

О. К. Слухаєвська, Л. М. Захарія
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

ІНТЕЛЕКТУАЛЬНІ ЗАСОБИ В СИСТЕМІ ПРАЦЕВЛАШТУВАННЯ

© Слухаєвська О. К., Захарія Л. М., 2018

У роботі представлено проект системи, яка здійснює двосторонній процес пошуку – кандидата на вакансію та автоматизований пошук вакансій для кандидата. Для цього постійно відслідковується інформація про наявні вакансії засобами web mining. Отримана інформація про нові вакансії класифікується з точки спорідненості інформації з раніше визначеними класами вакансій, які відіграють роль навчальної вибірки. Використання алгоритмів і методів машинного навчання дозволяє підвищити ефективність процесу підбору підходящої роботи та скорочення часу пошуку кандидатів на заявлені вакансії.

Запропонований підхід збільшує вплив індивідуальних якостей шукачів роботи на процес, а отже, і на результат роботи системи пошуку вакансій використовуючи інструменти та методи штучного інтелекту. Система сьогодні спроектована для ІТ сфери, як найрозвиненішої з погляду структуризації вимог до кандидатів та з урахуванням великої кількості пропозицій на цьому ринку праці.

Ключові слова: машинне навчання, штучний інтелект, вакансія, рекрутинг, робота, пошук роботи.

The paper considers the project of the system that carries out the bilateral process of search – candidate on a vacancy and automated search of vacancies for a candidate. For this purpose, information on available vacancies through web mining is constantly monitored. The obtained information on new vacancies is classified in terms of information related to previously defined classes of vacancies that play the role of a training sample.

The use of algorithms and methods of machine learning allows to increase the efficiency of the selecting suitable work and reducing the search time for candidates for vacancies process.

The proposed approach increases the influence of the individual qualities of recruiters on the process and, consequently, on the result of the job search system by employing tools and methods of artificial intelligence. The system is currently designed for the IT sector, as the most developed in terms of structuring the requirements for candidates and taking into account the large number of proposals in this labor market.

Key words: machine learning, artificial intelligence, recruiting, opening, job, search for a job.

Вступ

Сьогодні все більш затребуваними є результати всебічного аналізу даних, наявних саме у мережі інтернет. У результаті настання таких значних змін та підходів до отримання даних у інформаційному просторі, актуальним є вирішення питань використання результатів такого аналізу у різних галузях та сферах життя. На сьогодні вже існує багато готових інструментів, які надають можливість збирати, аналізувати дані та приводити їх до потрібного вигляду для подальшого опрацювання та/або збереження.

У цій статті буде розглянуто використання таких інструментів на прикладі організації функціонування сайту для рекрутингу.

Аналіз останніх досліджень та публікацій

Швидкий розвиток сучасних технологій та постійне вдосконалення алгоритмів машинного навчання та “еволюція” розробки штучного інтелекту стає основою для публікацій та досліджень за цією тематикою. Сфери застосування таких розробок у сучасному житті такі різноманітні, що систематизувати та виділити найбільш затребувані або найперспективніші серед них дуже складно і недоцільно. Власне, жодна із сфер життєдіяльності сучасної людини не унеможливує застосування технології штучного інтелекту. Розробники, науковці та дослідники приділяють увагу цим технологіям і визнають їхній значний вплив на майбутнє людства. Серед усіх доступних робіт у цьому сегменті знань, для аналізу питання щодо застосування штучного інтелекту у сучасних інформаційних системах було оцінено такі наукові роботи як Liu, Bing “Web Data Mining. Exploring Hyperlinks, Contents, and Usage Data, Second Edition” [1], Yaser S. Abu-Mostafa, Malik Magdon-Ismael, Hsuan-Tien Lin Learning from data [2], Vihar Kurama “Linear Algebra for Deep Learning” [3], Divyansh Dwivedi “Machine Learning for beginners” [4], Amid Fish “Lessons Learned Reproducing a Deep Reinforcement Learning Paper” [5], Drew Conway, John Myles White “Machine Learning for Hackers” [6]. Однак, думки дослідників лише частково розкривають питання застосування методів штучного інтелекту в інформаційних системах, оскільки перед будь-якою системою постають унікальні проблеми, вирішення яких передбачає внесення змін у стандартні алгоритми опрацювання даних. Отже вивчивши дані напрацювання, було отримано необхідні знання для постановки проблем та розробки проекту.

