

DOI: 10.17323/2587-814X.2023.2.71.84

# Интеллектуальный метод формирования списка требований профиля должности на основе нейросетевых моделей языка с использованием таксономии ESCO и корпуса онлайн-вакансий\*

И.Е. Николаев 

E-mail: ivan\_nikolaev@csu.ru

Челябинский государственный университет

Адрес: Россия, 454001, г. Челябинск, ул. Братьев Кашириных, д. 129

## Аннотация

Системы онлайн-рекрутмента за последние годы накопили огромное количество данных о реальном рынке труда. Особый интерес для исследования представляют данные о реальных требованиях рынка труда, содержащихся в текстах онлайн-вакансий, а также процесс их извлечения и структурирования для дальнейшего анализа и использования. Этап составления актуального списка требований для профиля должности в процессе подбора персонала является очень трудоемким и требует от HR-специалиста значительных усилий, связанных с мониторингом изменений целых отраслей и профессий, а также анализом востребованности и актуальности существующих на рынке требований. В данной статье автором предлагается концептуальная модель рекомендательной системы, позволяющая снизить нагрузку на HR-специалиста на этапе формирования актуального списка требований профиля должности в процессе подбора персонала. В основе модели предлагается использовать комбинацию графовой модели требований рынка труда на основе таксономии ESCO, адаптированной для русского языка, и интеллектуального метода формирования рекомендаций для составления актуального списка требований в процессе подбора персонала на основе нейросетевых моделей языка на архитектуре трансформеров, таксономии навыков ESCO и корпуса онлайн-вакансий российского рынка труда. Также в статье приводится концептуальный алгоритм работы рекомендательной системы и возможные варианты рекомендаций по актуализации списка требований профиля должности в процессе подбора персонала на основе анализа потребностей реального рынка труда.

\* Статья опубликована при поддержке Программы НИУ ВШЭ «Университетское партнерство»

**Ключевые слова:** анализ рынка труда, требования рынка труда, человеческие ресурсы, профиль должности, интеллектуальный анализ данных, обработка естественного языка, нейросетевые модели языка

**Цитирование:** Николаев И.Е. Интеллектуальный метод формирования списка требований профиля должности на основе нейросетевых моделей языка с использованием таксономии ESCO и корпуса онлайн-вакансий // Бизнес-информатика. 2023. Т. 17. № 2. С. 71–84. DOI: 10.17323/2587-814X.2023.2.71.84

## Введение

**В** настоящее время подавляющее число компаний закрывают существенную часть своих потребностей в кадрах посредством размещения онлайн-объявлений о вакансиях в системах онлайн-рекрутмента. В таких системах ежедневно генерируется и накапливается огромное количество структурированных и полуструктурированных данных о вакансиях и соискателях (резюме). Для примера можно привести две русскоязычные системы онлайн-рекрутмента hh.ru и superjob, первая насчитывает порядка 57 млн. резюме и более 40 млн. вакансий за период с 2010 по 2020 гг., вторая содержит более 12 млн. вакансий за период с 2010 по 2020 гг.

В этой связи, особенно актуальной становится проблема обработки и извлечения информации из данных онлайн-вакансий, поскольку ее решение позволит моделировать и понимать сложные явления на рынке труда (см., например, [1–6]).

Хотя системы онлайн-рекрутмента де-факто и стали основным источником для соискателей и менеджеров по подбору персонала, они по-прежнему демонстрируют недостатки в поиске, релевантности и точности, поскольку предложения о работе представлены на естественном языке и часто в нескольких синтаксически и лексически разных, но семантически близких формах. Это приводит к тому, что в процессе поиска поисковые запросы подвержены двусмысленности естественного языка и плохо сопоставляются с описанием должностей в онлайн-вакансиях. В частности, запросы, которые являются чрезмерно определенными или непоследовательными, часто не возвращают совпадений, в то время как соответствующие предложения о работе все еще можно было бы найти, если бы проблема согласованности или специфичности поисковых запросов была решена. Если точных совпадений не хватает, часто приходится принимать худшие альтернативы или идти на компромисс с первоначальными требованиями.

Еще одной проблемой является отсутствие инструментов, позволяющих человеку использовать извлеченные из онлайн-вакансий данные в своей профессиональной деятельности. К примеру, при разработке профиля должности в процессе формирования вакансии на замещение должности в компании специалисту HR-специалисту или начальнику соответствующего отдела необходимо потратить существенные усилия для того, чтобы отобрать несколько десятков, а то и сотен, аналогичных или близких вакансий, провести их анализ, извлечь из них список публикуемых требований и обязанностей, провести их анализ и ранжирование, и сверить их с должностными обязанностями той должности, для замещения которой разрабатывается новая вакансия.

Современные рекомендательные системы в целом, и в области рынка труда в частности, в значительной степени зависят от больших объемов ручной обработки данных и экспертных знаний, что делает их дорогостоящими, сложными в обновлении и подверженными ошибкам.

В статье предлагается концептуальная модель рекомендательной системы на основе следующих компонент:

- ◆ графовая модель требований рынка труда основе таксономии ESCO, адаптированной для русского языка;
- ◆ интеллектуальный метод формирования рекомендаций для составления списка требований в процессе подбора персонала на основе нейросетевых моделей языка с использованием таксономии навыков ESCO и корпуса онлайн-вакансий российского рынка труда;
- ◆ модель и концептуальный алгоритм работы автоматизированной рекомендательной системы для формирования рекомендации;
- ◆ возможные варианты рекомендаций по актуализации списка требований профиля должности на основе анализа потребностей рынка труда.

## 1. Анализ результатов предшествующих работ

В последние годы наблюдается возрастающий интерес к использованию методов искусственного интеллекта (ИИ) для анализа данных на рынке труда – «разведка рынка труда» (LMI, Labor Market Intelligence). LMI означает разработку и использование методов, алгоритмов и структур ИИ для анализа данных рынка труда, которые помогают с планированием политики и принятием решений [7–9].

Например, в работах [10, 11] исследования нацелены на создание рекомендательных систем, которые определяют соответствие между резюме соискателя и вакансией по отдельным компетенциям на уровне определения должности. Другие работы нацелены на определение востребованности определенных навыков [12], которые могут помочь соискателям определить свою образовательную траекторию или направление переподготовки и повысить свой уровень конкурентоспособности.

Благодаря быстрому развитию компьютерной лингвистики и инструментов для анализа текстов на естественном языке (NLP, Natural Language Processing), некоторые ученые пытаются анализировать изменения на рынке труда по текстам онлайн-вакансий на уровне отдельных компетенций [13–16]. Этот подход имеет много преимуществ, так как позволяет выявлять изменения на уровне конкретных профессий и специальностей, а также требований работодателей. Например, он позволяет осуществлять мониторинг онлайн-вакансий в разных регионах и странах в реальном времени, прогнозировать востребованность отдельных навыков, компетенций и технологий в рамках конкретных профессий или отраслей, а также быстро сравнивать аналогичные рынки труда в разных странах и регионах.

