
NLP: CREATING SEARCH ENGINE FOR JOB VACANCIES



Zolotarevych Oleksandra
Student, BP CS-4



Smysh Oleh, research supervisor
Master of Software Engineering,
assistant, PhD student CS-3

Contents →

1. Problem statement
2. Brief overview of used NLP techniques
3. Analysis
4. Design and Development
5. Conclusions

Problem statement

Create a search engine for vacancies written in **natural Ukrainian language** using **NLP techniques** to structure given information.

NLP methods →

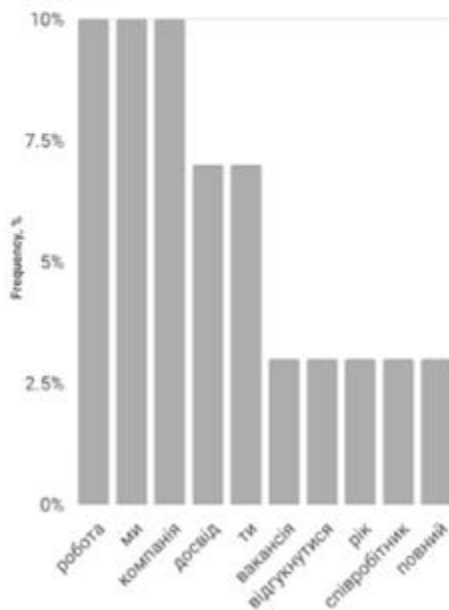
Tokenization	Lemmatization	Part of speech tagging
<p>This is the process of breaking a string of characters into words, punctuation, numbers - tokens.</p>	<p>The process of reducing inflected words to their word stem, base or root form</p>	<p>The process of identifying parts of speech, morphological features</p>

Analysis →

- Lemmatization, No conjunctions
- Lemmatization, conjunctions
- No lemmatization, conjunctions
- No lemmatization, No conjunctions

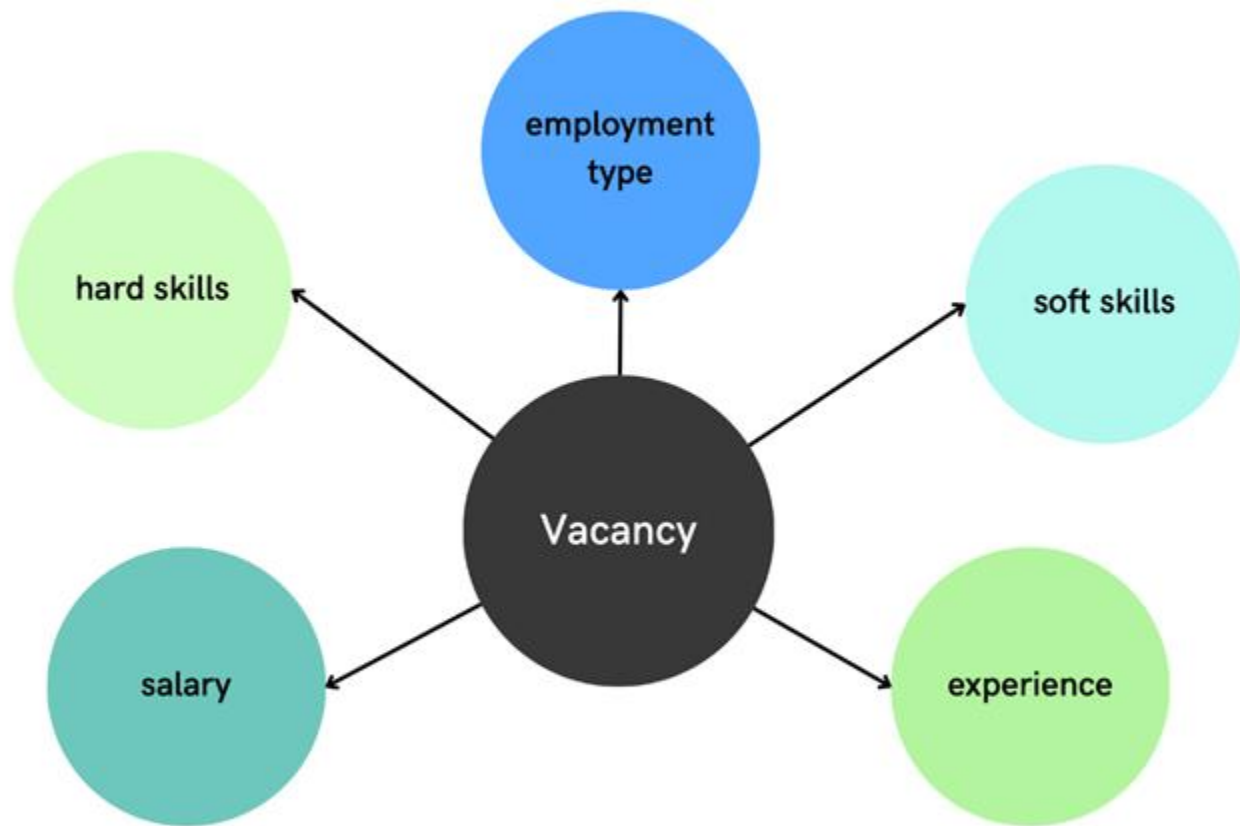
Sources: *work.ua, robota.ua, jooble.org*

