

УДК 331.543

*Е.С. Голомазова*

*Белорусский государственный экономический университет  
Республика Беларусь, Минск  
elena-golomazova@yandex.ru*

*А.А. Лушневский*

*Гродненский государственный университет им. Я. Купалы  
Республика Беларусь, Гродно  
a.lushnevsky@yandex.by*

## **АВТОМАТИЗАЦИЯ МЕТОДИКИ ВЫЯВЛЕНИЯ ПОТРЕБНОСТЕЙ РЫНКА ТРУДА НА ОСНОВЕ СРЕДСТВ PYTHON**

*Halamazava A.S. Belarus State Economic University, Republic of Belarus, Minsk.*

*Lushneuski A.A. Yanka Kupala State University of Grodno, Republic of Belarus, Grodno.*

**AUTOMATION OF THE METHOD OF LABOR MARKET NEEDS IDENTIFICATION ON THE BASIS OF PYTHON.** Rapid obsolescence of professional skills leads to the need to develop operational methodological support for labor market analysis. In this study, we have developed an automated method of content analysis of job advertisements using Python software. Implementation of the methodology allows to calculate the importance of certain skills for employers.

**KEYWORDS:** content analysis; jobs advertisements; automation of methods of labor market analysis.

*Быстрое устаревание профессиональных навыков приводит к необходимости развития оперативного методологического обеспечения анализа рынка труда. В данном исследовании описана авторская автоматизированная методика контент-анализа объявлений о работе с помощью программных средств Python. Реализация методики позволяет вычислить важность тех или иных навыков для работодателей.*

**КЛЮЧЕВЫЕ СЛОВА:** контент-анализ; объявления о вакансиях; автоматизация методики анализа рынка труда.

1,3 миллиарда из 3,5 работающих людей в мире обладает навыками, не соответствующими потребностям рынка труда [1]. Стремительное технологическое развитие и цифровизация экономики только усугубляет данную проблему. Так, согласно DESI-2019 каждый пятый работающий житель ЕС отметил изменение в течение

года своих функциональных обязанностей в связи с внедрением нового программного обеспечения или оборудования [2]. В данных условиях важным является отслеживание потребностей работодателей в знаниях и навыках с последующим ознакомлением с результатами исследования заинтересованных сторон – учреждений образования, государственных органов, обучающихся либо планирующих обучение лиц.

Одним из методов, позволяющих провести оперативное исследование рынка труда, является количественный контент-анализ объявлений о вакансиях с подсчетом частоты упоминания определенных навыков. С учетом автоматизации данный подход позволяет охватить достаточно большую выборку с точки зрения объема и разнообразия представленных секторов, профессий, географии работодателей с наименьшими затратами ресурсов по сравнению с опросами, анкетированием либо интервьюированием. Ключевым недостатком описываемого метода является его количественная природа и вытекающее из этого допущение, что часто упоминаемые в вакансиях навыки являются ценными для работодателей [3], что не всегда соответствует действительности. Нивелировать данный недостаток помогает расширение объема и улучшение репрезентативности выборки, а также увеличение периода сбора информации [4]. Реализация широкомасштабного исследования с помощью контент-анализа объявлений о работе невозможна без использования программного обеспечения. В связи с этим целью данной работы стала автоматизация методики исследования навыков, представленных в описаниях вакансий работодателей. Методика разрабатывалась для анализа навыков специалистов в области маркетинга информационных технологий [5, 6].

Для выбора оптимального средства автоматизации был проведен теоретический обзор, который показал, что Harper (2012) на основании изучения 70 научных исследований в области контент-анализа объявлений о работе выделил следующие программные средства, которые были использованы для кодирования текстов объявлений с последующим созданием словаря навыков (MS Access, MS Word, SimStat, WordStat, QDA Miner). Их использовали только 4 % исследовательских проектов [4]. Также в немногих из исследований рынка труда специалистов в области маркетинга, менеджмента и финансов [7-11 и др.] указывается программное обес-

печение, с помощью которого происходила обработка собранной информации из объявлений о вакансиях. В исследовании McArthur говорится об использовании программных средств NVivo, позволяющих производить кодирование текстов и расчет частоты упоминания определенной терминологии.