Проект Европейского центра по развитию профессионального образования (Cedefop) заслуживает особого внимания, поскольку его целью является сбор и классификация онлайн-вакансий для всей ЕС с помощью машинного обучения [17–19]. Кроме того, в рамках данного проекта проводятся исследования для выявления тенденций на рынке труда и прогнозирования востребованности отдельных навыков. Например, в работе [20] авторы, используя методы интеллектуального анализа текстов, анализируют литературу в категории «четвертая технологическая революция» и сравнивают результаты с новой версией классификации навы-

ков ESCO, чтобы определить в какой степени новая версия классификация навыков ESCO, создаваемая экспертами вручную, отражает тенденции, происходящие на реальном рынке труда.

## 2. Методы и материалы

### 2.1. Обзор европейской таксономии навыков ESCO

Способность извлекать ценные знания из больших объемов данных, таких как системы онлайн-рекрутмента, сильно зависят от наличия актуальных баз знаний, таксономии и тезаурусов. Такие ресурсы необходимы для эффективного применения методов машинного обучения и для решения большинства NLP и NLU задач (NLU, Natural Language Understanding).

В настоящее время в основе большого количества проектов по анализу рынка труда используется европейская классификация навыков ESCO. ESCO (европейские навыки, компетенции и профессии) – это многоязычная классификация европейских навыков, компетенций, квалификаций и профессий. В ней определяются и классифицируются навыки, компетенции, квалификации и профессии, соответствующие рынку труда ЕС, образованию и профессиональной подготовке, на 25 европейских языках. Система предоставляет профессиональные профили, показывающие взаимосвязи между профессиями, навыками, компетенциями и квалификацией. ESCO была разработана в открытом ИТ-формате, доступна через онлайн-портал для бесплатного использования.

ESCO структурирована на основе трех взаимосвязанных компонентов, представляющих базу данных с возможностью поиска на 28 языках. Этими основными элементами являются: а) профессиональные профили (профессии), б) навыки/компетенции/знания и в) квалификации, как показано на модели данных ESCO (рис. 1). Первый компонент – профессиональные профили (или профессии) содержит название, описание профессии и показывают, являются ли навыки и компетенции и знания необходимыми или необязательными (опциональными), и какие квалификации имеют отношение к каждой профессии. Второй компонент содержит информацию о знаниях, навыках и компетенциях, а также некоторые групповые концепции. В ESCO v.1 он содержит около 13500 концепций (а если включить альтернативные названия, то почти 100000 формулировок), организованных в иерархию, а также

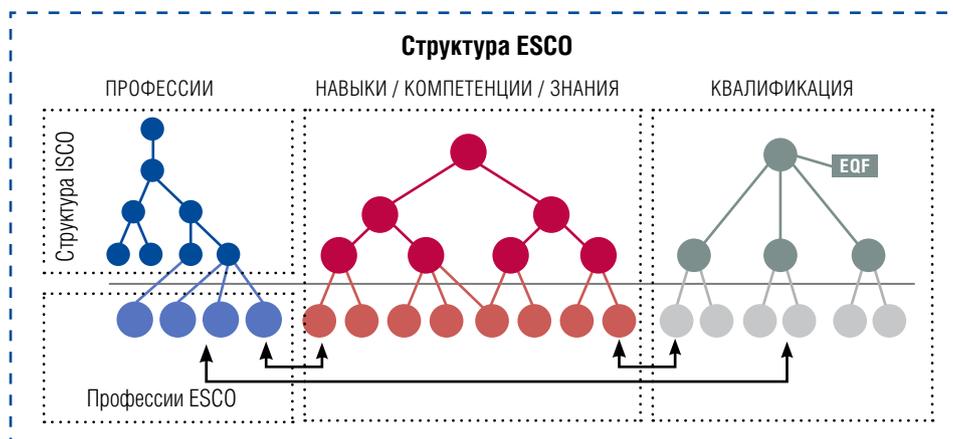


Рис. 1. Модель данных ESCO.

Источник: [https://ec.europa.eu/esco/portal/escopedia/ESCO\\_data\\_model](https://ec.europa.eu/esco/portal/escopedia/ESCO_data_model)

структурирован через связь с профессиями. Третий компонент – квалификаций позволяет государствам и присваивающим органам предоставлять данные о квалификациях, которые собираются в ESCO. Квалификации структурированы с использованием Европейской системы квалификаций (EQF) и ISCED Fields of Education and Training 2013.

В настоящий момент отдельными группами учебных предлагаются подходы и алгоритмы автоматического расширения таксономии навыков ESCO на основе открытых данных онлайн-вакансий [21, 22].

Например, для профессии «инженер по компьютерному оборудованию», шифр по ESCO 2152.1.1, в таксономии определено 22 альтернативных названия профессии, неполный список альтернативных названий приведен в *таблице 1*. Также для этой профессии определен ряд сущностей навык/компетенция и знания, которые необходимы для

этой профессии, каждый из которых также включает в себя список альтернативных названий на естественном языке. Так, например, для этой профессии определено: 47 основных и 25 дополнительных навыков/компетенций, 16 основных и 20 дополнительных знаний, Примеры названия для сущностей навыков/компетенций и знаний, а также их альтернативные названия представлены в *таблице 2*.

Важнейшее преимущество данной классификации заключается в том, что в ней используются формулировки названий профессии, названий навыков/компетенций и знаний, а также их альтернативные названия на естественном языке, что существенно упрощает и расширяет возможности ее применения для анализа текстов онлайн-вакансий реального рынка труда, которые также представлены на естественном языке, современными методами машинного обучения.

## 2. Модель рекомендательной системы формирования актуальных требований профиля должности

### 2.1. Обобщенная графовая модель требований рынка труда на основе классификации ESCO и данных из онлайн-вакансий

Представим модель рынка труда как ориентированный граф. За основу возьмем таксономию ESCO.

**Определение графовой модели рынка труда на основе таксономии ESCO.** Графовая модель представлена в виде кортежа из трех элементов  $E = (O, R, S)$ , где  $O = \{o_1, \dots, o_n\}$  – множество профессий,  $S = \{s_1, \dots, s_m\}$  – множество сущностей на-

Таблица 1.

#### Примеры альтернативных названий для профессии «инженер по компьютерному оборудованию»

Наименование
специалист по компьютерному оборудованию
инженер по компьютерной технике
инженер по аппаратной части ПК
специалист по ИТ-оборудованию

Таблица 2.