Lemmatization, no conjunctions
1 word



*Illustration of frequency analysis
for all 3 sources.*

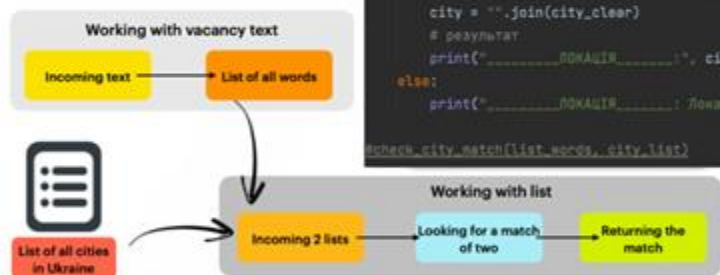
lemmatization, no conjunctions



Visualization of vacancy breakdown

Design and Development →

- Location
- Hard skills
- Soft skills
- Experience
- Salary
- Employment



```
# fowya cniannalnu
def check_city_match(l1, l2):

    # ПОСЛОТА З СЛІВНИ БАРОМ БАКАМЦІІ

    a_set = set(l1)
    b_set = set(l2)

    # перевірка на мстг
    if (a_set & b_set):
        city_match = str(a_set & b_set)
        # розбилення на мстг
        city_clear = [char for char in city_match if char.isalnum()]
        city = "".join(city_clear)
        # результат
        print("_____ДОКАЖІВ_____": city)
    else:
        print("_____ДОКАЖІВ_____": "Зовсім не сказано")

# check_city_match(list_words, city_list)
```

+

QA Model

Visual representation of structuring vacancy written in natural Ukrainian language

Віддалена робота

Внутрішній аудитор

50 000 грн

ROZETKA

5 днів тому

ROZETKA

Київ вул. Кирилівська, 82р-н. Подільський метро Тараса Шевченка

Віддалена робота

Гібридна

В офісі/на місці

Повна зайнятість

Найпопулярнішому інтернет-супермаркету Rozetka.ua потрібен «Внутрішній аудитор».

Що робити:

Контролювати правильність складання фінансової звітності. Контролювати договори. Контролювати діяльність та звітність на відповідність податковому законодавству.

Консультувати співробітників бухгалтерської служби та менеджменту з питань обліку, оподаткування та контролю діяльності. Приймати участь у розробці бізнес-процесів нових проектів.

Що для цього потрібно:

Вища освіта в галузі бухгалтерського обліку/фінансів/економіки. Досвід роботи в бухгалтерії від 5 років. Комунікабельність та стресостійкість вітається. Знання законодавства регулюючого ТЦО. Аналітичні здібності. Системний підхід до вирішення питань. Вміння та бажання працювати в команді.

З нашого боку: робота в команді професіоналів. Цікаві задачі.

