтеза названий новых направлений (компетенций) для корректируемых рабочих программ. Эти задачи решаются методами обработки естественного языка (NLP) [1], [2], образующих одно из направлений машинного обучения.

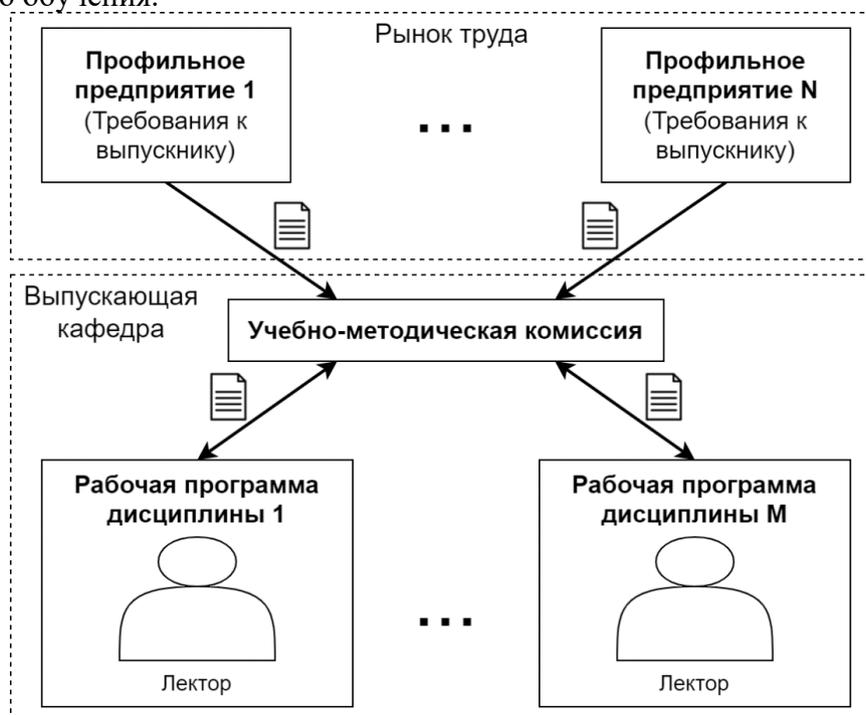


Рисунок 1 – Схема взаимодействия кафедры с профильными предприятиями по обновлению содержания рабочих программ дисциплин

Автоматизация решения этих задач является целью данной работы. Функционал системы автоматизации должен реализовывать интеллектуальный анализ требований предприятий и по критерию смысловой близости определять рабочую программу дисциплины, которую необходимо пополнить новыми знаниями (компетенциями), извлечёнными из текста рекомендаций предприятия. Для достижения поставленной цели выполнена формализация указанных задач, разработаны алгоритмы их решения и проведен ряд экспериментальных исследований.

Постановка общей задачи. Пусть имеется множество учебных программ дисциплин P , где каждый элемент $p \in P$ является текстовым документом, определяющим содержание и структуру учебного процесса по изучению конкретной учебной дисциплины с целью получения профессиональных знаний.

Пусть R будет множеством требований от предприятий, где каждый элемент $r \in R$ является также текстовым документом, в котором представлены необходимые для работы предприятия компетенции. Знания, содержащиеся в тексте t ($t \in P \vee t \in R$), представим как множество K при помощи функции извлечения знаний (назовём её $knwl$):

$$K = knwl(t).$$

Элементами множества K могут быть: понятия, методы, факты, закономерности, концепции, гипотезы и другие элементы знаний. Цель заключается в получении новых знаний K_{new} для учебной программы дисциплины $p \in P$ из требований предприятий $r \in R$:

$$K_{new} = new(r, p),$$

где new – функция получение новых знаний K_{new} для учебной программы p из требований предприятия r . Определение новых знаний нами сведено к определению разности множеств знаний из требований предприятия и соотнесенной с ними учебной программы дисциплины $p_{sim} \in P$:

$$K_{new} = K_r \setminus K_{p_{sim}},$$

где p_{sim} – это учебная программа, предмет изучения которой, наилучшим образом (по содержанию и сути) соответствует профессиональным требованиям предприятия r . Для установления такого соотношения между требованиями и дисциплиной необходимо определить специальную функцию $sim: R \rightarrow P$, которая каждому требованию $r \in R$ ставит в соответствие наиболее близкую в профессиональном плане учебную программу дисциплины $p \in P$. В таком случае определение новых знаний для найденной учебной программы дисциплины может быть представлено в следующем виде:

$$K_{new} = new(r, p_{sim}) = new(r, sim(r)) = knwl(r) \setminus knwl(sim(r)).$$

Таким образом, данная задача предполагает для известного множества учебных программ P и предложенного текстового документа с требованиями предприятия $r \in R$ определение наиболее подходящей дисциплины $p_{sim} \in P$ и выделение для неё новых знаний K_{new} . Как видно, возникают две следующие подзадачи:

- 1) определение функции sim соотнесения требований предприятия с одной из учебных программ;
- 2) определение функции $knwl$ извлечения знаний из текста.

Сопоставление требований предприятия и рабочей программы дисциплины. Чтобы сопоставить требование и дисциплину необходимо воспользоваться определённой метрикой, которая будет показывать степень отношения требования к каждой из дисциплин, и выбрать отношение с максимальным значением данной метрики [3].

Определим метрику $M(r, p)$, которая для каждого требования $r \in R$ и каждой учебной программы дисциплины $p \in P$ будет отражать их схожесть по смыслу в виде вещественного числа с областью значений $E(M(r, p)) = [-1; 1]$. Тогда задача соотнесения требований предприятия с учебной программой будет состоять в том, чтобы найти такую дисциплину p , для которой значение метрики $M(r, p)$ будет максимально, т.е.:

$$\forall r \in R \langle M(r, \text{sim}(r)) = \max_{p \in P} M(r, p) \rangle.$$

Так как учебные программы дисциплин и требования от предприятий являются текстовыми документами, то для сопоставления дисциплины и требования были применены методы обработки естественного языка (Natural Language Processing, NLP) [4]. В методе NLP первый шаг заключается в переходе от слов к числам, т. к. компьютер не умеет работать с текстом в явном виде, поэтому анализируемые тексты были оцифрованы. Для такого преобразования рассматривались два основных вида моделей числового представления текста:

- модели, основанные на подсчёте слов (BOW, TF-IDF [5]);
- модели вложений слов (Word2Vec, Doc2Vec, FastText, Glove).