**Примеры названия для сущностей навык/компетенция и знания для профессии «инженер по компьютерному оборудованию»**

Приоритетное название	Альтернативные названия	Тип
собирать аппаратные компоненты	сборка компьютерной техники, установка оборудования, сборка компьютерных комплектующих, сборка компонентов компьютера	навык / компетенция
установка программного обеспечения	установка компьютерного программного обеспечения, загрузка программного обеспечения, установка компьютерного программного обеспечения, загрузка программного обеспечения, ...	навык / компетенция
создавать технические планы	создавать планы относительно технических деталей, создавать промышленные планы, создавать технические чертежи	навык / компетенция
принципы электричества	электрический ток, напряжение, физика электричества, наука об электричестве, теория электричества, сопротивление, напряжение	знания
аппаратные компоненты	аппаратные компоненты системы, типы аппаратных компонентов, компоненты оборудования, компоненты для аппаратных систем, части для аппаратных систем, компоненты аппаратных систем, аппаратные части системы, типология аппаратных компонентов	знания

выков/компетенций и знаний, и  $R: O \cdot S \rightarrow B$  — отношение, которое связывает профессию  $o$  с навыком  $s$ , а именно  $r(o, s) = 1$ , если навык  $s$  связан с профессией  $o$  в ESCO, и 0 в противном случае.

Стоит отметить, что одна профессия может быть связана с несколькими сущностями навыков/компетенций и знаний, а одна сущность навыков/компетенций или знаний может относиться к нескольким профессиям (рис. 2).

**Определение онлайн-вакансии.** Онлайн-вакансия  $j$  представлена в виде кортежа  $V = (i, c, p, t)$ , где  $i \in N$  — уникальный идентификатор,  $c \in C$  — уникальный идентификатор отрасли,  $p \in P$  — уникальный идентификатор профессии вакансии,  $t \in T$  — текст требований из вакансии.

**2.2. Вычисление важности навыков для профессии**

Так как один и тот же навык в структуре ESCO может быть связан с несколькими профессиями, необходим инструмент, позволяющий оценить важность навыка для конкретной профессии.

Оценить важность навыков для каждой профессии можно с помощью инструмента RCA, первоначально использовавшегося в контексте иссле-

ований в США [23], где авторы использовали классификацию навыков O\*NET (американский аналог ESCO), чтобы учесть важность каждого навыка для каждой профессии. Важность (частота) навыков для профессий  $o_i \in O$  и навыков  $s_l \in S$  определяется по формуле (1), где  $I$  обозначает функцию индикатор. Функция  $rca$  вычисляется по формуле (2), где  $sf$  — частота навыка  $s_l$  для профессии  $o_i$ .

Чтобы получить более понятную меру, мы вычисляем нормализованный  $rca$  по формуле (3), нормализуя  $rca$  по отношению к максимальному значению, полученному для учитываемой профессии, так что наиболее востребованный навык для каждой профессии имеет нормализованный  $rca$ , равный 1.

$$sf(o_i, s_l) = \frac{\sum_{k=1}^m I(o_k = o_i) \cdot I(s_l = s_l)}{\sum_{k=1}^m I(o_k = o_i)}, \tag{1}$$

$$rca(o_i, s_l) = \frac{sf(o_i, s_l) / \sum_{j=1}^p sf(o_i, s_j)}{\sum_{k=1}^m sf(o_k, s_l) / \sum_{k=1}^m \sum_{j=1}^p sf(o_k, s_j)}, \tag{2}$$

$$rca_N(o_i, s_l) = \frac{rca(o_i, s_l)}{\max_j rca(o_i, s_l)}. \tag{3}$$

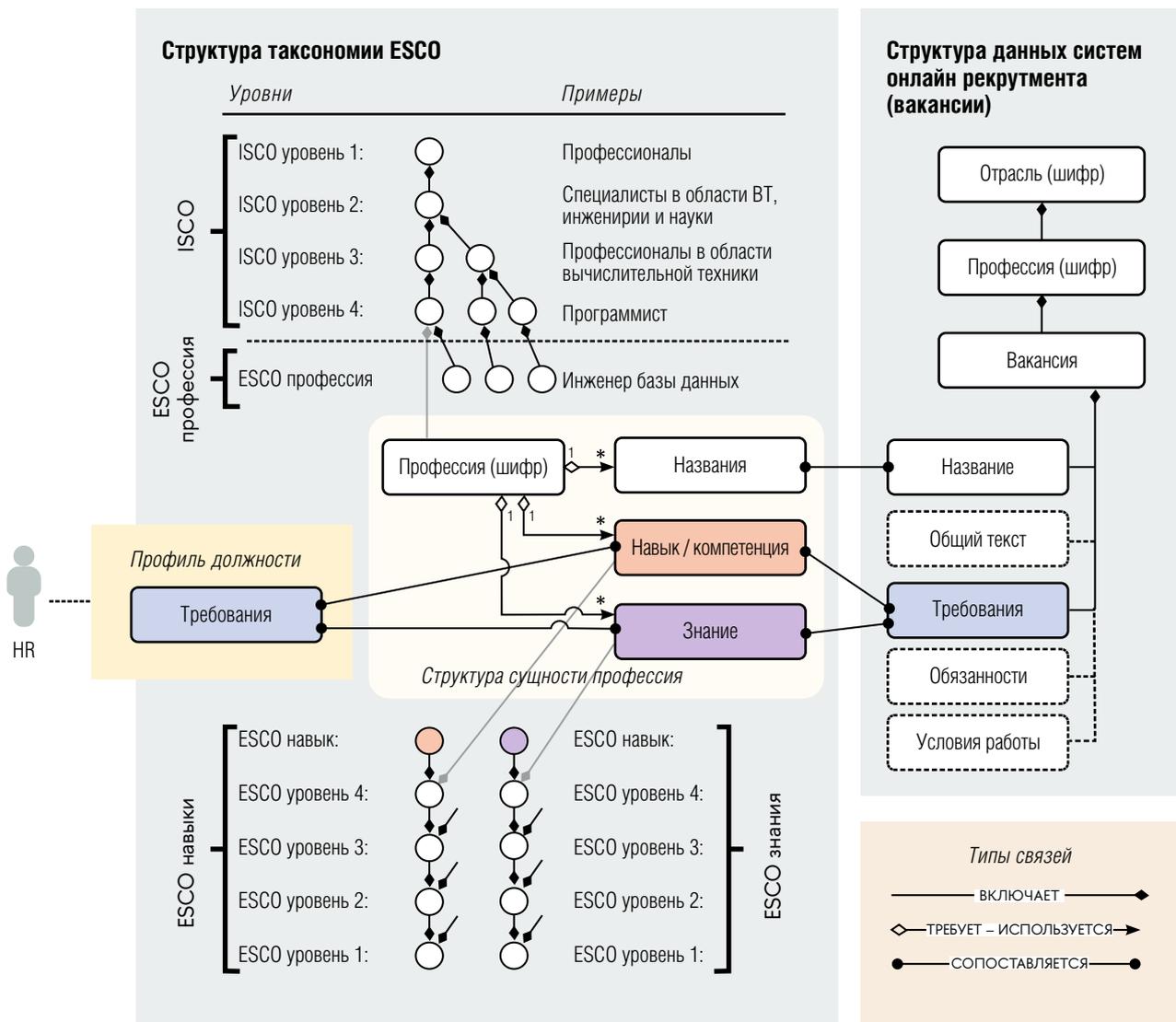


Рис. 2. Основные сущности и отношения в обобщенной графовой модели требований рынка труда на основе классификации ESCO.