Наиболее сложными этапами для автоматизации в области контент-анализа является токенизация и лемматизация текстов (выделение структурных единиц текста и приведение их к базовым формам, которые в последующем можно сопоставить со словарными формами), а также составление словаря (морфологические единицы, с которыми будут сравниваться выделенные структурные единицы из текста). Автоматизация этих процессов возможна с помощью систем OBIE (ontology-based information extraction) – систем, которые обрабатывают неструктурированный или полуструктурированный тексты на естественном языке через управляемый онтологиями механизм для извлечения определенных типов информации и представления результатов с использованием онтологий (структура терминов, их описание и правила вывода под запросы пользователей) [12]. Примеры реализации исследований объявлений о работе с помощью OBIE-системы GATE представлены в работах Sibarani et al. [13] и Dadzie et al. [14]. Также на основе OBIE-системы построен европейский проект Skills Online Vacancy Analysis Tool for Europe, собирающий и анализирующий информацию из объявлений о работе в 15 европейских странах [15].

Основным ограничением подобных исследований является качество онтологии, используемой для анализа. Исследователи отмечают, что существующие классификаторы необходимо расширять вручную терминами, полученными в ходе первичного анализа собранных объявлений [13]. Учитывая это, при развитии нашей методики мы будем использовать смешанный подход к составлению словаря навыков: исходя из содержания объявлений и принятой в литературе классификации навыков.

Для осуществления автоматизации процесса сбора и анализа данных мы сделали выбор в пользу языка программирования Python по ряду причин: свободный доступ к средствам языка, наличие широкого круга библиотек для проведения анализа, простота освоения. Это позволяет сделать предложенный механизм реализации исследования легко переносимым на другие области.

Предлагаемая методика контент-анализа объявлений о работе предполагает реализацию следующих этапов исследования:

1. Определение источников (интернет-порталы о трудоустройстве) и характера собираемой информации о представленных на рынке вакансиях.

2. Определение необходимого периода и частоты сбора информации.

3. Проведение тестового сбора данных для проверки корректности выставленных критериев сбора и сортировки информации.

4. Сбор данных, осуществляемый с помощью парсинга источников с последующей сортировкой данных по категориям. В нашем исследовании мы проводили сбор информации с портала jobs.tut.by и приняли такие категории для сортировки как: «Должность», «Работодатель», «Дата публикации объявления», «Требуемый опыт», «Зарплата», «Описание вакансии», «Ключевые слова».

5. Удаление дубликатов, а также вакансий, нерелевантных выбранной области анализа. Удаление дубликатов происходит с помощью функции `DataFrame.drop_duplicates`.

6. Создание словаря навыков, состоящего из четырех уровней: категория навыков – группа навыков – навык – ключевые слова для поиска. Формирование словаря происходит в несколько этапов:

6.1. Присвоение группы первому указанному ключевому навыку;

6.2. Проверка последующих навыков на принадлежность существующим группам. Если подходящей группы не обнаруживается, создается новая группа. Синонимичные навыки вносятся сразу в графу «Ключевые слова для поиска»;

6.3. Сопоставление созданных исходя из полученного материала групп с теоретическим концептом. Происходит уточнение названий навыков и групп, их содержания, а также объединение в категории;

6.4. Расширение словаря навыков с учетом существующих классификаций и отраслевых аналитических отчетов.

6.5. Расширение списка ключевых слов синонимами и различными вариантами написания навыков с помощью сервисов «Планировщик ключевых слов Google Ads» и «Яндекс WordStat». Подбор синонимов и вариантов написания осуществляется по принципу снежного кома.

6.7. Запись ключевых слов для поиска с помощью регулярных выражений (рисунок 1). Для проверки правильности написания регулярных выражений использовался сервис [regex101.com](http://regex101.com).

Яндекс.Директ	яндекс\директ.*?[директ.*? direct реклам.{1,3}? яндекс.*? яндекс реклам.*? реклам.{1,3}? yandex yandex реклам.*?
Google AdWords	adwords google ads эдваордс.*?[адвордс google реклам.*? реклам.{1,7}? google гугл реклам.*? реклам.{1,7}? гугл
Mail.ru Group	\bmail\.ru
Amazon Advertising	amazon advertis.*?[амазон advertis.*? amazon реклам.*? реклам.{1,7}? amazon реклам.{1,7}? амазон амазон реклам.*?