Для выбора конкретного векторного представления необходимо учитывать смысл метрики, по которой определяется отношение требования к дисциплине. Модели, основанные на подсчёте слов, не учитывают их порядок и смысл использования, а модели вложения слов как раз основываются на контексте слов и их смысле. В данной задаче в качестве метрики целесообразно использовать схожесть документов по смыслу, поэтому была выбрана модель вложений слов. Для расчёта метрики в этом случае достаточно оценить косинусное расстояние [3] между векторами документов требований и учебной программы дисциплины:

$$M(r, p) = \frac{\text{vec}(r) * \text{vec}(p)}{\|\text{vec}(r)\| * \|\text{vec}(p)\|},$$

где vec – функция преобразования текстового документа в вектор заданной длины.

Основу моделей вложений составляют нейронные сети и их обучение на немаркированных данных. Были рассмотрены ряд моделей этого класса. Модель Word2Vec предназначена для кодирования смысла слов в векторах фиксированной размерности при помощи нейронной сети, которая предсказывает вхождения слов в контексте целевых слов. Модель Doc2Vec является дальнейшим развитием идеи Word2Vec на целые документы. В этом случае при предсказании учитываются не только слова, но и вектор документа, который является дополнительным входным сигналом. Модель FastText является улучшением идеи Word2Vec за счёт предсказания не только слов, но и n -символьных грамм (последовательностей из n символов), таким образом, появляется возможность работы с ранее неизвестными словами и орфографическими ошибками в тексте, однако размер такой модели и её вычислительная сложность кратно возрастает по сравнению с Word2Vec.

В ходе анализа различных подходов и алгоритмов был выбран алгоритм Doc2Vec, который позволяет создать векторное представление документа любой длины. После этапа обучения данный алгоритм позволяет генерировать векторы для неизвестных ранее документов. Следовательно, необходимо обучить модель на документах всех учебных дисциплин, а затем использовать обученную модель для создания векторов документов требований от предприятий. После этого можно использовать полученные векторы для поиска схожих документов, что и необходимо на этом этапе решения задачи. Программная реализация данного подхода выполнена в виде конвейера NLP (рис. 2).

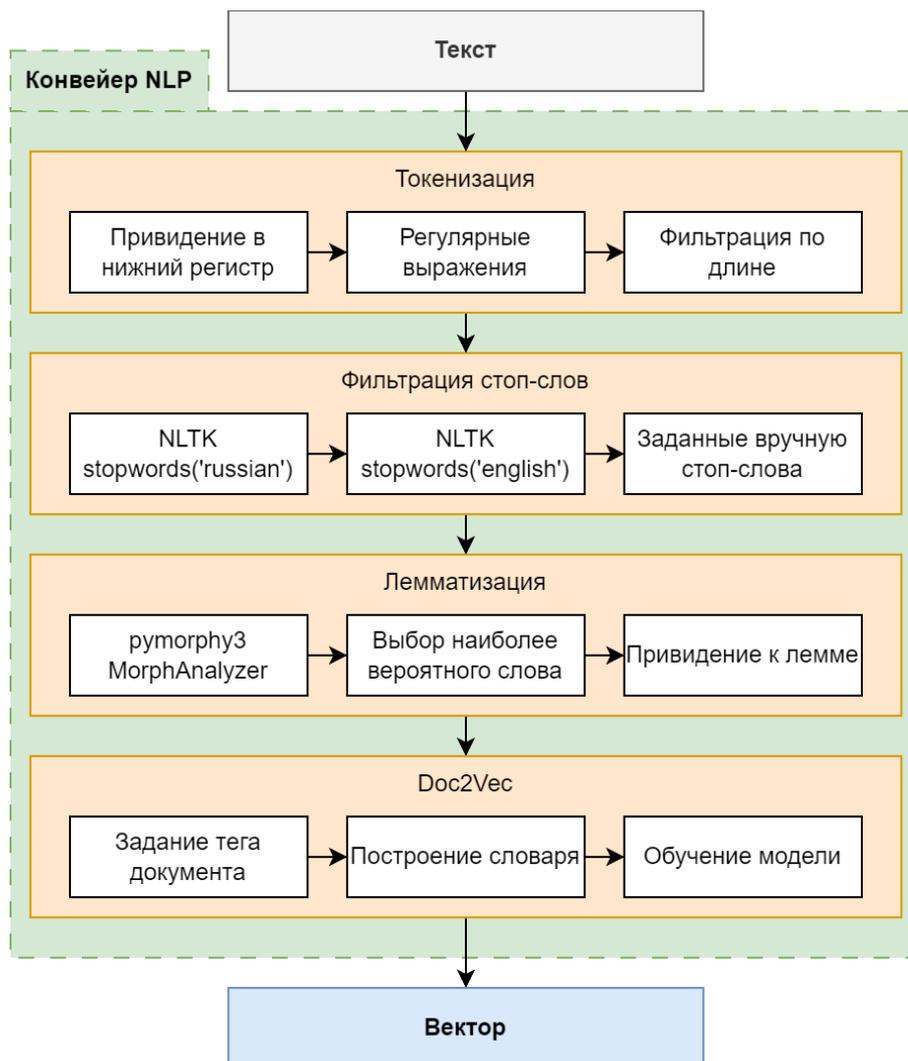


Рисунок 2 – Конвейер NLP для преобразования текстового документа в вектор вещественных чисел заданного размера n

В первую очередь исходный текст документа разделяется на токены. Для этого использовались регулярные выражения. Затем отфильтровываются стоп слова (предлоги, местоимения, междометия и т.п.), а также слишком короткие (менее 2 символов) и слишком длинные (более 25 символов) слова, чтобы уменьшить влияние очень частых и неинформативных слов. Далее проводилась лемматизация, т. е. приведение словоформы к леммам – их нормальным (словарным) формам [6].

Лемматизация снижает вариативность одного и того же слова, что уменьшает словарь и увеличивает точность анализа текста. После этого полученные токены передавались в модель Doc2Vec для преобразования документа в его смысловой вектор с заранее заданным размером n .

