Программа
профессиональной переподготовки

Data Science
(DS-16)

Москва
2025

ИТОГОВЫЙ ПРОЕКТ слушателя программы «Специалист по Data Science»

Утвержденная тема проекта: Анализ вакансий с портала хэдхантер (NLP)

Научный руководитель: Старший преподаватель, академический руководитель магистратуры Аналитика больших данных, НИУ ВШЭ Паточенко Е.А

Ценность задачи

Сегодня многие крупные компании активно внедряют искусственный интеллект (ИИ) в свои бизнес-процессы, и рекрутмент не является исключением. Например, алгоритмы обработки текстов на естественных языках (Natural Language Processing, NLP) помогают рекрутерам значительно экономить время на отборе кандидатов. Это, в свою очередь, создает дополнительный барьер для соискателей на первом этапе отбора резюме.

Создание системы автоматического сопоставления и анализа резюме с доступными вакансиями, поможет соискателям быстрее адаптироваться под требования компании.

Постановка задачи

1

Обеспечить доступ к данным с HH.ru

2

Создать простую языковую модель для оценки семантического сходства описаний резюме и вакансий

3

Сопоставления резюме с вакансиями на основе данных HH.ru с использованием технологий машинного обучения и обработки естественного языка

HeadHunter Semantic Similarity Analysis for Resume (HH SSAR)
Система смыслового сопоставления резюме и вакансий

Что умеет система:

1. Сбор данных с HH.ru через API и веб-скрейпинг
2. Обучение Word2Vec на русскоязычном корпусе Wikipedia
3. Интеллектуальное сопоставление с учетом семантики
4. Детальная аналитика по каждому критерию

Гибридный подход сопоставления:

1. Семантический анализ (Word2Vec)
2. Многокритериальный анализ

Извлечение и подготовка данных резюме и вакансий (HH.py)

Модуль HH.py отвечает за сбор и обработку данных с сайта HH.ru

Основные функции модуля:

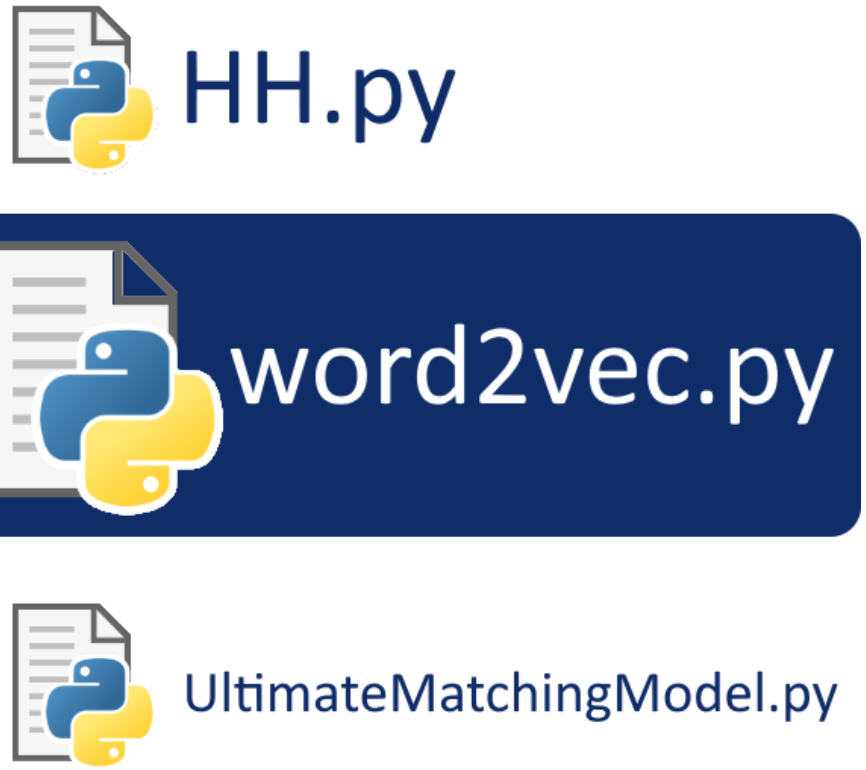
- Создает два класса для Резюме и Вакансии;
- Извлекает данные о вакансиях с использованием API;
- Извлекает данные о резюме методом парсинга HTML страницы;
- Подготавливает данные для дальнейшей обработки.

Обучение модели на основе искусственных нейронных сетей (word2vec.py)

Модуль word2vec.py отвечает за создание и обучение модели Word2Vec на русскоязычном корпусе Wikipedia для получения векторных представлений слов.

Основные функции модуля:

- Скачивает дампы статей с Википедии;
- Извлекает текст из дампов и обрабатывает;
- Обучает модель Word2Vec с заданными параметрами.