### 2.3. Формирование рекомендаций на основе семантического сопоставления первоначального списка требований из профиля должности с графовой моделью рынка труда

Идея метода формирования рекомендаций по актуализации списка требований при составлении профиля должности исходит из семантического сопоставления отдельных сущностей из первоначального списка требований и сущностей из графовой модели рынка труда.

В предлагаемом методе можно выделить следующих этапы:

1. Создание графовой модели рынка труда на основе классификации навыков ESCO.

2. Расширение графовой модели за счет информации из текстов онлайн-вакансий реального рынка труда.
3. Сопоставление исходного списка требований с сущностями графовой модели рынка труда.
4. Ранжирование результатов сопоставления на основе метрики RCA.
5. Формирование рекомендаций для внесения в исходный список требований.

Шаги 3–5 могут повторяться несколько раз, что позволит с каждой новой итерацией формировать более точный и актуальный список требований.

Модель рекомендательной системы представлена на рисунке 3.

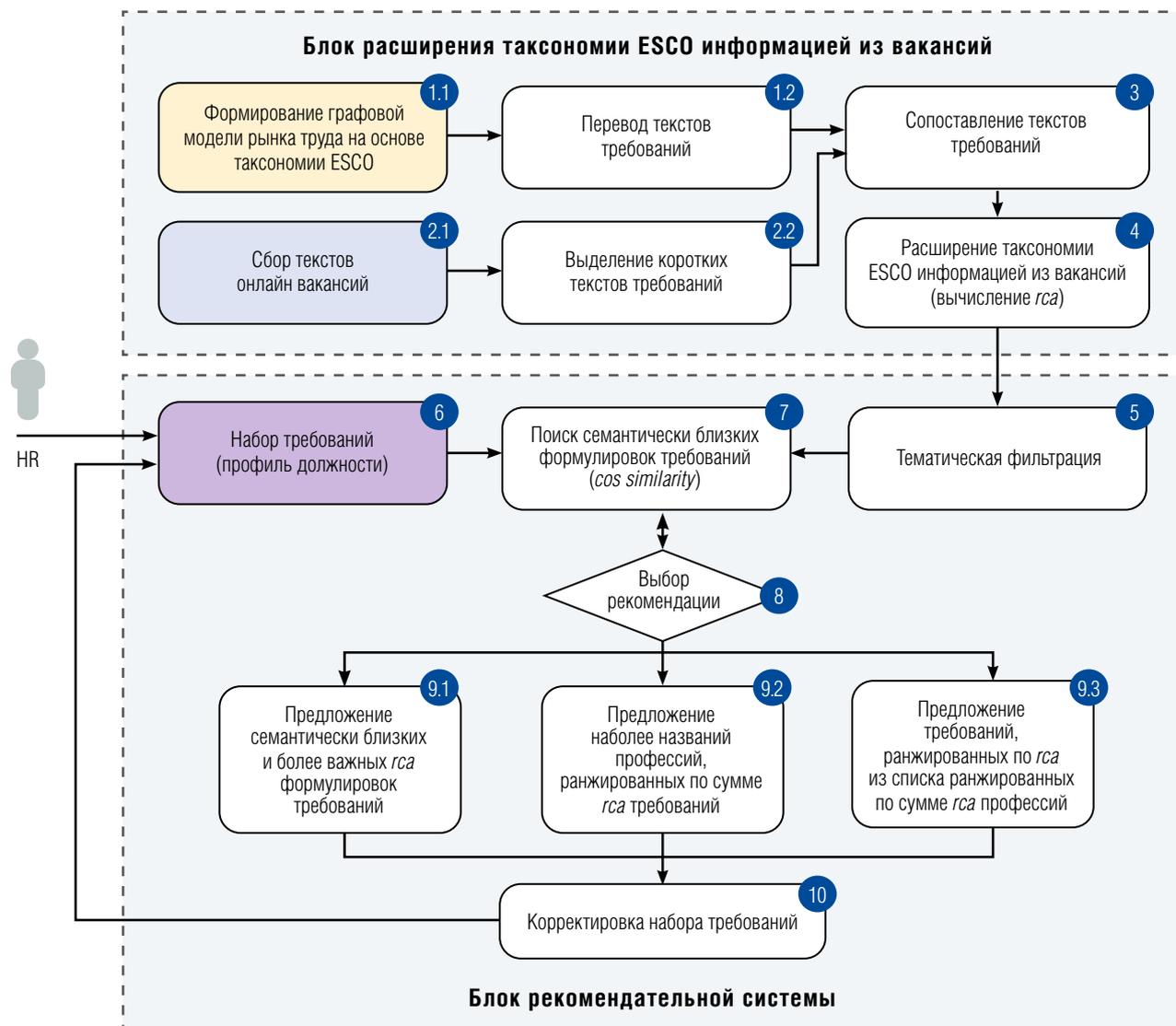


Рис. 3. Модель рекомендательной системы по актуализации списка требований при составлении профиля должности.

### 3. Концептуальный алгоритм функционирования системы

Концептуальный алгоритм – это абстрактное описание процесса решения задачи или выполнения определенного действия без указания конкретных инструкций или языка программирования. Это понятие используется в информатике, математике и других научных областях, где важен процесс решения задачи, а не конкретный код или программный язык. В научных публикациях данный термин может использоваться для обсуждения общих подходов к решению задач, без привязки к конкретным технологиям или реализациям.

Результатом работы стала разработка концептуального алгоритма работы рекомендательной системы:

1.1. Формирование графовой модели рынка труда на основе европейской таксономии навыков ESCO (см. описание графовой модели).

1.2. Адаптация графовой модели навыков рынка труда для русского языка. С помощью сервисов автоматического перевода переводятся все формулировки европейской таксономии навыков, при этом структура связей между сущностями таксономии сохраняется.

2.1. Сбор онлайн-вакансий из российских систем онлайн-рекрутмента. Многие системы онлайн-ре-

крутмента поддерживают открытый API для получения данных о вакансиях (например, [api.hh.ru](http://api.hh.ru), [api.superjob.ru](http://api.superjob.ru) и др.).

2.2. Выделение коротких текстов требований из текстов вакансий. Есть несколько возможных вариантов выделения коротких текстов требований из текстов вакансий:

- ◆ поиск прямых совпадения формулировок сущностей из графовой модели рынка труда с текстами из онлайн-вакансий;
- ◆ разработка правил и поиск совпадений на основе правил (например, использовать `yargy`-парсер из библиотеки `natasha` для `python` (<https://natasha.github.io/>);
- ◆ и третий вариант, подготовка обучающего датасета и обучение модели под задачу NER. Например, используя нейросетевые модели для анализа текстов на естественном языке из библиотеки `deeppavlov` (<https://deeppavlov.ai/>) разработанной инновационной ИИ-компанией – `iPavlov`, спин-офф МФТИ. Продолжение успешного проекта «Нейроинтеллект `iPavlov`», реализованного в рамках НТИ, при индустриальной поддержке Сбербанка.