Представление знаний в текстовом документе. Чтобы дополнить учебную программу дисциплины новыми знаниями необходимо, во-первых, выделить их из текста требований предприятия, во-вторых, определить, какие из них являются новыми для учебной программы. Под термином «знания», которые извлекаются из текста, понимается формализованный набор сведений, получаемый путём распознавания в тексте сущностей, отношений и другой фактической информации, представленной в структу-

рированной форме и пригодной для машинной интерпретации [7], [8]. Поэтому в качестве модели представления знаний, находящихся в тексте, лучше всего использовать семантическую сеть [9], [10] в виде орграфа $G(V_g, E_g)$, где:

V_g – множество вершин, представляющих собой слова, которые встречаются в исходном тексте;

E_g – множество ребер, которые представляют синтаксические связи между словами в исходном тексте. $E_g = \{(a, syn, b) | a, b \in V_g, syn \in S\}$, где S – множество допустимых синтаксических связей между словами.

Таким образом, множество знаний K может быть представлено как множество ребер такого орграфа $K = E_g$. Для построения модели знаний в такой форме, предполагающей извлечение понятий и связей из текста, применяются следующие известные методы синтаксического анализа: на основе правил, статистические, на основе внешних баз знаний, на основе машинного обучения и гибридные модели [10-14].

В данной работе авторы воспользовались методом на основе семантико-синтаксических правил [10]. В данном методе текст представляется в виде синтаксического дерева, которое также можно описать орграфом $T(V_t, E_t)$, где:

V_t – множество вершин, которые представляют собой синтаксические единицы – слова (текстовое представление, словарная форма, часть речи, род, число и т.д.), которые встречаются в исходном тексте;

E_t – множество ребер, которые представляют связи между синтаксическими единицами в исходном тексте $E_t = \{(a, syn, b) | a, b \in V_t, syn \in S\}$. Очень важно, что синтаксические связи в этом графе отражают смысловые связи, выраженные средствами языка.

Следует заметить, что простой отбор ребер с нужной синтаксической связью из дерева синтаксического анализа не позволит извлечь знания из сложных синтаксических конструкций, поэтому необходимо учитывать и транзитивные связи (рис. 3). Применение семантико-синтаксических правил позволяет извлекать структурированные знания из исходного текста. Например, как показано на данном рисунке, с их помощью можно получить прилагательные «кластерного» и «дискриминантного», которые описывают имя существительное «анализа».

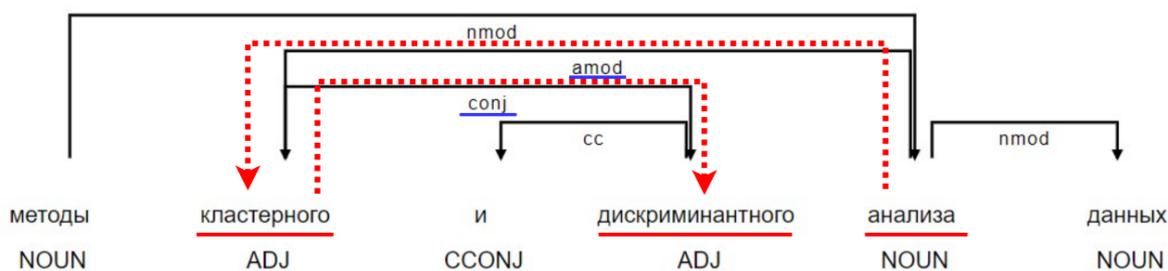


Рисунок 3 – Пример извлечения знаний из дерева синтаксического анализа текста с учётом транзитивных связей между словами

Для выявления недостающих знаний необходимо сравнить список знаний в тексте учебной программы дисциплины со списком знаний в тексте требований предприятия. Новыми будут те элементы знаний, которые содержатся только в тексте требований.

Для упрощения реализации операции сопоставления (сравнения) орграфов T для текстов требований и учебной программы дисциплины использовалась идеализированная модель знаний G , получаемая по семантико-синтаксическим правилам из графа T сохранением в нём только следующих связей S между словами:

- 1) существительное с прилагательным (amod);
- 2) существительное с существительным (nmod);
- 3) глагол с существительным (obj).

Эту операцию трансформирования полного орграфа T в сокращённый орграф G назовём $trans: G = trans(T)$.

Например, для предложения «Методы кластерного и дискриминантного анализа данных» элементы знаний, выделенные из этого текста, будут представлять собой список основных сущностей (слов) и связей между ними (табл.1), который можно представить в виде сокращённого орграфа G (рис. 4). В таблице хранится информация о трёх существительных («анализ», «данные» и «метод»), двух прилагательных («дискриминантный» и «кластерный») и связях между ними.

Таблица 1 – Табличное представление списка синтаксических связей в предложении «Методы кластерного и дискриминантного анализа данных»

Токен	Отношение/часть речи	Токен
дискриминантный	ADJ	дискриминантный
кластерный	ADJ	кластерный
анализ	NOUN	анализ
данные	NOUN	данные
метод	NOUN	метод
анализ	amod	дискриминантный
анализ	amod	кластерный
анализ	nmod	данные
метод	nmod	анализ

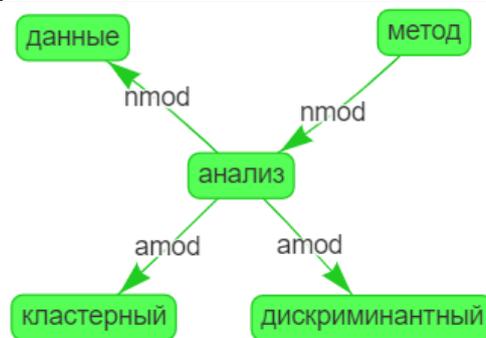


Рисунок 4 – Графическое представление орграфа G

Из этих связей можно сконструировать смысловые словосочетания, которые присутствуют в тексте требований и являются элементами новых знаний.

Выделение новых смысловых фрагментов в тексте требований. С целью обновления учебных программ дисциплин необходимо определять элементы знаний (последовательности слов) – новые смысловые фрагменты, т. е. выделять ближайший контекст для найденных в тексте требований новых слов [10]. В начале выделяются смысловые фрагменты путём модификации [13] дерева синтаксического анализа T_r :

$$T'_r = frag(T_r) ,$$

где $frag$ – функция объединения слов по семантико-синтаксическим правилам в смысловые фрагменты, результат которой наглядно показан на рис. 5.