Прохання в резюме вказувати очікуваний рівень заробітної плати.

```
↑ ЛОКАЦІЯ: Київ
↗ ІНСТРУМЕНТИ: не вказано
● ДОСВІД:
Досвід роботи в бухгалтерії від 5 років.

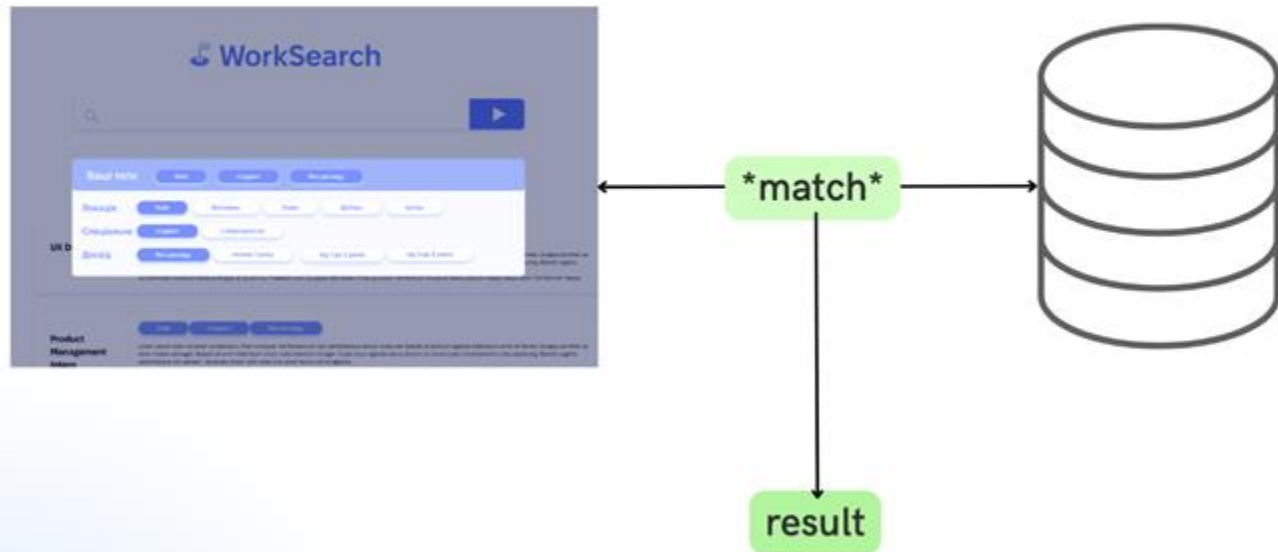
Знання міжнародних стандартів бухгалтерського обліку та звітності, досвід їх практичного застосування.

💰 ЗАРОБІТНА ПЛАТА: 50 000 грн
    Валюта: ['UAN']
🔧 ЗАЙНЯТІСТЬ:
Повна зайнятість

👉 ГНУЧКІ НАВИЧКИ: ['Комунікативний', 'Стресостійкість']
```

