

Інформаційна технологія парсингу резюме

Вінницький національний технічний університет

Анотація

У роботі проаналізовано актуальність розробки інтелектуального модуля для парсингу резюме. В даній роботі використання технології NER для класифікації тексту на поіменовані сутності визначення тону повідомлення, а також визначення скілів, що належать кандидату.

Ключові слова: інтелектуальний модуль, поіменовані сутності, парсинг, аналіз тону, скілли.

Abstract

The paper analyzes the relevance of developing an intelligent module for resume parsing. In this paper, the use of NER technology for the classification of text into named entities determines the tone of the message, as well as the determination of skills belonging to the candidate.

Keywords: intellectual module, named entities, parsing, tone analysis, skills.

Вступ

У зв'язку зі стрімким зростанням Інтернет-рекрутингу, рекрутингові системи стикаються з великою кількістю особистих резюме, що потребують уваги. Щоб привернути увагу рекрутерів, багато кандидатів різноманітними способами форматують свої резюме, використовуючи різний розмір шрифту, кольори та таблиці. Хоча існують контрольовані та правило-засновані методи для виділення інформації з резюме, вони сильно залежать від ієрархічної структури та великих обсягів маркованих даних, які зібрати складно.

В рамках дослідження було враховано визначення тону повідомлення в резюме з використанням методів обробки природної мови. Такий аналіз дозволяє виявляти емоційний настрій та ставлення кандидата до різних аспектів його професійної діяльності. Це може бути корисно при відборі кандидатів, оскільки допомагає встановити їхню мотивацію, комунікативність та відповідність корпоративній культурі. Крім того, для покращення процесу відбору було використано онтологію даних, що дозволяє "вгадувати" скіли кандидата на основі аналізу його резюме. Шляхом автоматичного визначення ключових слів та понять у тексті, система може ідентифікувати навички кандидата, що не вказані у резюме.

Таким чином, синтаксичний аналізатор резюме в поєднанні з аналізом тону повідомлення та використанням онтології даних створює потужний інструмент для ефективного відбору та аналізу кандидатів у процесі рекрутингу

Результати дослідження

Штучний інтелект (ШІ), заснований на машинному навчанні та глибокому навчанні, може допомогти в оцінці резюме людини та навіть створити градацію відповідності резюме до певної позиції у фірмі. Синтаксичний аналізатор резюме — це система глибокого навчання/AI, яка визначає повну інформацію з резюме, аналізує, зберігає та упорядковує.

Синтаксичний аналізатор резюме — програмний продукт, який перетворює неструктуровані дані в структуровану форму. Це компонент, який автоматично розділяє інформацію за різними областями та параметрами, як-от контактна інформація, освітня кваліфікація, досвід роботи, навички, досягнення, професійні сертифікати, щоб швидко допомогти вам визначити найбільш релевантні резюме на основі ваших критеріїв.

Всі документи, що подаються на вхід системи, мають пройти попередню обробку, яка полягає у екстрактуванні необробленого тексту та приведенні до формату, який сприймає нейронна мережа та SVM класифікатор. А той текст, на якому навчається нейронна мережа та SVM класифікатор, ще повинні піддаватися попередній обробці – видаленню пунктуації, видаленню слів, які не передають семантику тексту, приведенню до нижнього регістру та визначенню частини мови слова. Після цього відбувається навчання нейронної мережі та SVM класифікатора на класифікацію тексту.

Архітектура моделі — це функція, яка підключає екземпляр моделі, який потім можна використовувати в компоненті конвеєра або як рівень більшої мережі.

Люди не починають думати з чистого листа щомиті. Ми не викидаємо з голови все і не починаємо думати з нуля. Наші думки мають постійність.

Традиційні нейронні мережі не мають цієї властивості, і в цьому їхній головний недолік. Уявімо, наприклад, що ми хочемо класифікувати події, що відбуваються у фільмі. Незрозуміло, як традиційна нейронна мережа могла використовувати міркування про попередні події фільму, щоб отримати інформацію про наступні.

Вирішити цю проблему допомагають рекурентні нейронні мережі (Recurrent Neural Networks, RNN) [1]. Це мережі, що містять зворотні зв'язки та дозволяють зберігати інформацію. Зворотні зв'язки надають рекурентним нейронним мережам певну загадковість. Проте, якщо подумати, вони не дуже відрізняються від звичайних нейронних мереж. Рекурентну мережу можна розглядати як кілька копій однієї і тієї ж мережі, кожна з яких передає інформацію наступної копії.

Одна з привабливих ідей RNN полягає в тому, що вони потенційно вмюють пов'язувати попередню інформацію з поточним завданням, наприклад, знання про попередній кадр відео можуть допомогти в розумінні поточного кадру. Якби RNN мали таку здатність, вони були б надзвичайно корисні. На жаль, у міру зростання відстані між інформацією на нейронах – RNN втрачають здатність пов'язувати інформацію.

Глибоке навчання [2]– це метод машинного навчання, який передбачає самостійне вибудовування (тренування) загальних правил як штучної нейронної мережі з прикладу даних під час процесу навчання. Навчання нейронної мережі, особливо систем обробки природної мови, зазвичай проводиться з учителем, т. е. на конкретних прикладах даних із попередньо визначеними їм результатами. Поняття «глибоке навчання» визначає навчання про «глибоких» штучних нейронних мереж. Крім вхідного та вихідного шару, вони складаються із сотень додаткових «прихованих» шарів між видимими шарами для введення та виведення. Отримана матриця прихованого шару є матрицею вхідних даних для наступного шару. У цьому випадку лише матриця вихідних даних останнього шару містить результат.