3. Сопоставление текстов требований из онлайн-вакансий с сущностями графовой модели требований рынка труда. На данном шаге предполагается использовать простое сопоставление и поиск совпадений между нормализованными (приведенных к нормальной форме) текстами требований, извлеченных из текстов онлайн-вакансий и сущностями графовой модели.

Одним из возможных улучшений данного шага можно предложить использование русскоязычной версии тезауруса `ruwordnet` [24], который содержит отношения гипонимов, гиперонимов, а также словарь синонимов для русского языка. Использование данной информации позволит расширить список вариации текстов требований при сопоставлении.

4. Для всех формулировок графовой модели считается параметр *rsa*, который фактически обозначает важность навыка для конкретной профессии.

Стоит отметить, что расширение таксономии ESCO может происходить не просто статистическими данными, но и может представлять собой более сложный процесс, например, поиск новых формулировок навыков/компетенций или знаний и их интегрирование в существующую графовую модель [21,22].

5. Тематическая фильтрация. На данном шаге, используя инструменты тематического моделирования, пользователю могут быть предложены термины и понятия, автоматически сгруппированные по темам. Пользователь может выбрать список слов или понятий, которые в обязательном порядке должны содержаться или отсутствовать в текстах итоговой выдачи. Возможность применения такого тематического фильтра была продемонстрирована автором в статье [25].

6. Пользователем системы (например, HR-специалистом) формируется первоначальный набор требований и подается на вход системы.

7. Сопоставление текстов требований из запроса пользователя и сущностей из графовой модели рынка труда. В данном случае предполагается использовать более интеллектуальный процесс сопоставления, который можно разбить на два этапа: с помощью современных нейросетевых моделей, построенных на архитектуре трансформеров (`RuBERT`, `Robert` и другие) для текстов требований пользователя и для сущностей графовой модели получаются векторные представления. Затем, используя конусную меру близости, попарно сопоставляются векторные представления. Далее отсекаются все тексты требований, которые лежат за пределами некоторого допустимого расстояния (определяется экспериментально), и которые являются наиболее семантически близкими к тексту исходного требования. Так определяются семантически близкие тексты требований из графовой модели для всех текстов из исходного списка требований пользователя

Алгоритм выбора наиболее эффективной нейросетевой модели, которая бы позволила генерировать наилучшие (с точки зрения компактности) векторные представления для семантически близких текстов требований был рассмотрен автором в статье [26].

8. Пользователь выбирает какой тип рекомендации он хотел бы сформировать для исходного списка требований.

9.1. Семантически близкие тексты из графовой модели, отобранные на шаге 7 ранжируются по параметру *rsa*. Отранжированные формулировки демонстрируются пользователю, с указанием параметра *rsa* и профессии. Фактически на этом этапе пользователь получает возможность выбрать наиболее важные и семантически близкие требования, и принимает решение о добавлении предлагаемых требований к своему первоначальному списку.

9.2. Семантически близкие тексты из графовой модели, отобранные на шаге 6, ранжируются по параметру  $rca$ . Для совокупности требований исходного списка отбираются формулировки с наибольшим  $rca$ . По всем требованиям и профессиям считается сумма. Профессии ранжируются по сумме  $rca$  для исходного списка требований (рис. 4). Из графовой модели отбираются  $n$  названия профессии на основе наибольшей суммы  $rca$  для исходного списка требований пользователя и демонстрируются пользователю.

9.3. Также как в 9.2 вычисляется сумма  $rca$ . Из графовой модели отбираются  $n$  профессии на основе наибольшей суммы  $rca$  для исходного списка требований пользователя (рис. 4). Из каждой отобранной профессии отбираются  $m$  наиболее важных по  $rca$  текстов навыков/компетенций или знаний и демонстрируются пользователю.

Шаг 10. После изучения предложенных рекомендаций пользователь системы может выбрать варианты, которые будут добавлены в первоначальный список требований.

Шаги с 5 по 10 могут повторяться неоднократно, что позволит итеративно уточнять и улучшать первоначальный список требований, заданных пользователем.

Примеры возможных вариантов модификации набора требований профиля должности на основе сформированных рекомендаций:

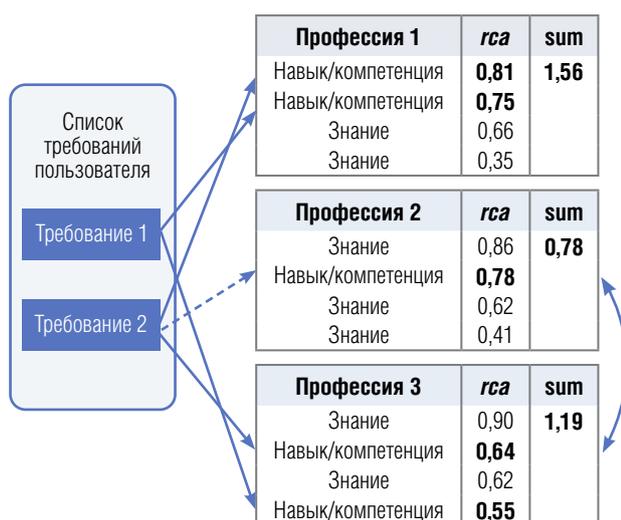


Рис. 4. Ранжирование профессии по сумме  $rca$  для исходного списка требований.

- ♦ Предложить новую альтернативную формулировку для существующего требования на основе более высокого  $rca$ .
- ♦ Предложить включить новое требование в профиль должности на основе высокого  $rca$  с учетом взаимосвязи с существующими требованиями.
- ♦ Рекомендовать исключить требование из профиля должности на основе низкого  $rca$  с учетом взаимосвязи с существующими требованиями.
- ♦ Разбить требования на категории по принадлежности к профессиям
- ♦ Предложить название должности исходя из списка требований профиля должности.
- ♦ Ранжирование требований по их востребованности на основе показателя  $rca$ .

Список рекомендаций может быть скорректирован в сторону расширения путем добавления новых функциональных блоков в систему.

Вычислительная сложность всей системы оценивается как невысокая. Наиболее трудоемкие этапы, такие как извлечение сущностей знаний, навыков и компетенций из текстов вакансий, а также их векторного представления происходят один раз и могут происходить в фоновом режиме. Тематическая фильтрация и подсчет  $rca$  являются относительно простыми вычислительными операциями. Наиболее сложной, с точки зрения вычислительной сложности, представляет собой операция ранжирования большого количества требований относительно друг друга на основе косинусной меры близости. Для выполнения высокопроизводительного поиска похожих векторов требований планируется использовать библиотеку FAISS (Facebook AI Similarity Search) [26, 27]. Данная библиотека предоставляет собой набор алгоритмов для индексации больших наборов векторов и быстрого поиска ближайших соседей в этих наборах. Библиотека была разработана Facebook AI Research и распространяется на условиях лицензии Apache 2.0.