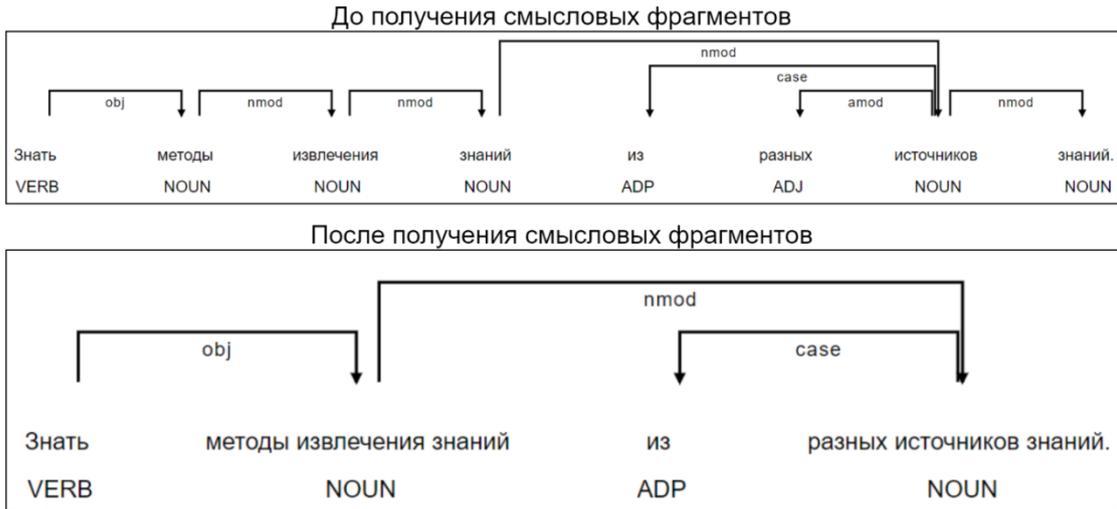


Рисунок 5 – Получение смысловых фрагментов на основе существительных

Далее получаем идеализированную модель текста требований G_r' , чтобы путём сравнения с соответствующей моделью учебной программы, извлечь новые элементы знаний:

$$G_r' = trans(T_r')$$

Так как при получении множества знаний K_r' из модифицированного дерева синтаксического анализа T_r' вершиной может быть целое словосочетание, то такая вершина может быть рассмотрена как элемент знания:

$$K_r' = \{(K_a, syn, K_b) | a, b \in V_r', syn \in S\}.$$

В таком случае для вычисления разности множеств $K_r' \setminus K_{p_sim}$, т. е. нахождения нового элемента знания, необходимо определить операцию проверки принадлежности элемента множества K_r' к множеству K_{p_sim} (рис. 6).

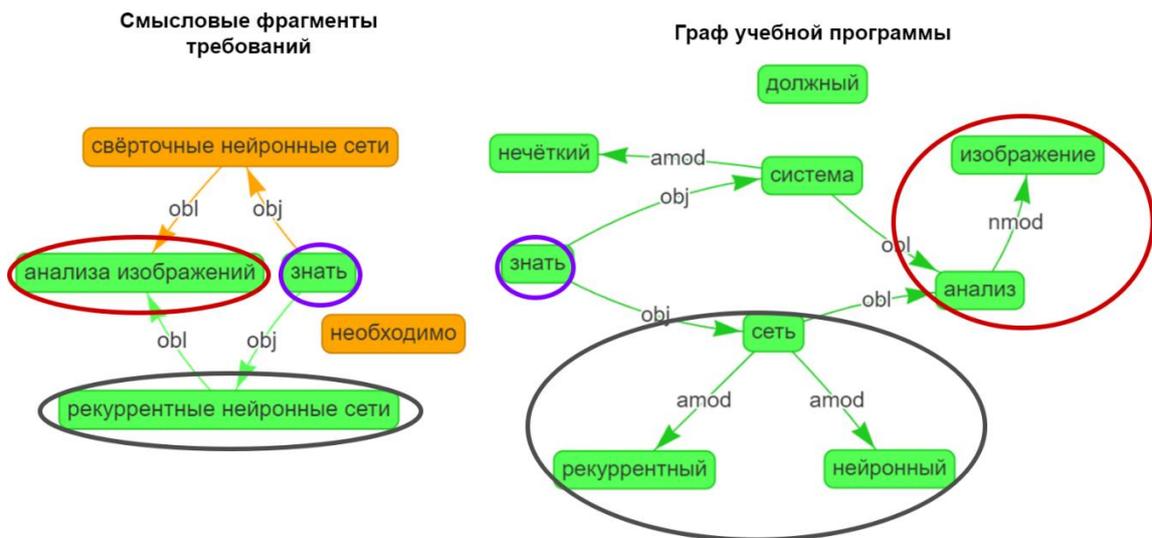


Рисунок 6 – Определение новых элементов знаний из требований предприятия: новый фрагмент выделен жёлтым цветом

Для этого достаточно проверить факт, что вершины K_a и K_b являются подмножествами K_{p_sim} , а также факт существования такого ребра $\langle m, syn, n \rangle \in K_{p_sim}$, который бы связывал понятия из K_a и K_b такой же связью syn :

$$\langle K_a, syn, K_b \rangle \in K_{p_sim} \Leftrightarrow K_a, K_b \subseteq K_{p_sim} \wedge (\exists \langle m, syn, n \rangle \in K_{p_sim} : m \in K_a \wedge n \in K_b).$$

Агентно-ориентированная структура системы. Как было уже отмечено, выпускающая кафедра университета и профильные предприятия, при взаимодействии по рассмотренным вопросам, образуют сложную распределённую систему, в которой участники индивидуальны по своей структуре, их деятельность интеллектуальна и динамична (рис. 1) [15], [16]. Поэтому программная модель этой системы строилась в виде взаимодействия интеллектуальных агентов [17], [18], реализующих соответствующие роли на основе принципа ограниченной рациональности (рис. 7) [19], [20]. В многоагентной модели взаимодействуют три типа агентов, которые выполняют следующие роли: агент Предприятие – предоставляет в виде текста свои рекомендации кафедре по улучшению подготовки студентов; агент Методист – принимает от агентов Предприятие тексты с требованиями и по ним определяет дисциплины и новые знания, которые посылает соответствующим агентам Лектор; агенты Лектор – предоставляют рабочие программы агенту Методист по его запросу и принимают от него новые знания в виде новых методов, технологий, инструментов, навыков и методик, извлечённых из требований предприятий, которые в дальнейшем могут быть использованы лектором для обновления соответствующих учебных программ дисциплин.