Постановка проблеми

Важливою проблемою користувача, який шукає роботу або здійснює пошук кандидата на певну посаду, є відсутність уніфікованої системи, яка допомогла б описати наявні професійні навички або вимоги до навичок кандидата, якого шукає роботодавець. Також однією із проблем є відсутність прозорості процесу у комунікації між рекрутером та кандидатом. Процес підбору працівників складається з певних кроків та етапів, що можуть мати визначені часові паузи та періоди, і весь цей час, і кандидат і його резюме перебувають у так званому “актуальному” стані для рекрутера. Звідси виникає проблема, коли рекрутеру треба швидко згадати, на якому етапі зупинилися переговори з конкретним кандидатом, дату проведення інтерв'ю, або на якій стадії розгляд кандидатури претендента з точки зору роботодавця.

Під час обробки резюме кандидатів на посаду працівнику з кадрових питань складно оцінити заочно певні особисті якості та нахильності людини. Кожен рекрутер має набір стандартних тестів, які допомагають йому визначитись із такими якостями кандидата. Тести, які використовують рекрутери, можуть бути уніфіковані та внесені до системи і пропонуються для проходження тестування визначеним кандидатам.

Одним із шляхів нейтралізації впливу таких проблем на процес заповнення вакансії є автоматизація опису затребуваних та наявних професійних навичок: система вже містить великий набір навичок, тому користувачі мають обрати необхідні під час пошуку або створення вакансій.

Завдяки засобам штучного інтелекту можливо ефективно вирішити проблему скорочення часу пошуку вакансії або кандидата. Спираючись на те, за якими саме критеріями користувач опрацьовує пошук, система автоматично пропонує вакансії, які можуть бути цікавими для користувача. Одночасно, аналізуючи інформацію про успішно укомплектовані вакансії, система підлаштовується самостійно і змінює так зване коло кандидатів, резюме яких, найімовірніше, відповідають умовам подібних вакансій.