### Заключение

Разработка рекомендательной системы для формирования актуального списка требований профиля должности на основе анализа реальных требований рынка труда является важным шагом в развитии HR-технологий и позволит: значительно сократить трудозатраты HR специалистов, более точно и системно определять требования к кандидатам на различные должности; определить потенциальные возможности для переподготов-

ки работников, лучше понимать, как изменения в требованиях рынка труда влияют на компании и их персонал, формировать более гибкие и адаптивные стратегии привлечения и управления персоналом.

Предложенная концептуальная модель рекомендательной системы включает в себя:

- ◆ Графовую модель требований рынка труда на основе таксономии ESCO, адаптированной для русского языка;
- ◆ Интеллектуальный метод формирования рекомендаций для составления списка требований в процессе подбора персонала на основе нейросетевых моделей языка с использованием таксономии навыков ESCO и корпуса онлайн-вакансий российского рынка труда. В рамках метода предлагается использовать нейросетевые модели языка, построенные на архитектуре трансформеров (модели семейства BERT) для оценки семантической близости сущностей первоначального списка требований с графовой моделью рынка труда;
- ◆ Модель и концептуальный алгоритм работы автоматизированной рекомендательной системы для формирования рекомендации.

В статье также приводятся возможные варианты рекомендаций по актуализации списка требований профиля должности на основе анализа потребностей реального рынка труда.

Для улучшения данной модели автором дополнительно планируется: разработка метода извлечения отдельных коротких текстов знаний и навыков из текстов требований онлайн-вакансий реального рынка труда; разработка системы автоматического расширения графовой модели текстами знаний и навыков; интеграция расширенной графовой модели требований рынка труда через международную систему классификации занятий (ISCO, International Classification System Of Occupations) с общероссийскими классификаторами (ОКЗ – общероссийский классификатор занятий, ОКВЕД – общероссийский классификатор видов экономической деятельности), и профессиональными стандартами РФ. Кроме того, отдельной и важной задачей является разработка метода оценки качества и исследование эффективности использования рекомендательной системы на основе предложенной модели в различных секторах экономики и на различных рынках труда. ■

### Литература

1. Boselli R., Cesarini M., Mercorio F. Using machine learning for labour market intelligence // *Machine Learning and Knowledge Discovery in Databases: European Conference*. 2017. P. 330–342. [https://doi.org/10.1007/978-3-319-71273-4\\_27](https://doi.org/10.1007/978-3-319-71273-4_27)
2. Colombo E., Mercorio F., Mezzanzanica M. AI meets labor market: Exploring the link between automation and skills // *Information Economics and Policy*. 2019. Vol. 47. P. 27–37. <https://doi.org/10.1016/j.infoecopol.2019.05.003>
3. Giabelli A., Malandri L., Mercorio F. GraphLMI: A data driven system for exploring labor market information through graph databases // *Multimedia Tools and Applications*. 2022. Vol. 81. P. 3061–3090. <https://doi.org/10.1007/s11042-020-09115-x>
4. Mezzanzanica M., Boselli R., Cesarini M., Mercorio F. A model-based approach for developing data cleansing solutions // *Journal of Data and Information Quality*. 2015. Vol. 5. No. 4. P. 1–28. <https://doi.org/10.1145/2641575>
5. Xu T., Zhu H., Zhu C., Li P., Xiong H. Measuring the popularity of job skills in recruitment market: A multi-criteria approach // *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018. Vol. 32. No. 1. <https://doi.org/10.1609/aaai.v32i1.11847>
6. Zhang D., Liu J., Zhu H., Liu Y., Wang L., Wang P., Xiong H. Job2Vec: Job title benchmarking with collective multi-view representation learning // *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2019. P. 2763–2771. <https://doi.org/10.1145/3357384.3357825>
7. Важность LMI / Комиссия Великобритании по трудоустройству и профессиональным навыкам. 2015. [Электронный ресурс] <https://www.gov.uk/government/publications/the-importance-of-labour-market-intelligence> (дата обращения 30.01.2023).
8. Mezzanzanica M., Mercorio F. Big data enables labor market // *Encyclopedia of Big Data Technologies*. 2019. P. 226–236. [https://doi.org/10.1007/978-3-319-63962-8\\_276-1](https://doi.org/10.1007/978-3-319-63962-8_276-1)
9. Eggertsson T. *Economic behavior and institutions: Principles of neoinstitutional economics*. Cambridge University Press, 1990. <https://doi.org/10.1017/CBO9780511609404>
10. Qin C., Zhu H., Xu T., Zhu C., Jiang L., Chen E., Xiong H. Enhancing person-job fit for talent recruitment: An ability-aware neural network approach // *The 41st international ACM SIGIR conference on research & development in information retrieval*. 2018. P. 25–34. <https://doi.org/10.1145/3209978.3210025>
11. Zhu C., Zhu H., Xiong H., Ma C., Xie F., Ding P., Li P. Person-job fit: Adapting the right talent for the right job with joint representation learning // *ACM Transactions on Management Information Systems*. 2018. Vol. 9. No. 3. P. 1–17. <https://doi.org/10.1145/3234465>