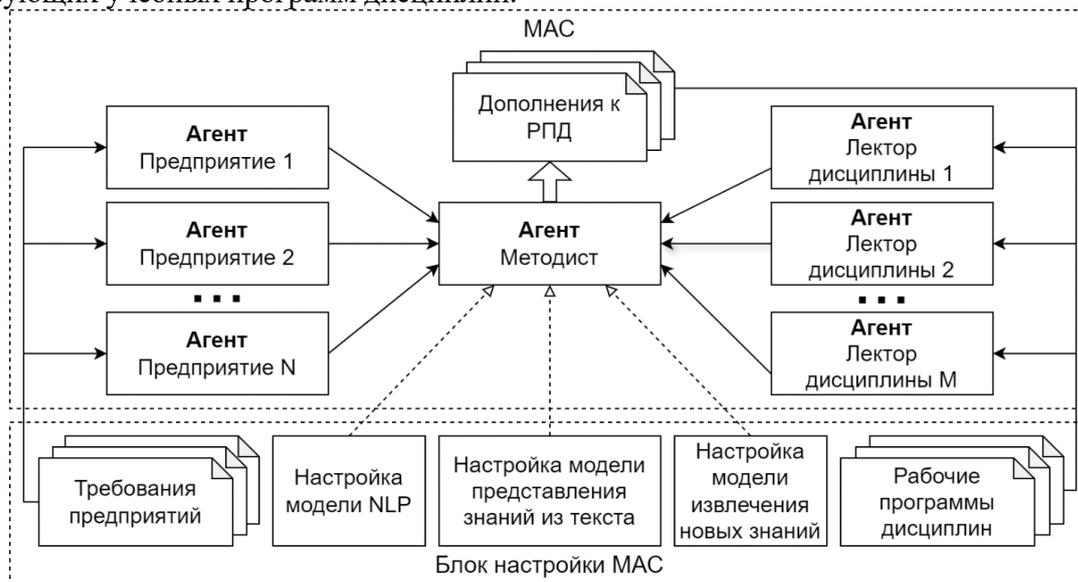


Рисунок 7 – Структура многоагентной системы актуализации содержания учебных программ дисциплин путем смыслового анализа требований предприятий

В модели используются две базы данных: одна для хранения текстов рабочих программ дисциплин кафедры, другая – для текстов рекомендаций предприятий по улучшению подготовки студентов. Для настройки программного агента Методист, имеющего BDI архитектуру [21], предусмотрена установка исходных значений ряда параметров для алгоритмов NLP, представления и извлечения знаний.

Модельный эксперимент по определению новых знаний. В качестве примера были проведены эксперименты по извлечению новых знаний из текста требований предприятия к специалисту по интеллектуальному анализу данных для одноимённой учебной дисциплины. Результат выделения новых знаний представлены в виде графа на рис. 8.

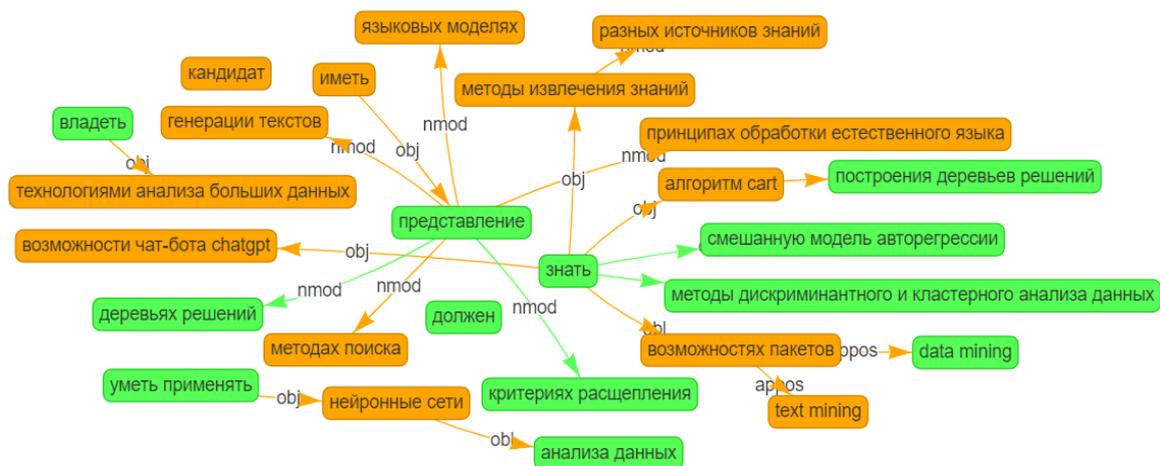


Рисунок 8 – Смысловые фрагменты требований для дисциплины «Интеллектуальный анализ данных»

Зелёным выделены смысловые фрагменты, которые уже известны в дисциплине «Интеллектуальный анализ данных», а оранжевым – неизвестные (новые), т.е. зелёные связи между фрагментами присутствуют в учебной программе, а оранжевые - отсутствуют.

Таким образом, система выделила следующие новые смысловые фрагменты (элементы знаний): *разных источников знаний, технологиями анализа больших данных, языковых моделях, методы извлечения знаний, принципах обработки естественного языка, кандидат, text mining, генерации текстов, нейронные сети, методах поиска, иметь, возможностях пакетов, возможности чат-бота chatgpt, алгоритм cart.*

В таком виде смысловые фрагменты передаются лектору дисциплины как рекомендация к обновлению учебной программы или добавлению новых разделов, обеспечивающих формирование у студентов современных компетенций.

Выводы

В работе математически обоснована модель роли ключевого агента Методист, выполняющего анализ требований предприятий с целью выявления из них новых знаний для дополнения и корректировки учебных программ дисциплин подготовки специалистов. Анализ требований и синтез новых знаний решаются путём компьютерной обработки текстов на естественном языке методами машинного обучения. В работе рассмотрено сопоставление текстов требования предприятия и учебных программ на основе векторного представления текста моделью Doc2Vec и косинусного подобию.

Для извлечения знаний из текста предложено использовать синтаксические правила, а для выявления недостающих знаний – формирование смысловых фрагментов на основе существительных. Рассмотрено представление извлечённых знаний в виде графа знаний.

