

УДК 004.891.2

КЛАСТЕРИЗАЦИЯ ВАКАНСИЙ ПО КЛЮЧЕВЫМ НАВЫКАМ КАНДИДАТОВ В СФЕРЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Зайчикова А.И. (Университет ИТМО)

Научный руководитель – доцент, Графеева Н.Г.

(Университет ИТМО)

Введение. Полный реестр профессиональных стандартов востребован во многих организациях, таких как государственные учреждения, учреждения социального обслуживания, высшие учебные заведения, крупные промышленные предприятия, для эффективного и своевременного получения информации о профессиях и профессиональных навыках. Традиционным решением в различных странах мира является стандартизация и унификация профессиональной структуры общества с помощью стандартов классификации профессий [1]. Стандарты и должностные инструкции, такие как содержание профессиональной деятельности, требования к навыкам и квалификационные требования, вместе образуют систему профессионального портрета [2]. Репрезентативными стандартами являются ONET в Соединенных Штатах [3], ESCO в Европейском союзе [4], NOC в Канаде [5] и SOCGD в Китае [6].

В России также есть свой реестр профессиональных стандартов [7], который охватывает 1521 профессию. Эти стандарты в основном способны всесторонне и объективно описать профессии в России. Однако, по сравнению с Европой и Соединенными Штатами, российская система профессионального портрета имеет некоторые недостатки. Одним из них является отсутствие профессиональных навыков или квалификационных требований по некоторым профессиям, в частности очень скудно описаны профессии в сфере информационных технологий.

Другим существенным недостатком является то, что актуализация данных серьезно запаздывает, больше половины профессиональных портретов не обновлялись уже более 10 лет, что для многих профессий является не допустимым, поскольку появление новых технологий и различные научные открытия, как следствие, ведут за собой изменения требований, предъявляемых к кандидатам. Разумеется, главной причиной отсутствия актуальных данных является тот факт, что процесс пересмотра стандартов – это ручная работа, которая отнимает значительное количество времени, рабочую силу и много ресурсов. Поэтому возникает необходимость предоставления информации другим путем и в режиме реального времени.

Сегодня онлайн-рекрутмент стал важным каналом как для рекрутеров, распространяющих информацию об имеющихся вакансиях, так и для лиц, находящихся в поиске работы. Содержание вакансий обычно содержит описание информации о работе, требования к специальности, образованию, опыту, профессиональным навыкам и т.д. Извлекая информацию о требованиях из вакансий и объединяя ее в профессиональную информацию, ее можно использовать в качестве эффективного дополнения к традиционному профессиональному профилированию.

Основная часть. Работа состоит из четырех основных этапов.

На первом этапе формируется датасет из json-объектов вакансий с помощью официального API ресурса Head-Hunter [8].

На втором этапе из вакансий выделяются явные и неявные признаки (ключевые навыки кандидатов). Явные признаки извлекаются из стандартных полей json-объектов. Неявные признаки извлекаются из текстов требований по вакансии путем решения задачи распознавания именованных сущностей с помощью модели W2NER – классификатор попарных отношений слов в предложении. Данная модель использует предобученную BERT и двухстороннюю LSTM для получения представления входного предложения в виде тензора. Далее свёрточная сеть преобразует тензор в двумерное представление попарных отношений

слов в предложении. Затем происходит классификация полученных результатов и извлечение предсказанных именованных сущностей из матрицы отношений.

На третьем этапе признакам присваиваются веса с учетом данных из реестра профессиональных стандартов. Создается список, содержащий ключевые понятия и их принадлежности к конкретным вакансиям, иными словами, производится векторизация вакансий.

На четвертом этапе, используя список ключевых понятий, проводится кластерный анализ перечня вакансий следующими методами: иерархическая, k-means, DBSCAN, GMM. Качество использованных моделей определяется на основе двух наборов данных: (1) информация из реестра профессиональных стандартов; (2) экспертные данные от кадровых сотрудников. По результатам кластеризации составляются топы ключевых навыков кандидатов и определяются диапазоны оплаты труда для полученных кластеров (специальностей в сфере информационных технологий).

Выводы. Предлагаемый подход позволит автоматизировать анализ потребности в специалистах, путем анализа перечня вакансий, что позволит сократить трудоёмкость выполнения и повысить качество анализа потребности специалистов.

Список использованных источников:

1. Emmel A., Cosca T. Occupational classification systems: Analyzing the 2010 standard occupational classification (soc) revision //Bureau of Labour Statistics. – 2010.
2. Executive Office of the President, Office of Management and Budget. Standard Occupational Classification Manual. United States; 2018.
3. Peterson N. G. et al. Understanding work using the Occupational Information Network (O*NET): Implications for practice and research //Personnel psychology. – 2001. – Т. 54. – №. 2. – С. 451-492.
4. <https://esco.ec.europa.eu/en>
5. <https://noc.esdc.gc.ca/>
6. National Working Committee on Revision of Occupational Classification. Occupational classification of the People's Republic of China (2015 Edition). China: China labor and social security press. 2015; 7– 8.
7. <https://profstandart.rosmintrud.ru/>
8. <https://dev.hh.ru/>