12. Xu T., Zhu H., Zhu C., Li P., Xiong H. Measuring the popularity of job skills in recruitment market: A multi-criteria approach // Proceedings of the AAAI conference on artificial intelligence. 2018. Vol. 32. No. 1. <https://doi.org/10.1609/aaai.v32i1.11847>
13. Vinel M., Ryazanov I., Botov D., Nikolaev I. Experimental comparison of unsupervised approaches in the task of separating specializations within professions in job vacancies // Artificial Intelligence and Natural Language. AINL 2019. Communications in Computer and Information Science. Ustalov, D., Filchenkov, A., Pivovarova, L. (eds). Vol. 1119. Springer, Cham, 2019. [https://doi.org/10.1007/978-3-030-34518-1\\_7](https://doi.org/10.1007/978-3-030-34518-1_7)
14. Nikolaev I., Ryazanov I., Botov D. The comparison of distributive semantics models applied to the task of short job requirements clustering for the Russian labor market // 8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020). Atlantis Press, 2020. P. 295–301. <https://doi.org/10.2991/aisr.k.201029.056>
15. Colombo E., Mercurio F., Mezzanzanica M. Applying machine learning tools on web vacancies for labour market and skill analysis // Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence. Technology Policy Institute Conference on the Economics and Policy Implications of Artificial Intelligence. [Электронный ресурс] [https://techpolicyinstitute.org/wp-content/uploads/2021/03/Colombo\\_paper.pdf](https://techpolicyinstitute.org/wp-content/uploads/2021/03/Colombo_paper.pdf) (дата обращения 30.05.2023).
16. O’Kane L., Narasimhan R., Nania J., Taska B. Digitalization in the German labor market: Analyzing demand for digital skills in job vacancies // Gütersloh: Bertelsmann Stiftung, 2020.
17. Real-time labour market information on skill requirements: Setting up the EU system for online vacancy analysis / Cedefop. [Электронный ресурс] <https://www.cedefop.europa.eu/it/about-cedefop/public-procurement/real-time-labour-market-information-skill-requirements-setting-eu> (дата обращения 30.01.2023).
18. Boselli R., Cesarini M., Mercurio F., Mezzanzanica M. Classifying online job advertisements through machine learning // Future Generation Computer Systems. 2018. Vol. 86. P. 319–328. <https://doi.org/10.1016/j.future.2018.03.035>
19. Boselli R., Cesarini M., Marrara S., Mercurio F., Mezzanzanica M., Pasi G., Viviani M. WoLMIS: a labor market intelligence system for classifying web job vacancies // Journal of intelligent information systems. 2018. Vol. 51. P. 477–502. <https://doi.org/10.1007/s10844-017-0488-x>
20. Chiarello F., Fantoni G., Hogarth T., Giordano V., Baltina L., Spada I. Towards ESCO 4.0 – Is the European classification of skills in line with Industry 4.0? A text mining approach // Technological Forecasting and Social Change. 2021. Vol. 173. Article 121177. <https://doi.org/10.1016/j.techfore.2021.121177>
21. Giabelli A., Malandri L., Mercurio F., Mezzanzanica M., Seveso A. NEO: A tool for taxonomy enrichment with new emerging occupations // The Semantic Web – ISWC 2020. Lecture Notes in Computer Science. 2020. Vol. 12507. P. 568–584. [https://doi.org/10.1007/978-3-030-62466-8\\_35](https://doi.org/10.1007/978-3-030-62466-8_35)
22. Malandri L., Mercurio F., Mezzanzanica M., Nobani N. Meet: A method for embeddings evaluation for taxonomic data // 2020 International Conference on Data Mining Workshops (ICDMW). P. 31–38. <https://doi.org/10.1109/ICDMW51313.2020.00014>
23. Alabdulkareem A., Frank M.R., Sun L., AlShebli B., Hidalgo C., Rahwan I. Unpacking the polarization of workplace skills // Science advances. 2018. Vol. 4. No. 7. Article eaao6030. <https://doi.org/10.1126/sciadv.aao6030>
24. Тезаурус русского языка RuWordNet. [Электронный ресурс] <https://ruwordnet.ru/ru> (дата обращения 30.05.2023).
25. Nikolaev I., Botov D., Dmitrin Y., Klenin J., Melnikov A. Use of topic modelling for improvement of quality in the task of semantic search of educational courses // 21st International Workshop on Computer Science and Information Technologies (CSIT 2019). Atlantis Press, 2019. P. 104–111. <https://doi.org/10.2991/csit-19.2019.18>
26. Николаев И.Е., Мельников А.В. Сравнение нейросетевых моделей на архитектуре трансформеров в контексте задачи оценки компактности векторных представлений семантически близких текстов требований европейской классификации навыков ESCO // Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника. 2022. № 3. С. 19–29. <https://doi.org/10.14529/ctcr220302>.
27. Faiss. A library for efficient similarity search and clustering of dense vectors. [Электронный ресурс] <https://faiss.ai/> (дата обращения 30.05.2023).
28. Johnson J., Douze M., Jégou H. Billion-scale similarity search with gpus // IEEE Transactions on Big Data. 2019. Vol. 7. No. 3. P. 535–547.

## Об авторе

### Николаев Иван Евгеньевич

старший преподаватель, кафедра информационных технологий и экономической информатики, Челябинский государственный университет, 454001, г. Челябинск, ул. Братьев Кашириных, 129;

E-mail: [ivan\\_nikolaev@csu.ru](mailto:ivan_nikolaev@csu.ru)

ORCID: 0000-0002-9686-2435

# An intelligent method for generating a list of job profile requirements based on neural network language models using ESCO taxonomy and online job corpus

**Ivan E. Nikolaev**

E-mail: [ivan\\_nikolaev@csu.ru](mailto:ivan_nikolaev@csu.ru)

Chelyabinsk State University

Address: 129, Kashirin Brothers Str., Chelyabinsk 454001, Russia

## Abstract

Online recruitment systems have accumulated a huge amount of data on the real labor market in recent years. Of particular interest to the study are the data on the real requirements of the labor market contained in the texts of online vacancies, as well as the process of extracting and structuring them for further analysis and use. The stage of compiling an up-to-date list of requirements for a position profile in the recruitment process is very time-consuming and requires a large amount of effort from an HR specialist related to monitoring changes in entire industries and professions, as well as analyzing relevance of existing requirements on the market. In this article, the author proposes a conceptual model of a recommendation system that allows one to reduce the burden on an HR specialist at the stage of forming an up-to-date list of requirements for a position profile in the recruitment process. The model is based on a combination of the following components: a graph model of labor market requirements based on the ESCO taxonomy adapted for the Russian language; and an intelligent method of forming recommendations for compiling an up-to-date list of requirements in the recruitment process based on neural network models of the language on the architecture of transformers, ESCO skills taxonomy and corpus online vacancies of the Russian labor market. The article also provides a conceptual algorithm for the work of the recommendation system and possible options for recommendations on updating the list of requirements of the position profile in the recruitment process based on an analysis of the needs of the real labor market.

**Keywords:** labor market analysis, labor market requirements, human resources, job profile, data mining, natural language processing, neural network language models

**Citation:** Nikolaev I.E. (2023) An intelligent method for generating a list of job profile requirements based on neural network language models using ESCO taxonomy and online job corpus. *Business Informatics*, vol. 17, no. 2, pp. 71–84. DOI: [10.17323/2587-814X.2023.2.71.84](https://doi.org/10.17323/2587-814X.2023.2.71.84)