Экспериментальные исследования разработанного алгоритмического и программного обеспечения для агента Методист показали возможность его применения в составе мультиагентной имитационной модели с целью улучшения адаптации учебного процесса к условиям реального рынка труда.

Таким образом, новизна предложенного подхода заключается в том, что разработаны методы и алгоритмы смыслового анализа текстов рекомендаций профильных предприятий с целью извлечения из них новых знаний, позволяющих своевременно обновлять учебные программы подготовки специалистов в соответствии с изменениями конъюнктуры рынка труда.

Список литературы

1. Федяев, О. И. Машинное обучение агентов для моделирования процесса обновления рабочих программ дисциплин с учётом требований предприятий / О. И. Федяев, Н. В. Мелещенко // Инжиниринг предприятий и управление знаниями (ИП&УЗ-2024) : сборник научных трудов XXVII Российской научной конференции : в 2 т., Москва, 28–29 ноября 2024 года. – Москва: Российский экономический университет имени Г.В. Плеханова, 2024. – С. 337-344. – EDN PFRTKS.
2. Мелещенко, Н. В. Определение семантической эквивалентности текстов требований предприятий и рабочих программ дисциплин / Н. В. Мелещенко, О. И. Федяев // Программная инженерия: методы и технологии разработки информационно-вычислительных систем (ПИИВС-2024) : V Международная научно-практическая конференция, сборник материалов и докладов, Т.1, г. Донецк, 27-28 ноября 2024 г. – Донецк, ФГБОУ ВО «ДонНТУ», 2024.- С. 158-165.
3. Андриевская, Н. К. Гибридная интеллектуальная мера оценки семантической близости / Н. К. Андриевская // Проблемы искусственного интеллекта. – 2021. – № 1(20). – С. 4-17. – EDN ZDZKGG.
4. Бурлаева, Е. И. Проект построения алгоритма классификации текстовых документов / Е. И. Бурлаева, В. Н. Павлыш // Проблемы искусственного интеллекта. – 2017. – № 4(7). – С. 24-31. – EDN YWZSON.
5. Бурлаева, Е. И. Сравнение некоторых методов машинного обучения для анализа текстовых документов / Е. И. Бурлаева, С. А. Зори // Проблемы искусственного интеллекта. – 2019. – № 1(12). – С. 42-51. – EDN WAKKYA.
6. Пикалев, Я. С. Разработка системы нормализации текстовых корпусов / Я. С. Пикалев // Проблемы искусственного интеллекта. – 2022. – № 2(25). – С. 64-78. – EDN CNHKBN.
7. Построение графов знаний нормативной документации на основе семантического моделирования и автоматического извлечения терминов / Д. И. Муромцев, И. А. Шилин, Д. А. Плюхин [и др.] // Научно-технический вестник информационных технологий, механики и оптики. – 2021. – Т. 21, № 2. – С. 256-266. – DOI 10.17586/2226-1494-2021-21-2-256-266. – EDN KCXQLE.
8. Методы и модели извлечения знаний из медицинских документов / Р. Х. Зулкарнеев, Н. И. Юсупова, О. Н. Сметанина [и др.] // Информатика и автоматизация. – 2022. – Т. 21, № 6. – С. 1169-1210. – DOI 10.15622/ia.21.6.4. – EDN ASOOVS.
9. Semantic Text Analysis Using Artificial Neural Networks Based on Neural-Like Elements with Temporal Signal Summation / A. Kharlamov, E. Samaev, D. Kuznetsov, D. Pantiukhin // Problems of Artificial Intelligence. – 2023. – No. 3(30). – P. 4-27. – DOI 10.34757/2413-7383.2023.30.3.001. – EDN YDFBKG.
10. Кленин, Ю. Д. Активное обучение для извлечения знаний из описаний образовательных курсов в условиях малых объёмов данных / Ю. Д. Кленин // Онтология проектирования. – 2019. – Т. 9, № 4(34). – С. 522-535. – DOI 10.18287/2223-9537-2019-9-4-522-535. – EDN NOWPRQ.
11. Кобышев, К. С. Анализ и классификация алгоритмов извлечения отношений из текстовых данных / К. С. Кобышев, С. А. Молодяков // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. – 2021. – № 5. – С. 71-79. – DOI 10.37882/2223-2966.2021.05.15. – EDN KXLLZK.
12. Мусаев, А. А. Обзор современных технологий извлечения знаний из текстовых сообщений / А. А. Мусаев, Д. А. Григорьев // Компьютерные исследования и моделирование. – 2021. – Т. 13, № 6. – С. 1291-1315. – DOI 10.20537/2076-7633-2021-13-6-1291-1315. – EDN HDOMZN.
13. Николаев, И. Е. Метод извлечения знаний и навыков/компетенций из текстов требований вакансий / И. Е. Николаев // Онтология проектирования. – 2023. – Т. 13, № 2(48). – С. 282-293. – DOI 10.18287/2223-9537-2023-13-2-282-293. – EDN DLDZBI.
14. Герасименко, Е. М. Алгоритм поиска и приобретения знаний на основе технологий обработки и анализа текстов на естественном языке / Е. М. Герасименко, Ю. А. Кравченко, Д. А. Шаненко // Известия ЮФУ. Технические науки. – 2024. – № 5(241). – С. 88-102. – DOI 10.18522/2311-3103-2024-5-88-102. – EDN GMUMBY.
15. Словохотов, Ю. Л. Распределенный интеллект мультиагентных систем. Ч. 1. Основные характеристики и простейшие формы / Ю. Л. Словохотов, Д. А. Новиков // Проблемы управления. – 2023. – № 5. – С. 3-22. – DOI 10.25728/ru.2023.5.1. – EDN CFQKJZ.
16. Словохотов, Ю. Л. Распределенный интеллект мультиагентных систем. Ч. 2. Коллективный интеллект социальных систем / Ю. Л. Словохотов, Д. А. Новиков // Проблемы управления. – 2023. – № 6. – С. 3-21. – DOI 10.25728/ru.2023.6.1. – EDN FYEXEW.
17. Листопад, С. В. Архитектуры интеллектуальных агентов сплоченных гибридных интеллектуальных многоагентных систем / С. В. Листопад, И. А. Кириков // Системы и средства информатики. – 2022. – Т. 32, № 2. – С. 81-91. – DOI 10.14357/08696527220208. – EDN HXTARF.
18. Листопад, С. В. Базовая архитектура рефлексивно-активных систем искусственных гетерогенных интеллектуальных агентов / С. В. Листопад // Информатика и ее применения. – 2024. – Т. 18, № 3. – С. 89-96. – DOI 10.14357/19922264240311. – EDN UNTQBV