**Рисунок 1 – Пример графы "Ключевые слова для поиска" с регулярным выражением**

7. Тестирование словаря на всей выборке и его доработка с учетом установленных аномально больших либо маленьких результатов. К примеру, в нашем случае были выявлены следующие причины их появления:

- двойной учет навыков (например, в первоначальный словарь навыку «SEO» были поставлены в соответствие ключевые слова «seo» и «оптимизация». Однако эти термины в описаниях вакансий зачастую встречаются вместе в виде понятия «seo оптимизация». Чтобы избежать двойной учет навыка на одну вакансию в словарь был добавлен более точный термин «поисковая оптимизация»);

- неточная формулировка ключевых слов, приводящая к учету более общих терминов (например, по запросу «mail» – определяющему сервисы компании Mail Group – происходил учет понятий, в которые входит слово «mail», например, «email»);

- учет навыков из разных групп по одному ключевому слову (например, запрос «ключевые слова» может встречаться, как в навыке «SEO», так и в навыке «Контекстная реклама»).

8. Анализ частоты встречаемости навыков в вакансиях. Сформированный словарь ключевых слов используется для подсчета ключевых слов в описаниях вакансий по такому принципу: скрипт находит в описании вакансии слова и словосочетания, соответствующие ключевым словам в словаре; в случае совпадения, соответ-

ствующему навыку приплюсовывается одно значение [16]. Расчет осуществляется с помощью синтаксиса регулярного выражения. Пример кода представлен на рисунке 2.

```
from collections import Counter
frequency_skills_exp = Counter()
n = 0
for job in description_exp:
    count = 0
    for item in dic_skill:
        pattern = r"(?i)[^a-яA-Яa-zA-Z0-9_|^]%s[^a-яA-Яa-zA-Z0-9_|$]" % dic_skill[item]
        try:
            if re.search(pattern, job):
                frequency_skills_exp[item] += 1
                count += 1
        except re.error:
            print("Проблема с элементом: " + dic_skill[item])

exp.iat[n, 3] = count
n += 1

frequency_skills_exp
```

**Рисунок 2 – Код для расчета частоты встречаемости навыков в описаниях вакансий**

9. Анализ частоты встречаемости групп навыков. В случае упоминания в описании вакансии нескольких навыков из группы, учёт группы происходит все равно один раз на вакансию. Это необходимо для того, чтобы в последующем можно было подсчитать частотность встречаемости групп навыков в числе вакансий.

10. Расчет частотных показателей:

- частотность по числу навыков – процент упоминания того или иного цифрового навыка или группы в общем числе упомянутых навыков или групп;

- частотность по числу вакансий – процент упоминания того или иного цифрового навыка или группы в общем числе вакансий.

11. Статистический анализ. Для статистического анализа результатов мы использовали следующие подходы:

- описательная статистика (функция `DataFrame.describe`);
- проверка распределения на нормальность (`scipy.stats.normaltest`);
- проверка дисперсий на гомогенность (`scipy.stats.levene`);
- преобразование Бокса-Кокса (`scipy.stats.boxcox`);

- непараметрические методы: критерий Спирмена (`scipy.stats.spearmanr`) U-тест Манна-Уитни (`scipy.stats.mannwhitneyu`), тест Краскела-Уоллиса (`scipy.stats.kruskal`);
- критерий Хи-квадрат Пирсона без и с поправкой Йейтса (`scipy.stats.chisquare`, `scipy.stats.chi2_contingency`);
- точный тест Фишера (`scipy.stats.fisher_exact`).

Встроенная функция точного теста Фишера предназначена для расчета показателя для таблицы размером  $2 \times 2$ . В нашем случае необходимо было рассчитать данный показатель для таблицы  $2 \times 3$ . В связи с этим была написана функция, представленная на рисунке 3.

```
def exact_fisher(observed, expected):
    a = observed[0]
    b = observed[1]
    c = observed[2]
    d = expected[0]
    e = expected[1]
    f = expected[2]
    numerator = (factorial(a+d)*factorial(b+e)*factorial(c+f) /
                 factorial(a)*factorial(d)*factorial(b)*factorial(e)*factorial(c)*factorial(f))
    denominator = factorial(a+b+c+d+e+f) / factorial(a+b+c)*factorial(d+e+f)
    exact_fisher = numerator / denominator
    return exact_fisher
```

