

THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):
(https://youtu.be/JWrU8_GirTA)
- Link slides (dạng .pdf đặt trên Github):
(<https://github.com/tringuyen2020/CS2205.MAR2024.git>)
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*

<ul style="list-style-type: none">● Họ và Tên: Nguyễn Bảo Trí● MSSV: 230201059 	<ul style="list-style-type: none">● Lớp: CS2205.MAR2024● Tự đánh giá (điểm tổng kết môn): 7/10● Số buổi vắng: 3● Số câu hỏi QT cá nhân: 0● Link Github: https://github.com/tringuyen2020/CS2205.MAR2024.git
--	---

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

XỬ LÝ CV BẰNG KỸ THUẬT XỬ LÝ NGÔN NGỮ TỰ NHIÊN (NLP)

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

CV PROCESSING BY NATURAL LANGUAGE PROCESSING (NLP)
TECHNIQUE

TÓM TẮT (Tối đa 400 từ)

Bản mô tả công việc (Job Description hay JD) được coi là một hoạt động chính của bộ phận HR và là một nhiệm vụ rất khó để tìm được nhân tài có thể làm việc hiệu quả. Các nhà tuyển dụng cần có chi phí tài chính, thời gian và công sức lớn để chọn ra những ứng viên tốt nhất.

Tương tự, người tìm việc khó có được thông tin về việc tuyển dụng hay yêu cầu của công ty. Mô hình này đề xuất trích xuất thông tin từ CV và xếp hạng theo sở thích của các công ty liên kết bằng kỹ thuật Xử lý ngôn ngữ tự nhiên (Natural Language Processing hay NLP). Nó làm cho quá trình tuyển dụng trở nên dễ dàng và hiệu quả. Một CV chứa nhiều khối thông tin khác nhau bên trong và bất kỳ trình phân tích cú pháp nào cũng cần rút ra những khối thông tin này như là trình độ học vấn, kinh nghiệm, dự án, địa chỉ, thông tin cá nhân...

Do đó, về cơ bản, chúng tôi sẽ xây dựng một cổng thông tin việc làm trong đó người dùng hoặc nhân viên hoặc người tìm việc sẽ tải lên CV của họ và sử dụng kỹ thuật NLP, trình phân tích cú pháp sẽ chuyển thông tin cần thiết và tạo thông tin CV có cấu trúc theo nhu cầu và yêu cầu của công ty cũng sẽ xếp hạng CV tùy theo năng lực của người tìm việc. Cổng thông tin việc làm sẽ bao gồm ba mô-đun như Nhân viên, Quản trị viên và Công ty. Mô-đun Quản trị viên sẽ quản lý các mô-đun Công ty và Nhân viên.

GIỚI THIỆU (Tối đa 1 trang A4)

Các doanh nghiệp và cơ quan tuyển dụng xử lý rất nhiều hồ sơ hàng ngày. Do đó, cần có một hệ thống thông minh tự động có thể lấy ra tất cả thông tin quan trọng từ các CV không có cấu trúc và chuyển đổi tất cả chúng sang định dạng có cấu trúc chung, sau đó có thể được xếp hạng cho một vị trí công việc cụ thể. Chúng tôi đã xem xét nhiều tài liệu nghiên cứu và chúng tôi đã thu hẹp lại sáu tài liệu nghiên cứu được đề cập dưới đây.