19. Чернышев, С. А. Классификация общих шаблонов проектирования мультиагентных систем / С. А. Чернышев // Программные продукты и системы. – 2022. – № 4. – С. 670-679. – DOI 10.15827/0236-235X.140.670-679. – EDN XYYQOG.
20. Макаренко, С. И. Семантическая совместимость человеческих агентов при обеспечении интереса в сетевых системах / С. И. Макаренко // Журнал радиоэлектроники. – 2022. – № 1. – DOI 10.30898/1684-1719.2022.1.1. – EDN TQGBHN.
21. Шилов, Н. В. Алгоритмы для BDI-агентов, основанные на знаниях / Н. В. Шилов, Н. О. Гаранина // Моделирование и анализ информационных систем. – 2020. – Т. 27, № 4. – С. 442-453. – DOI 10.18255/1818-1015-2020-4-442-453. – EDN NRGOSJ.

References

1. Fedyaev O. I., Meleshchenko N. V. Machine learning of agents for Modeling the Process of Updating Academic Programs Considering Enterprises Requirements // Engineering of Enterprises and Knowledge Management (IP&UZ-2024): Proceedings of the XXVII Russian Scientific Conference: In 2 Volumes, Moscow, November 28–29, 2024. – Moscow: Plekhanov Russian University of Economics, 2024. – Pp. 337-344.
2. Meleshchenko N. V., Fedyaev O. I. Determining the Semantic Equivalence of Enterprise Requirements and Academic Program Texts // Software Engineering: Methods and Technologies for Developing Information and Computing Systems (PIVS-2024): V International Scientific and Practical Conference, Proceedings and Reports, Vol. 1, Donetsk, November 27-28, 2024. – Donetsk, Donetsk National Technical University, 2024. – Pp. 158-165.
3. Andrievskaya N. K. Hybrid Intelligent Measure for Assessing Semantic Similarity // Problems of Artificial Intelligence. – 2021. – No. 1(20). – Pp. 4-17.
4. Burlaeva E. I., Pavlysh V. N. Project for Building a Text Document Classification Algorithm // Problems of Artificial Intelligence. – 2017. – No. 4(7). – Pp. 24-31.
5. Burlaeva E. I., Zori S. A. Comparison of Some Machine Learning Methods for Text Document Analysis // Problems of Artificial Intelligence. – 2019. – No. 1(12). – Pp. 42-51.
6. Pikalev Y. S. Development of a Text Corpus Normalization System // Problems of Artificial Intelligence. – 2022. – No. 2(25). – Pp. 64-78.
7. Muromtsev D. I., Shilin I. A., Plyukhin D. A. [et al.]. Building Knowledge Graphs of Regulatory Documentation Based on Semantic Modeling and Automatic Term Extraction // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. – 2021. – Vol. 21, No. 2. – Pp. 256-266. – DOI 10.17586/2226-1494-2021-21-2-256-266.
8. Zulkarnaev R. Kh., Yusupova N. I., Smetanina O. N. [et al.]. Methods and Models for Knowledge Extraction from Medical Documents // Informatics and Automation. – 2022. – Vol. 21, No. 6. – Pp. 1169-1210. – DOI 10.15622/ia.21.6.4.
9. Kharlamov A., Samaev E., Kuznetsov D., Pantiukhin D. Semantic Text Analysis Using Artificial Neural Networks Based on Neural-Like Elements with Temporal Signal Summation // Problems of Artificial Intelligence. – 2023. – No. 3(30). – Pp. 4-27. – DOI 10.34757/2413-7383.2023.30.3.001.
10. Klenin Y. D. Active Learning for Knowledge Extraction from Educational Course Descriptions in Small Data Conditions // Ontology of Designing. – 2019. – Vol. 9, No. 4(34). – Pp. 522-535. – DOI 10.18287/2223-9537-2019-9-4-522-535.
11. Kobyshev K. S., Molodyakov S. A. Analysis and Classification of Algorithms for Relation Extraction from Text Data // Modern Science: Current Problems of Theory and Practice. Series: Natural and Technical Sciences. – 2021. – No. 5. – Pp. 71-79. – DOI 10.37882/2223-2966.2021.05.15.
12. Musaev A. A., Grigoriev D. A. Review of Modern Technologies for Knowledge Extraction from Text Messages // Computer Research and Modeling. – 2021. – Vol. 13, No. 6. – Pp. 1291-1315. – DOI 10.20537/2076-7633-2021-13-6-1291-1315.
13. Nikolaev I. E. Method for Extracting Knowledge and Skills/Competencies from Job Requirement Texts // Ontology of Designing. – 2023. – Vol. 13, No. 2(48). – Pp. 282-293. – DOI 10.18287/2223-9537-2023-13-2-282-293.
14. Gerasimenko E. M., Kravchenko Y. A., Shanenko D. A. Algorithm for Knowledge Search and Acquisition Based on Natural Language Text Processing and Analysis Technologies // Izvestiya SFedU. Engineering Sciences. – 2024. – No. 5(241). – Pp. 88-102. – DOI 10.18522/2311-3103-2024-5-88-102.
15. Slovokhotov Y. L., Novikov D. A. Distributed Intelligence of Multi-Agent Systems. Part 1. Main Characteristics and Simplest Forms // Problems of Control. – 2023. – No. 5. – Pp. 3-22. – DOI 10.25728/pu.2023.5.1.
16. Slovokhotov Y. L., Novikov D. A. Distributed Intelligence of Multi-Agent Systems. Part 2. Collective Intelligence of Social Systems // Problems of Control. – 2023. – No. 6. – Pp. 3-21. – DOI 10.25728/pu.2023.6.1.