При проектуванні ми повинні дотримуватись основних рекомендацій до вибору параметрів архітектури глибоких нейронних мереж, що дозволяють забезпечити найбільшу ефективність роботи мережі і точність одержуваних результатів. Дані рекомендації складені на основі аналізу існуючих найбільш ефективних архітектур мереж.

Рекомендації до архітектури вхідного шару мережі:

- Текст з розширення UTF-8.
- Документ може подаватись з будь якою кількістю стрічок.
- Текст у документі повинен бути англійською мовою.

Зважаючи на швидкий розвиток штучного інтелекту, бажання роботодавців отримати якнайбільшу користь від кандидата, а також на необхідність визначати достойних кандидатів на конкуренті посади – такий сервіс був би дуже корисний для будь якого бізнесу, що включає у себе роботу людей з різних сфер.

На сьогоднішній день існує кілька аналогів сервісів з парсингу резюме. Ось перелік деяких з них:

Affinda: The Enterprise Grade Resume Parser [3] - Affinda надає продукти штучного інтелекту корпоративного рівня та індивідуальні рішення AI, використовується для формату PDF, оптимізує та стандартизує процес розгляду резюме. Архітектура глибокого навчання Affinda оцінює резюме та CV, витягуючи попередньо визначені текстові поля в формат Excel. Їх продукт автоматично визначає та виділяє ключові компоненти резюме відповідно до вибраних клієнтом критеріїв — зосереджуючись на найбільш доречних словах і фразах. До недоліків цього сервісу можна віднести низьку результативність, а також те, що даний сервіс є платним і досить дорогим.

Також розглянемо ще один продукт ALEX Resume Parser [4], це синтаксичний аналізатор резюме, який використовує кілька стратегій штучного інтелекту, включаючи методи обробки природної мови та розпізнавання шаблонів, щоб аналізувати відповідну інформацію з резюме. Як граматичний аналізатор резюме, ALEX надає значення термінам (словам і реченням) на основі контексту, в якому вони з'являються. До недоліків цього сервісу можна віднести невисоку достовірність роботи, прив'язаність до шаблону документа, а також те, що він являється платним.

Nanonets: The Efficient Resume Parsing Solution [5] - Nanonets пропонує передову службу аналізу резюме, яка використовує передові алгоритми штучного інтелекту для вилучення та впорядкування відповідної інформації з резюме. Завдяки спрощеному підходу Nanonets спрощує й автоматизує процес

перегляду резюме для компаній будь-якого розміру. Використовуючи найсучасніші методи машинного навчання, Nanonets точно визначає та витягує ключові дані з резюме, такі як контактна інформація, історія освіти, досвід роботи, навички тощо. Він розумно аналізує неструктурований текст і перетворює його на структуровані дані, дозволяючи рекрутерам і спеціалістам з кадрів легко переглядати та порівнювати інформацію про кандидатів. До недоліків можна віднести те, що сервіс може не визначати певні частини резюме через компактний опис, але конкретні обмеження чи потенційні недоліки служба Nanonets прямо не згадує.

Висновки

Після порівняльного аналізу сервісів парсингу резюме, таких як Affinda, Nanonets та ALEX Resume Parser, можна зробити висновок, що новий сервіс, що вміє екстрагувати скілли та інші поля з резюме, визначати семантичний тон повідомлення та вгадувати скілли, що належать кандидату, пропонує прогресивні технології для ефективного аналізу резюме. У порівнянні з Affinda та ALEX Resume Parser, новий сервіс відрізняється високою точністю та швидкістю обробки великого обсягу резюме. Враховуючи його здатність працювати з різними форматами резюме, цей сервіс є більш гнучким та універсальним.

Загалом, новий сервіс парсингу резюме представляє перспективну інформаційну технологію, яка може значно полегшити процес аналізу резюме та покращити якість рекрутингових процедур.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ларінов А. LSTM – мережі довгої та короткострокової пам'яті. Хабр. URL: <https://habr.com/company/wunderfund/blog/331310/>.
2. Мальцев. Огляд принципів глибокого навчання як динамічної теорії штучного інтелекту. URL: https://tech.vernadskyjournals.in.ua/journals/2021/6_2021/16.pdf
3. What Should You Look for in an Enterprise Resume Parser?. Affinda. URL: <https://www.affinda.com/recruitment-ai/what-should-you-look-for-in-an-enterprise-resume-parser>.
4. HireAbility ALEX Resume Parser offers outstanding accuracy. *HireAbility Resume Parsing and Job Parsing Solutions*. URL: <https://www.hireability.com/products/alex-cv-resume-parser/>.
5. Kurama V. How to OCR Resumes using Intelligent Automation. Nanonets. URL: <https://nanonets.com/blog/ocr-for-resume-parsing-deep-learning/>.

Олійник Нікіта Юрійович — студент групи 1KN-22м, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, email: delmark1904@gmail.com.

Колесницький Олег Костянтинович — доцент кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця.

Nikita Oliinyk Y. — student of group 1KN-22m, faculty of intellectual information technologies and automation, Vinnytsia National Technical University, Vinnytsia, email: delmark1904@gmail.com.

Kolesnytskyj Oleg K. — docent of the Computer Sciences Dpt., Vinnytsia National Technical University, Vinnytsia