## References

1. Boselli R., Cesarini M., Mercorio F. (2017) Using machine learning for labour market intelligence. *Machine Learning and Knowledge Discovery in Databases: European Conference*, pp. 330–342. [https://doi.org/10.1007/978-3-319-71273-4\\_27](https://doi.org/10.1007/978-3-319-71273-4_27)
2. Colombo E., Mercorio F., Mezzanzanica M. (2019) AI meets labor market: Exploring the link between automation and skills. *Information Economics and Policy*, vol. 47, pp. 27–37. <https://doi.org/10.1016/j.infoecopol.2019.05.003>
3. Giabelli A., Malandri L., Mercorio F. (2022) GraphLMI: A data driven system for exploring labor market information through graph databases. *Multimedia Tools and Applications*, vol. 81, pp. 3061–3090. <https://doi.org/10.1007/s11042-020-09115-x>
4. Mezzanzanica M., Boselli R., Cesarini M., Mercorio F. (2015) A model-based approach for developing data cleansing solutions. *Journal of Data and Information Quality*, vol. 5, no. 4, pp. 1–28. <https://doi.org/10.1145/2641575>
5. Xu T., Zhu H., Zhu C., Li P., Xiong H. (2018) Measuring the popularity of job skills in recruitment market: A multi-criteria approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1. <https://doi.org/10.1609/aaai.v32i1.11847>
6. Zhang D., Liu J., Zhu H., Liu Y., Wang L., Wang P., Xiong H. (2019) Job2Vec: Job title benchmarking with collective multi-view representation learning. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2763–2771. <https://doi.org/10.1145/3357384.3357825>
7. *The importance of LMI* (2015) UK Commission for Employment and Skills. Available at: <https://www.gov.uk/government/publications/the-importance-of-labour-market-intelligence> (accessed 30.01.2023).
8. Mezzanzanica M., Mercorio F. (2019) Big data enables labor market intelligence. *Encyclopedia of Big Data Technologies*, pp. 226–236. [https://doi.org/10.1007/978-3-319-63962-8\\_276-1](https://doi.org/10.1007/978-3-319-63962-8_276-1)
9. Eggertsson T. (1990) *Economic behavior and institutions: Principles of neoinstitutional economics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511609404>
10. Qin C., Zhu H., Xu T., Zhu C., Jiang L., Chen E., Xiong H. (2018) Enhancing person-job fit for talent recruitment: An ability-aware neural network approach. *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*, pp. 25–34. <https://doi.org/10.1145/3209978.3210025>
11. Zhu C., Zhu H., Xiong H., Ma C., Xie F., Ding P., Li P. (2018) Person-job fit: Adapting the right talent for the right job with joint representation learning. *ACM Transactions on Management Information Systems*, vol. 9, no. 3, pp. 1–17. <https://doi.org/10.1145/3234465>
12. Xu T., Zhu H., Zhu C., Li P., Xiong H. (2018) Measuring the popularity of job skills in recruitment market: A multi-criteria approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1. <https://doi.org/10.1609/aaai.v32i1.11847>
13. Vinel M., Ryazanov I., Botov D., Nikolaev I. (2019) Experimental comparison of unsupervised approaches in the task of separating specializations within professions in job vacancies. *Artificial Intelligence and Natural Language. AINL 2019. Communications in Computer and Information Science* (eds. D. Ustalov, A. Filchenkov, L. Pivovarova), vol. 1119. Springer, Cham. [https://doi.org/10.1007/978-3-030-34518-1\\_7](https://doi.org/10.1007/978-3-030-34518-1_7)
14. Nikolaev I., Ryazanov I., Botov D. (2020) The comparison of distributive semantics models applied to the task of short job requirements clustering for the Russian labor market. *Proceedings of the 8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020)*. Atlantis Press, pp. 295–301. <https://doi.org/10.2991/aisr.k.201029.056>
15. Colombo E., Mercorio F., Mezzanzanica M. (2018) Applying machine learning tools on web vacancies for labour market and skill analysis. *Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence. Technology Policy Institute Conference on the Economics and Policy Implications of Artificial Intelligence*. Available at: [https://techpolicyinstitute.org/wp-content/uploads/2021/03/Colombo\\_paper.pdf](https://techpolicyinstitute.org/wp-content/uploads/2021/03/Colombo_paper.pdf) (accessed 30 May 2023).
16. O’Kane L., Narasimhan R., Nania J., Taska B. (2020) *Digitalization in the German labor market: Analyzing demand for digital skills in job vacancies*. Gütersloh: Bertelsmann Stiftung.
17. *Real-time labour market information on skill requirements: Setting up the EU system for online vacancy analysis* (2016) Cedefop. Available at: <https://www.cedefop.europa.eu/it/about-cedefop/public-procurement/real-time-labour-market-information-skill-requirements-setting-eu> (accessed 30 January 2023).
18. Boselli R., Cesarini M., Mercorio F., Mezzanzanica M. (2018) Classifying online job advertisements through machine learning. *Future Generation Computer Systems*, vol. 86, pp. 319–328. <https://doi.org/10.1016/j.future.2018.03.035>
19. Boselli R., Cesarini M., Marrara S., Mercorio F., Mezzanzanica M., Pasi G., Viviani M. (2018) WoLMIS: A labor market intelligence system for classifying web job vacancies. *Journal of Intelligent Information Systems*, vol. 51, pp. 477–502. <https://doi.org/10.1007/s10844-017-0488-x>
20. Chiarello F., Fantoni G., Hogarth T., Giordano V., Baltina L., Spada I. (2021) Towards ESCO 4.0 – Is the European classification of skills in line with Industry 4.0? A text mining approach. *Technological Forecasting and Social Change*, vol. 173, article 121177. <https://doi.org/10.1016/j.techfore.2021.121177>
21. Giabelli A., Malandri L., Mercorio F., Mezzanzanica M., Seveso A. (2020) NEO: A tool for taxonomy enrichment with new emerging occupations. *The Semantic Web – ISWC 2020. Lecture Notes in Computer Science*, vol. 12507, pp. 568–584. [https://doi.org/10.1007/978-3-030-62466-8\\_35](https://doi.org/10.1007/978-3-030-62466-8_35)

22. Malandri L., Mercurio F., Mezzanica M., Nobani N. (2020) Meet: A method for embeddings evaluation for taxonomic data. *Proceedings of the 2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 31–38. Sorrento, Italy. <https://doi.org/10.1109/ICDMW51313.2020.00014>
23. Alabdulkareem A., Frank M.R., Sun L., AlShebli B., Hidalgo C., Rahwan I. (2018) Unpacking the polarization of workplace skills. *Science Advances*, vol. 4, no. 7, eaao6030. <https://doi.org/10.1126/sciadv.aao6030>
24. *The Russian language Thesaurus RuWordNet*. Available at: <https://ruwordnet.ru/ru> (accessed 30 May 2023).
25. Nikolaev I., Botov D., Dmitrin Y., Klenin J., Melnikov A. (2019) Use of topic modelling for improvement of quality in the task of semantic search of educational courses. *Proceedings of the 21st International Workshop on Computer Science and Information Technologies (CSIT 2019)*, pp. 104–111. Atlantis Press. <https://doi.org/10.2991/csit-19.2019.18>
26. Nikolaev I., Melnikov A. (2022) Comparison of neural network models based on transformer architecture in the context of the task of evaluating the compactness of vector representations of semantically similar texts of the requirements of the European ESCO skills classification. *Bulletin of the South Ural State University. Series: Computer Technology, Control, Radio Electronics*, vol. 22, no. 3, pp. 19–29.
27. Faiss. *A library for efficient similarity search and clustering of dense vectors*. Available at: <https://faiss.ai/> (accessed 30 May 2023).
28. Johnson J., Douze M., Jégou H. (2019) Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547.

### About the author

#### Ivan E. Nikolaev

Senior Lecturer, Department of Information Technology, Institute of Information Technologies, Chelyabinsk State University, 129, Kashirin Brothers Str., Chelyabinsk 454001, Russia;

E-mail: [ivan\\_nikolaev@csu.ru](mailto:ivan_nikolaev@csu.ru)

ORCID: 0000-0002-9686-2435