**Рисунок 3 – Код для расчета точного критерия Фишера для таблицы размерностью  $2 \times 3$**

Размер эффекта для непараметрических методов был рассчитан с помощью онлайн-калькулятора [17], позволяющего осуществлять перевод показателей размера эффекта для непараметрических тестов в показатель  $d$  Коэна и эта-квадрат. Размер эффект при анализе с помощью критерия Хи-квадрат Пирсона оценивался с помощью коэффициента  $V$  Крамера. Для данного показателя не существует встроенной функции в Python, в связи с этим нами была написана собственная функция (рисунок 4).

```
import math
from scipy.stats import chi2_contingency

def cramers_v(observed):
    chisquare, p, df, expected = chi2_contingency(observed)
    rows, columns = observed.shape
    N = rows*columns
    m = min(rows-1, columns-1)
    V = math.sqrt(chisquare / N*m)
    return V
```

**Рисунок 4 – Код для расчета коэффициента  $V$  Крамера**

## 12. Описание полученных результатов, разработка рекомендаций.

Представленная методика позволяет автоматизировать большинство этапов контент-анализа объявлений о вакансиях, начиная от этапа сбора информации и заканчивая статистическим анализом. Наличие встроенных статистических функций позволяет провести обработку и анализ данных. Однако составление и корректировка словаря остается достаточно трудоемким ручным процессом. Решение данной задачи средствами машинного обучения все равно требует участия исследователя для проверки корректности работы созданных алгоритмов. В тоже время программные средства Python позволяют реализовать различные модели машинного обучения. Благодаря широкому перечню функций статистического анализа и возможности машинного обучения Python превосходит ПО GATE для проведения контент-анализа объявлений о работе.

### Литература:

1 A global challenge for one billion workers [Electronic resource]. – Mode of access: [https://worldskills2019.com/application/files/2715/6690/3205/ENG\\_BCG\\_Mas\\_Unikum\\_August\\_17.pdf&ved=2ahUKEwjH7cnb\\_eDkAhVImYsKHSWVCaAQFjAAegQIBBAB&usg=AOvVaw0DI5quriwBdu4rZLdPxXqI](https://worldskills2019.com/application/files/2715/6690/3205/ENG_BCG_Mas_Unikum_August_17.pdf&ved=2ahUKEwjH7cnb_eDkAhVImYsKHSWVCaAQFjAAegQIBBAB&usg=AOvVaw0DI5quriwBdu4rZLdPxXqI). – Date of access: 21.09.2019.

2 DESI 2019 – Human capital – digital inclusion and skills [Electronic resource]. – Mode of access: <https://ec.europa.eu/digital-single-market/en/human-capital>. – Date of access: 03.09.2019.

3 Wellman, N. The employability attributes required of new marketing graduates [Electronic resource]. / N. Wellman // *Marketing Intelligence & Planning*. – 2010. – Vol. 28, № 7. – P. 908–930. – Mode of access: <https://doi.org/10.1108/02634501011086490>. – Date of access: 20.07.2019.

4 Harper, R. The collection and analysis of job advertisements: a review of research methodology / R. Harper // *Library and Information Research*. – 2012. – Vol. 36, № 112. – P. 29–54.

5 Голомазова, Е.С. Исследование конкурентоспособности выпускников высшей школы с учетом требований работодателей / О.А. Морозевич, Е. С. Голомазова // *The economic discourse : Intern. coll. of sci. papers*. – 2016. – № 1. – С. 21–30.

6 Голомазова, Е.С. Развитие оперативного подхода к исследованию востребованных рынком труда компетенций / О.А. Морозевич,

Е.С. Голомазова // Научные труды Белорусского государственного экономического университета. – Мн. : БГЭУ, 2018. – Вып. 11. – С. 287-296.

7 Wellman, N. ARE MARKETING DEGREES FIT FOR PURPOSE? Re-shaping the undergraduate marketing curriculum to better equip early career marketing practitioners for employment and career progression [Electronic resource]. – 2017. – Mode of access: <https://www.researchgate.net/publication/322697679>. – Date of access: 09.07.2019.