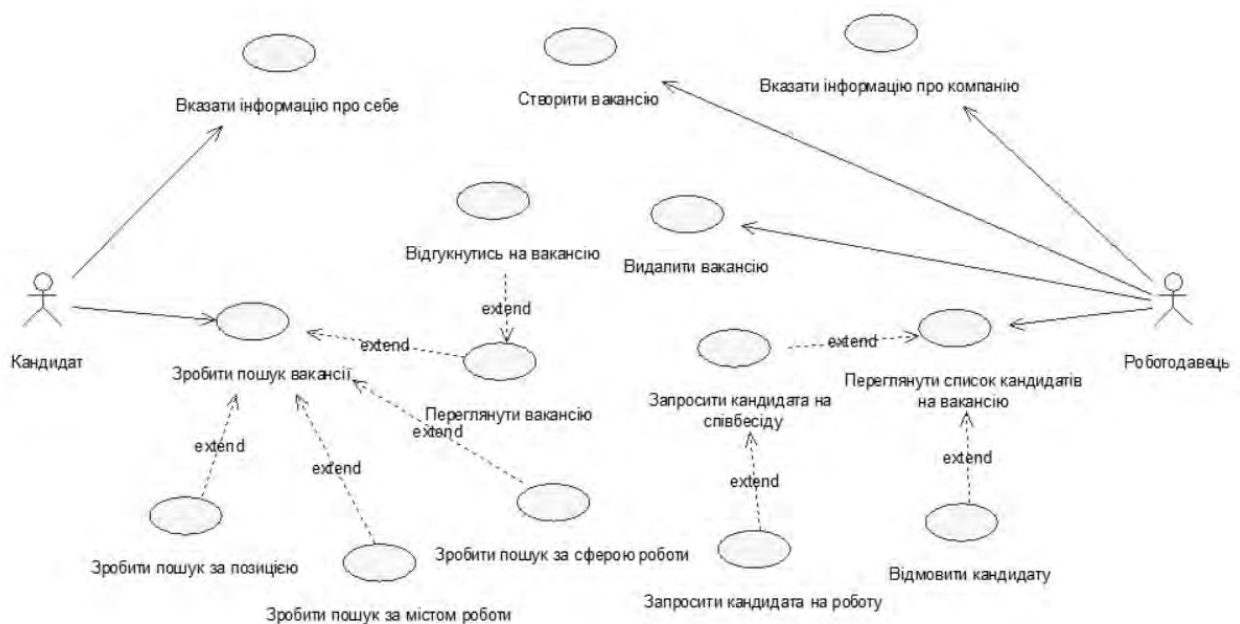
Виділення основних вимог до системи

Для такої системи було виділено два основні типи користувачів: кандидат та роботодавець. Система має по-різному взаємодіяти з кожним типом користувача, надаючи лише необхідні інструменти для роботи.

У результаті проведеного обстеження предметної області “Пошук роботи” отримано такий перелік реквізитів, які належать до інформаційних об’єктів Кандидат, Компанія-роботодавець:

- прізвище та ім’я кандидата;
- досвід роботи кандидата;
- місце проживання кандидата;
- сфера попередньої роботи кандидата;
- позиція кандидата;
- назва компанії-роботодавця;
- опис компанії-роботодавця;
- вакансія роботодавця;
- деталі вакансії – місто, сфера роботи, позиція.

На підставі складеного переліку реквізитів було складено діаграму варіантів використання, яку представлено на рисунку.



Діаграма варіантів використання

Визначені сценарії використання є основними функціональними вимогами до системи, які мають підтримуватись для її успішного функціонування:

- заповнити інформацію в профілі;
- згенерувати резюме;
- розмістити вакансію;
- зробити пошук вакансії;
- відгукнутись на вакансію;
- видалити вакансію;
- переглянути список кандидатів на вакансію;
- запросити кандидата на співбесіду;
- запросити кандидата на роботу;
- відмовити кандидату.

Головною метою розробки та впровадження такого проекту є:

- скорочення терміну укомплектування вакансій;
- досягнення ефективного використання наявної у системі інформації та підвищення якості автоматичного підбору роботи;

– для кандидата: надання зручного засобу для пошуку роботи при докладанні мінімуму зусиль. Користувач має заповнити профіль тільки один раз, щоб забезпечити себе пошуком вакансій самі у тій сфері, яка його цікавить;

– для роботодавця: окрім стандартного функціоналу розміщення вакансій, є можливість додаткового тестування потенційних кандидатів та конфіденційний обмін інформацією про кандидатів між роботодавцями, можливість синхронізації та оптимізації графіків і планів роботи осіб, відповідальних за підбір персоналу.

Використовуючи цей сервіс, кандидат на посаду не витрачає час на написання резюме, а використовує вже наявні анкетні дані, які надають можливість скласти резюме автоматично на сайті. Таким самим чином рекрутер має можливість згенерувати резюме для обраного кандидата.

Обмеження, накладені на систему, та їх обґрунтування

Обмеження, які вводяться для коректного функціонування системи пошуку роботи, пов'язані з необхідністю уніфікувати категорії зіставлення даних про наявну вакансію з інформацією про навички і професійні якості кандидата, які вніс до системи шукач роботи. Система містить обмежений набір навичок, що дозволяє поділити їх на певні категорії та представити у вигляді ієрархічного дерева, що, своєю чергою, дає змогу використовувати на них алгоритми кластеризації для розширеного пошуку. Також заздалегідь визначений набір навичок дозволяє і рекрутерам, і кандидатам маніпулювати одними термінами, що вносить ясність у процес комунікації.

Перелік сайтів, які використовуватимуться для пошуку вакансій, буде обмеженим, оскільки наразі в Україні є визначені лідери у ніші рекрутингу. Надалі планується збільшити перелік сайтів, а згодом повністю автоматизувати цей процес, відмовившись від заздалегідь визначених сайтів.