17. Listopad S. V., Kirikov I. A. Architectures of Intelligent Agents in Cohesive Hybrid Intelligent Multi-Agent Systems // Systems and Means of Informatics. – 2022. – Vol. 32, No. 2. – Pp. 81-91. – DOI 10.14357/08696527220208.
18. Listopad S. V. Basic Architecture of Reflexive-Active Systems of Artificial Heterogeneous Intelligent Agents // Informatics and Its Applications. – 2024. – Vol. 18, No. 3. – Pp. 89-96. – DOI 10.14357/19922264240311.
19. Chernyshev S. A. Classification of Common Design Patterns for Multi-Agent Systems // Software Products and Systems. – 2022. – No. 4. – Pp. 670-679. – DOI 10.15827/0236-235X.140.670-679.
20. Makarenko S. I. Semantic Compatibility of Human Agents in Ensuring Interoperability in Network-Centric Systems // Journal of Radio Electronics. – 2022. – No. 1. – DOI 10.30898/1684-1719.
21. Shilov N. V., Garanina N. O. Knowledge-Based Algorithms for BDI Agents // Modeling and Analysis of Information Systems. – 2020. – Vol. 27, No. 4. – Pp. 442-453. – DOI 10.18255/1818-1015-2020-4-442-453.

RESUME

O.I. Fedyaev, N.V. Meleshchenko

Role models of agents of the system for modeling the process of updating academic disciplines in accordance with the requirements of enterprises

One of the important problems of the higher education system is the discrepancy between the competencies acquired by graduates in educational institutions and the necessary professional skills for employment. Thus, there is a need for regular and professionally oriented cooperation between educational institutions and enterprises to solve this problem. The participants in this process (enterprises, the methodological commission of the department, lecturers) form a distributed system characterized by territorial remoteness, autonomy and functional heterogeneity. The interaction of participants in this system is carried out at the level of semantic analysis of text documents (recommendations of enterprises, work programs of disciplines) and the synthesis of names of new areas (competencies) for adjusted work programs. These tasks are solved by natural language processing (NLP) methods. Automating the solution of these tasks is the purpose of this work.

The tasks of text data mining have been set. The initial objects of analysis are the following text documents in docx format: requirements (recommendations) from enterprises and curricula of the disciplines of the graduating department. Requirements analysis and synthesis of new knowledge are solved by computer text processing in natural language using machine learning methods. The paper considers the comparison of texts of enterprise requirements and educational programs based on the vector representation of the meaning of the text by the Doc2Vec model and cosine similarity. Syntactic rules were used to extract knowledge from the text, and the formation of semantic fragments based on nouns was used to identify new knowledge. The extracted knowledge was represented as a knowledge graph.

The structure of a multi-agent system has been developed for the intelligent analysis of enterprise requirements in order to update the work programs of the department's disciplines. The roles of artificial agents are formally described using mathematical apparatus. The correctness of algorithms for semantic analysis of texts and extraction of new knowledge has been experimentally confirmed. The system allows you to define a work program that is similar in meaning to each recommendation of the enterprise and extract new knowledge from it for the innovation of the relevant curricula of the department.

Conclusion: The novelty of the proposed approach lies in the fact that methods and algorithms for semantic analysis of the texts of recommendations of specialized enterprises have been developed in order to extract new knowledge from them, allowing timely updating of training programs for specialists in accordance with changes in labor market conditions.

РЕЗЮМЕ

О.И. Федяев, Н.В. Мелешенко

Рольевые модели агентов системы моделирования процесса обновления учебных дисциплин с учётом требований предприятий

Одной из важных проблем системы высшего образования является несоответствие между компетенциями, получаемыми выпускниками в учебных заведениях, и необходимыми профессиональными умениями для трудоустройства. Поэтому возникает необходимость в регулярной и профессионально-ориентированной кооперации учебных заведений и предприятий для решения данной проблемы. Участники этого процесса (предприятия, методическая комиссия кафедры, лекторы) образуют распределённую систему, для которых характерна территориальная удалённость, автономность и функциональная неоднородность. Взаимодействие участников в этой системе осуществляется на уровне смыслового анализа текстовых документов (рекомендации предприятий, рабочие программы дисциплин) и синтеза названий новых направлений (компетенций) для корректируемых рабочих программ. Эти задачи решаются методами обработки естественного языка (NLP). Автоматизация решения этих задач является целью данной работы.

Выполнена постановка задач интеллектуального анализа текстовых данных. Исходными объектами анализа являются следующие текстовые документы в формате docx: требования (рекомендации) от предприятий и учебные программы дисциплин выпускающей кафедры. Анализ требований и синтез новых знаний решаются путём компьютерной обработки текстов на естественном языке методами машинного обучения. В работе рассмотрено сопоставление текстов требования предприятия и учебных программ на основе векторного представления смысла текста моделью Doc2Vec и косинусного подобия. Для извлечения знаний из текста использовались синтаксические правила, а для выявления новых знаний – формирование смысловых фрагментов на основе существительных. Извлечённые знания представлялись в виде графа знаний.

Разработана структура многоагентной системы для интеллектуального анализа требований предприятий с целью обновления рабочих программ дисциплин кафедры. Роли искусственных агентов формально описаны с использованием математического аппарата. Экспериментально подтверждена правильность алгоритмов смыслового анализа текстов и извлечения новых знаний. Система позволяет определять близкую по смыслу рабочую программу для каждой рекомендации предприятия и извлекать из неё новые знания для инновации соответствующих учебных программ кафедры.

Новизна предложенного подхода заключается в том, что разработаны методы и алгоритмы смыслового анализа текстов рекомендаций профильных предприятий с целью извлечения из них новых знаний, позволяющих своевременно обновлять учебные программы подготовки специалистов в соответствии с изменениями конъюнктуры рынка труда.

Федяев Олег Иванович – к.т.н., доцент, Федеральное государственное бюджетное образовательное учреждение «Донецкий национальный технический университет», г. Донецк. *Область научных интересов:* искусственный интеллект, нейронные сети, компьютерное зрение, многоагентные системы. Эл. почта: olegfedyayev@mail.ru. Адрес: 283001, РФ, ДНР, г. Донецк, ул. Артёма, 58. Телефон: +7949 334 91 32.

Мелешенко Николай Владимирович – аспирант, Федеральное государственное бюджетное образовательное учреждение «Донецкий национальный технический университет», г. Донецк. *Область научных интересов:* искусственный интеллект, машинное обучение, многоагентные системы. Эл. почта: meleshenko.nikolay@gmail.com. Адрес: 283001, РФ, ДНР, г. Донецк, ул. Артёма, 58. Телефон: +7949 478 44 70.

Статья поступила в редакцию 21.01.2025.