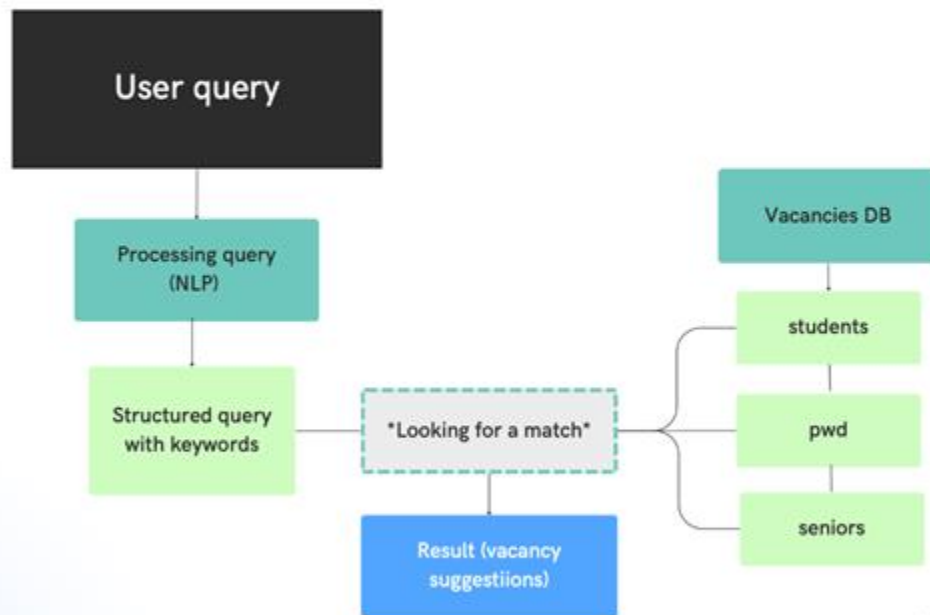


Search based on match

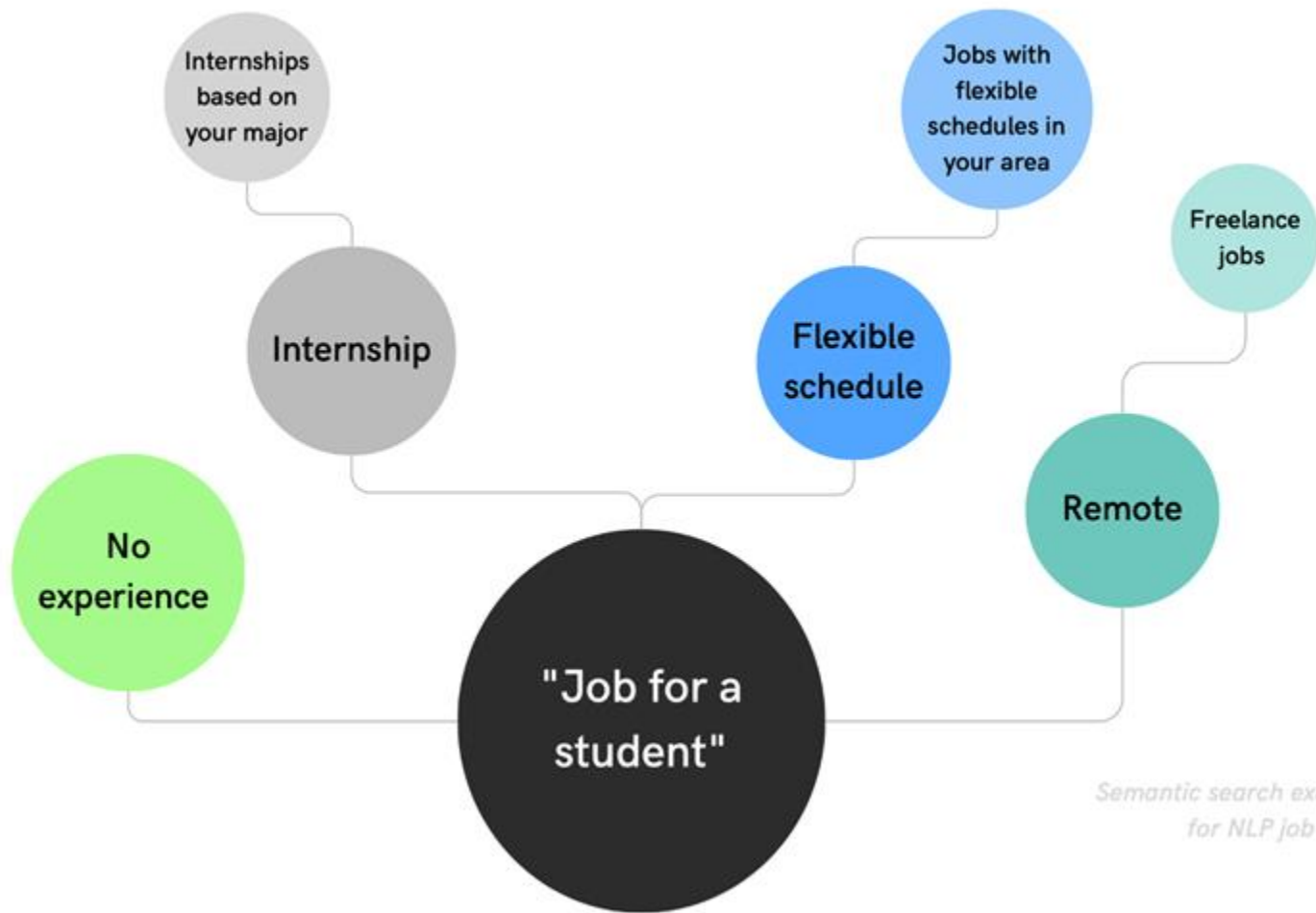


Visualization of match based search

Realization of semantic search in a job search engine



Visualization of semantic principle search



*Semantic search example schema
for NLP job search engine*

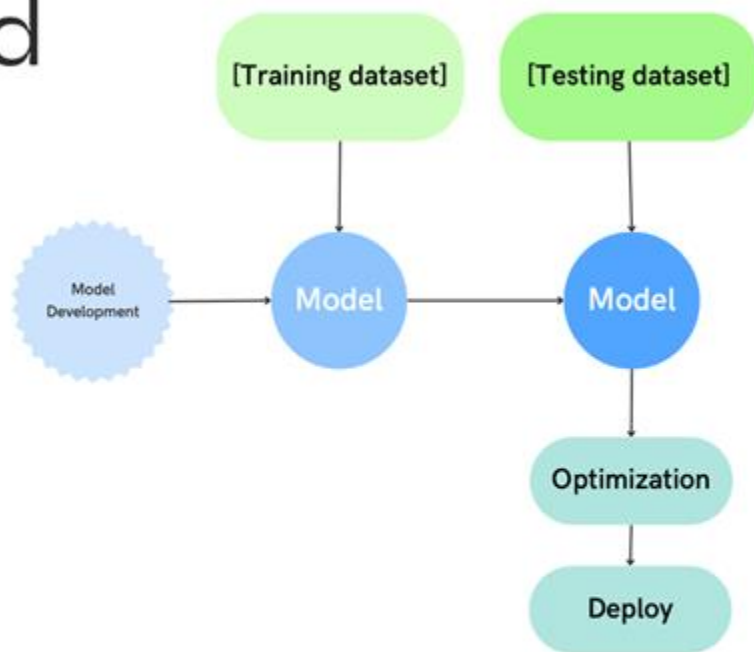
Semantic Markup

Marker \ Group	STUDENTS	PWD	PROFESSIONALS
Experience	0 - 1 years	1+	5+ years
Employment type	Flexible, part-time	Flexible, full-time	Full-time
Location	Remote, On-site	Remote, On-site	On-site
Compensation	\$ - \$\$	\$ - \$\$\$	\$\$ - \$\$\$

ML – Supervised learning

Naive Bayes TRAIN	0.885233	0.881579	0.876448
Naive Bayes TEST	0.794778	0.807818	0.792141

Model evaluation



Depiction of model development algorithm

Approbation

- Leipzig University, Germany, Chance for Science Conference 22/04/2022
- Leipzig University, Germany, Chance for Science Conference 08/09/2022
- XXIII All-Ukrainian scientific and technical conference "State, achievements and prospects of information systems and technologies", 21/04/2023

As a result of this work, the field of NLP was researched, was performed analysis and based on it methods of the parser were optimized. To ensure the realization of the semantic search, a model was established.



**Асистент
(помічника
менеджера з
продажу)**



6 000 — 10 000 \$/тиждень топу DigitalMarketing/владана робота/певна зайнятість/Доброго дня, колеги! Ми шукаємо Помічника менеджера з продажу / Стюарта з продажу / Sales Assistant/Про компанію Digital — це агентство Google Partners. Наші основні