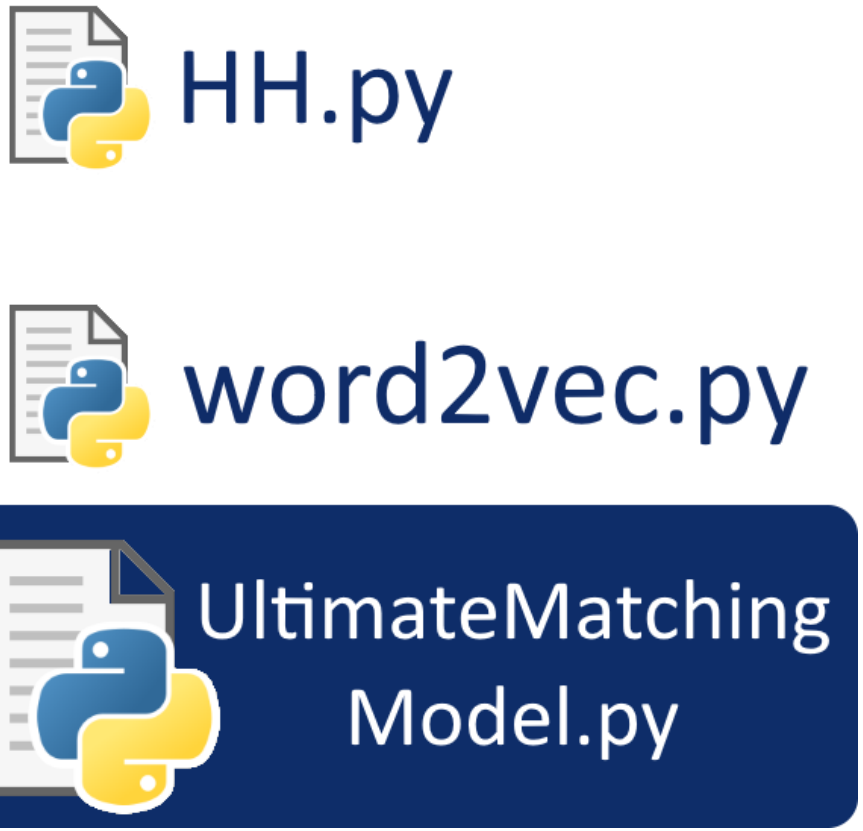


Сопоставления резюме с вакансиями (UltimateMatchingModel.py)

UltimateMatchingModel.py - это основной модуль для сопоставления резюме с вакансиями, которое объединяет семантический анализ и многокритериальную оценку для получения итогового рейтинга соответствия.

Основные функции модуля:

- Выполняет два алгоритма сопоставления;
- Выдает результат анализа.



FRONTEND

Главное меню

HH
Word2Vec
UltimateMatcher
Выход

OUTPUT

```
РЕЗУЛЬТАТЫ СОПОСТАВЛЕНИЯ
==========================
Обработано резюме: 5
Всего найдено совпадений: 23
==========================
📄 РЕЗЮМЕ: Python разработчик (Data Science)
  ID на HeadHunter: e0030b08ff0ccd25890039ed1f4d706b6f636e
  Ссылка: https://hh.ru/resume/e0030b08ff0ccd25890039ed1f4d706b6f636e
  Найдено совпадений: 8
  ----------------------------------------------------------------------

  1. 💼 Python разработчик (Machine Learning)
     🏢 Компания: Яндекс
     📍 Регион: Москва
     📊 Сходство: 87.3% (0.9) ( 🟢 ОТЛИЧНОЕ)
     🎯 Опыт: От 3 до 6 лет
     🔗 Ссылка: https://hh.ru/vacancy/12345678
     📝 Компоненты совпадения: - semantic: 0.85 - multi_criteria: 0.89
```

BACKEND

HH.py

вакансии | резюме

word2vec.py

gensim.model

UltimateMatchingModel.py

Семантическое сходство

Многокритериальная оценка

КОНСОЛИДАЦИЯ ДАННЫХ

ИТОГОВАЯ ОЦЕНКА С УЧЕТОМ ВЕСОВ

= cosine_similarity(resume, vacancy)*weight +
sum(criteria_scores)*weight

Планируемые обновления HeadHunter Semantic Similarity Analysis for Resume (HH SSAR)

1

Telegram Bot для быстрого доступа

2

Добавление функции адаптивного обучения системы для обновления весов

3

Добавление функции машинного обучение для предсказания релевантности и улучшения точности


ВЫСШАЯ • ШКОЛА • ЭКОНОМИКИ •