Так само обмеження накладені на перелік тестів, що пропонуються для проходження шукачам роботи. Тестування кандидатів допомагає здійснити оцінку здібностей та якостей кандидатів і є одним з етапів проходження відбору та підвищує його ефективність. Таке обмеження пов'язане з тим, що тестування кандидатів може проводитись за багатьма методиками і їхній перелік постійно зростає. Але результати тестування не є самою ціллю функціонування системи, а тільки доповнюють картину професійних та особистих якостей кандидатів.

Перелік методик, за якими може проводитись тестування кандидатів, в згодом може бути змінений, зважаючи на потребу та актуальність для користувачів рекрутерів.

Резюме, яке можна згенерувати на сайті, має заздалегідь визначену структуру. Це дає змогу уніфікувати вигляд резюме, що робить процес їхнього перегляду швидким та зручним для рекрутера.

Структуризація даних

Для точного зіставлення даних між вакансіями та профілями кандидатів (резюме), використовуються однакові набори заздалегідь визначених категорій та навичок, які до них належать.

Роботодавець під час заповнення вакансії обирає необхідну галузь роботи, відповідні навички та інструменти.

Кандидат заповнює:

– досвід роботи у певній галузі, навички та/або інструменти, якими він користувався;
– бажана галузь роботи та навички, які були отримані самостійно або з попередніх місць роботи.

Як з однієї, так з іншої сторони, можливо вказати необхідний/наявний рівень володіння навичками. Система надає можливість обрати з трьох наявних рівнів: початківець, досвідчений, експерт.

Процес пошуку вакансій

Процес пошуку вакансій направлений не тільки на зменшення терміну укомплектування вакансії, але і на підвищення якості самого результату працевлаштування. Значною мірою на результат підбору кандидатів для вакансії в цій системі впливають деталізовані вимоги до кандидатів на посаду, що зіставляються з так само деталізованими даними з резюме шукачів роботи. Впровадження такої системи дозволяє здійснювати якісніший та індивідуальний підбір роботи. Деталізація вказаних критеріїв відбувається шляхом здійснення самооцінювання кандидатами рівня наявного досвіду роботи та професійних здібностей за визначеною шкалою. Внесення користувачами цієї інформації до системи відбувається за запропонованим готовим набором категорій, що дають розширену картину професійних та особистих якостей кандидата, та дають змогу не тільки сформулювати уяву про кандидата для організації інтерв'ю з рекрутером, але й значно впливають на результат автоматичного підбору роботи.

Система має можливість автоматично підбирати потенційно цікавих вакансій для кандидата. Процес підбору вакансій здійснюється на основі тих даних, які кандидат надав про себе, заповнюючи профіль на сайті. Якщо немає вакансій, які в точності відповідають даним, наданим кандидатом, то критерії пошуку розширюються за допомогою кластеризації:

- спочатку в пошук включаються суміжні категорії, якщо потрібний результат не був досягнений, то включаються категорії, які перебувають на рівень вище. Так триває доти, поки результат не буде досягнений або не закінчатся категорії, які можна включити до пошуку;

- такі дані, як вік кандидата, роки досвіду, локація також отримають розширені критерії пошуку для досягнення кращого результату. Межі для віку кандидата та його років досвіду розширяться як числові дані. Для локації буде використаний такий самий алгоритм як для категорії (заснований на ієрархії).

Що більше даних було зазначено у профілі, то точніший пошук буде виконаний.

Також на процес підбору вакансій має вплив результат укомплектування аналогічних вакансій у минулому. Під час створення вакансії, на основі вибраних категорій відбувається процес її кластеризації, в результаті якого дана вакансія потрапляє в певну групу аналогічних їй. Якщо при успішному закритті вакансії рекрутер був вказав на кандидата з числа користувачів системи, то на основі його профілю створюється або коректується оптимальний профіль кандидата для поточної кластеризованої групи вакансій. Завдяки такому процесу, усім користувачам, які володіють знаннями та навичками наближеними до оптимального профілю, будуть запропоновані вакансії з поточної кластеризованої групи. Процес підбору кандидатів відносно оптимального профілю відбувається за допомогою наївного класифікатора Баеса. Застосування цього алгоритму дає змогу врахувати ймовірність того, що наявні вакансії можуть бути укомплектовані з врахуванням успішності закриття вакансій такого кластеру.