STT	Tác giả và Năm phát hành	Tiêu đề của tài liệu	Quan sát và Nhận xét
1	Nirali Bhaliya, Jay Gandhi, Dheeraj Kumar Singh 2020,IJITEE	Trích xuất CV có liên quan dựa trên NLP bằng Machine Learning	<ul style="list-style-type: none">● Phân tích cú pháp với giới hạn chỉ định● Chỉ số tương đồng về bộ kỹ năng
2	Dr.Parkavi A,Pooja Pandey,Poornima J,Vaibhavi G S Kaveri B W 2019, IJARBEST	Hệ thống tuyển dụng điện tử thông qua phân tích CV, kiểm tra tâm lý và phân tích mạng xã hội	<ul style="list-style-type: none">● Khai thác văn bản được sử dụng để tạo điểm số● Quét web
3	Ayishathahira C H,Sreejith C,Raseek C 2018, International CET Conference on Control, Communication, and Computing (IC4)	Sự kết hợp giữa mạng thần kinh và trường ngẫu nhiên có điều kiện để phân tích cú pháp CV hiệu quả	<ul style="list-style-type: none">● Phân loại CV thành 3 đoạn● Trích xuất 23 trường dữ liệu khác nhau
4	Papiya Das,Manjusha Pandey and Siddharth Swarup Rautaray 2018,IJITCS	Mô hình phân tích cú pháp CV sử dụng Quy trình trích xuất thực thể và Công cụ dữ liệu lớn	<ul style="list-style-type: none">● Chuyển đổi CV không có cấu trúc thành CV có cấu trúc● Hỗ trợ nhiều ngôn ngữ● Ánh xạ ngữ nghĩa cho kỹ năng, bảng việc làm, nhà tuyển dụng

5	Satyaki Sanyal, Neelanjana Ghosh, Souvik Hazra, Soumyashree Adhikary 2017, IJESC	Lọc từ khóa chuyển đổi các định dạng khác nhau của CV thành văn bản	<ul style="list-style-type: none"> ● Lọc từ khóa ● Chuyển đổi các định dạng khác nhau của CV thành văn bản
6	Vinaya R. Kudatarkar, Manjula Ramannavar, Dr. Nandini S. Sidnal 2015 IJIRAE	Phương pháp phân tích văn bản phi cấu trúc để đánh giá định tính sơ yếu lý lịch	<ul style="list-style-type: none"> ● Phân tích văn bản đơn giản ● Đánh giá chất lượng toàn diện ● Thước đo chất lượng cho sơ yếu lý lịch

MỤC TIÊU

- Đơn giản hóa quy trình tuyển dụng bằng cách sử dụng Optical Character Recognition (OCR) và NLP để xử lý CV tự động nhằm cung cấp những ứng viên chất lượng cho các công ty.
- Giảm thiểu sự bất công và phân biệt đối xử trong quá trình lọc CV.
- Các CV sẽ được phân loại và xếp hạng dựa trên dữ liệu từ các mạng xã hội như LinkedIn.