8 Rosenstreich, D What attributes do employers seek in Marketing Graduates? / D. Rosenstreich, R. Bedggood, G. Priday // ANZMAC Conference, Brisbane, 1-3 December 2014.

9 Liu, X Designing a Marketing Analytics Course for the Digital Age [Electronic resource]. / X. Liu, A. Burns // Marketing Education Review. – 2018. – № 28:1. – P. 28-40. – Mode of access: <https://doi.org/10.1080/10528008.2017.1421049>. – Date of access: 09.07.2019.

10 Klézl, V. Assessing the personal traits and skill required for purchasing jobs / V. Klézl, J. Vasek, M. Kotrle, S. Saghiri // Purchasing and Supply Chain Management: Fostering Innovation Conference : IPSERA 2018., Athens, Greece.

11 Schlee, R.P. Job Requirements for Marketing Graduates: Are There Differences in the Knowledge, Skills, and Personal Attributes Needed for Different Salary Levels? [Electronic resource]. / R.P. Schlee, G.L. Karns // Journal of Marketing Education. – 2017. – № 39(2). – P. 69–81. – Mode of access: <https://doi.org/10.1177/0273475317712765>. – Date of access: 09.07.2019.

12 Konys, A. Towards Knowledge Handling in Ontology-Based Information Extraction Systems / A. Konys. // Procedia Computer Science. – 2018. – № 126. – P. 2208–2218.

13 Sibarani, E.M. Ontology-guided Job Market Demand Analysis: A Cross-Sectional Study for the Data Science field [Electronic resource] / E.M. Sibarani, S. Scerri, C. Morales, S. Auer, // Semantics 2017 : Proceedings of the 13th International Conference on Semantic Systems, Collarana, 2017 ; ed.: R. Hoekstra [et al.] – USA, New York. – P. 25-32. – Mode of access: <https://doi.org/10.1145/3132218.3132228>. – Date of access: 09.07.2019.

14 Dadzie, A.-S. Structuring visual exploratory analysis of skill demand [Electronic resource] / A.-S. Dadzie, E.M. Sibarani, I. Novalija, S.

Scerri // Journal of Web Semantics. – 2018. – Vol. 49. – P. 51-70. – Mode of access: <https://doi.org/10.1016/j.websem.2017.12.004>. – Date of access: 09.07.2019.

15 Skills-OVATE: skills online vacancy analysis tool for Europe [Electronic resource]. – Mode of access: <https://www.cedefop.europa.eu/en/data-visualisations/skills-online-vacancies>. – Date of access: 03.09.2019.

16 Применение Python для анализа смежных востребованных навыков у разработчиков [Электронный ресурс]. – Режим доступа: <https://m.habr.com/ru/post/423745>. – Дата доступа: 07.07.2019.

17 Lenhard, W. Calculation of Effect Sizes [Electronic resource] / W. Lenhard, A. Lenhard – Mode of access: [https://www.researchgate.net/publication/305280567\\_Computation\\_of\\_Effect\\_Sizes](https://www.researchgate.net/publication/305280567_Computation_of_Effect_Sizes). – Date of access: 03.09.2019.

**УДК 339.138:004.7.**

***О.И. Гринько***

*Белорусский государственный экономический университет  
Республика Беларусь, Минск  
[info@digivox.by](mailto:info@digivox.by)*

## **СПОСОБЫ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ МЕДИЙНОЙ РЕКЛАМЫ В УСЛОВИЯХ СОВРЕМЕННОГО РЫНКА ИНТЕРНЕТ-КОММУНИКАЦИЙ**

***Grinko O.I.** Belarus State Economic University, Republic of Belarus, Minsk.*

**WAYS OF IMPROVING EFFECTIVENESS OF MEDIA ADVERTISING ON THE MODERN INTERNET COMMUNICATIONS MARKET.** The media advertising market has faced a number of difficulties over the past few years: ad-blocking technologies, low performance indicators, increasing competition in Media Advertising Networks, etc. But the most difficult challenge is the decreasing audience's interest in this type of communication. Based on data from analytic marketing agencies, this article discusses the ways of getting out of situation using modern communication technologies, which also contribute to the sustainable development of brand awareness.

**KEYWORDS:** display ads; media advertising; marketing communications; brand awareness; content marketing; blockchain.