Особливістю системи є те, що кандидатам поруч із вакансіями, запропонованими у цій системі, пропонуються там само вакансії з інших сайтів. Для організації такого пошуку на сторонніх ресурсах використовується автоматизований збір інформації засобами Web Mining.

Для пошуку на сторонніх ресурсах відбувається парсинг веб-сторінок – веб-скрапінг. Під час веб-скрапінгу відсилається запит до веб-ресурсів і, отримавши результат, відбувається парсинг HTML-документу на токени та пошук необхідних даних по складеному дереву. Отримані результати відображаються користувачу у прийнятній формі разом із посиланням на оригінал вакансії.

У майбутньому можливе підключення веб-краулерів, які самі зможуть шукати ресурси для парсингу.

Веб-краулер – це програма, яка призначена для перебору сторінок у мережі Інтернет і занесення їх до бази даних пошукової системи, де використовується веб-краулер.

Наразі визначено, що для задоволення потреб системи можна використовувати краулер таких типів:

- сфокусований краулер – завантажує лише ті сторінки, які відповідають обраній тематиці;

– інкрементний краулер – традиційний вид краулеру, який періодично оновлює усі документи, які вже зберігаються.

Висновки

Процес підбору працівників і успішне укомплектування вакансій має бути взаємовигідним і для працівника, і для роботодавця. Отже, у позитивному результаті зацікавлені обидві сторони. На вирішення актуальних проблем у сфері найму працівників і скерований цей проект з розробки системи пошуку роботи. Основною відмінністю цієї системи є її унікальність, швидкість реалізації та невелика вартість реалізації порівняно з глобальними проектами, оскільки будуть використовуватись готові інструменти, які просто потребують необхідного налаштування.

Розробка цієї системи може мати успішну реалізацію, оскільки проблема, яку вирішує цей проект, є актуальною, а підхід до її вирішення має інноваційний характер.

Проект поєднує вже добре відомі функції, які є у аналогів, та пропонує нові, які значно полегшують процес пошуку роботи/кандидата. Так само проект є інноваційним, оскільки такий процес пошуку є унікальним у визначеній сфері, адже використовуються такі інструменти машинного навчання як парсинг веб-сторінок, кластеризація ресурсів та веб-майнінг.

Також проект має великий потенціал для розвитку та вдосконалення як і з бізнес-сторони (впровадження нового функціоналу, такого як інтеграція з сторонніми ресурсами, партнерство з компаніями-роботодавцями та компаніями по пошуку кандидатів та ін.) так і з інноваційної (вдосконалення існуючих алгоритмів для швидшого та точнішого пошуку, повна автоматизація процесу пошуку за допомогою використання веб-краулерів).

1. L. Bing. *Web Data Mining. Exploring Hyperlinks, Contents, and Usage Data, Second Edition* / L. Bing // “Springer”. – 2016. – 851 p. 2. Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin. *Learning from data* / Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin // “Williams”. – 2015. – 542 p. 3 Vihar Kurama. *Linear Algebra for Deep Learning* / Vihar Kurama // *Towards Data Science*. – 2017. – 1–12 p. 4 Divyansh Dwivedi. *Machine Learning for beginners* / Divyansh Dwivedi. // *Towards Data Science*. – 2016. – 1–7 p. 5. Amid Fish. *Lessons Learned Reproducing a Deep Reinforcement Learning Paper* / Amid Fish // *Amid Fish. Blog*. – 2018. – 1–39 p. 6. Drew Conway, John Myles White. *Machine Learning for Hackers* / Drew Conway, John Myles White // O’Reilly. – 2014. – 746 p.