напрями: контекстна реклама, підтримка та веб-аналітика. Ми працюємо з клієнтами з України, Австралії, США, Канади, Європи, англієцькою та українською мовами. Ми не працюємо з інформаційними та контентними проектами, інформацією, MLM, форекс, казино.

**Помічник
менеджера з
продажу**



Ми, компанія "Силь-ком", - виробник та постачальник на ринку України комплектуючих і аксесуарів до мобільних пристроїв (телефонів). Ми - одні із лідерів на ринку України і наша продукція завжди залишається продуктом першої необхідності. Найголовніше, що кожен з нас зараз може зробити для себе та країни, - це зробити свій внесок в розвиток, висловити та підтримати себе та свою сім'ю, обрати стабільний та перспективний сегмент сфери своєї діяльності. Найцініше, що в нас є, - це люди поруч, так давайте об'єднаємося та досягнемо вершин розвитку та конкурентоспроможності, енергійності і зацікавленості у саморозвитку, ми запрошуємо тебе в нашу команду на посаду асистента нашої продажів.

**Асистент
департаменту
аналітичного
обладнання**



З 2004 року ми забезпечуємо сучасними рішеннями лабораторії в таких галузях як харчова промисловість, охорона здоров'я, екологія, енергетика, наука, криміналістика, матеріалознавство, фармація та виробництво НАСА/МКС. Розкриваємо потенціал інноваційних технологій задля здоров'я та безпеки нашій Україні. Асистент департаменту аналітичного обладнання/Основні відповідальності документування в 10 довідок, комерційні пропозиції, розуміння технічного термінологічного документа з акцентом на українську мову/контакти з деякими клієнтами.

Visualization

Thank you