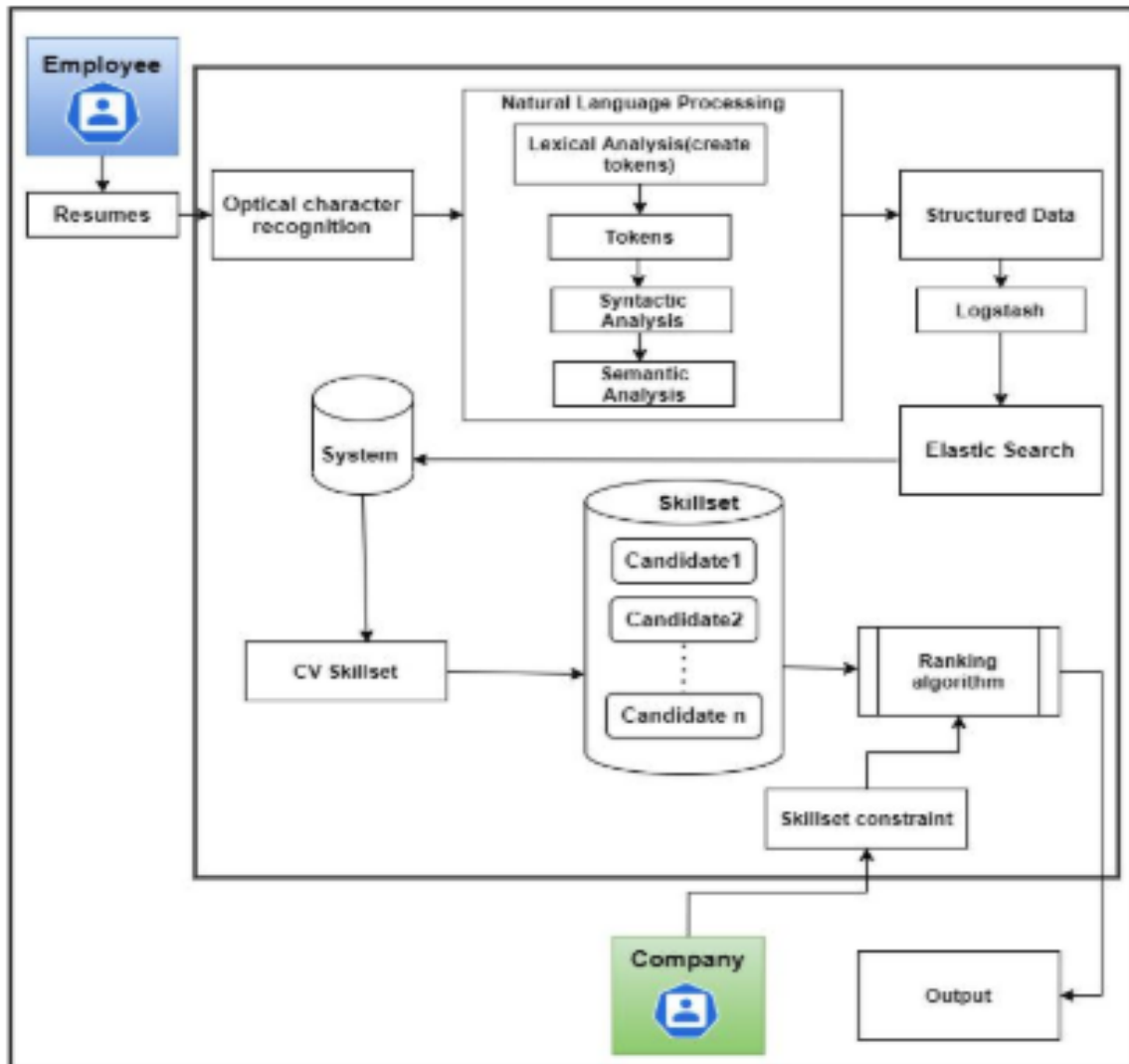
NỘI DUNG VÀ PHƯƠNG PHÁP

Nội dung:

Trong đề xuất này, chúng tôi đang sử dụng kỹ thuật NLP để phân tích CV theo các công ty cụ thể. Một cổng thông tin việc làm chung cho người sử dụng lao động cũng như nhân viên được cung cấp để đăng ký và tạo công việc. Các CV nhận được sẽ được phân tích cú pháp và xếp hạng theo yêu cầu của công ty.

Ngoài ra, mục tiêu khác của chúng tôi là trích xuất dữ liệu từ mạng xã hội như LinkedIn để ứng tuyển, điều này sẽ giúp quá trình tuyển dụng dễ dàng nhận được các đơn đăng ký chất lượng từ nhiều khu vực khác nhau bằng cách tránh bất công và phân biệt đối xử.

Kiến trúc hệ thống:



Phương pháp:

Để thực hiện phân tích cú pháp CV, cả nhân viên và công ty đều cần cung cấp CV chi tiết của nhân viên và các ràng buộc về kỹ năng yêu cầu công việc của công ty.

Hoạt động của kiến trúc hệ thống như sau:

1. Cung cấp web portal cho bộ phận HR để xác định các hạn chế và bộ kỹ năng cần thiết của công ty, dựa vào đó các ứng viên sẽ được đánh giá.
2. Ứng viên cũng cần sử dụng web portal để tạo tài khoản và tải CV lên. CV do ứng viên tải lên sẽ được nạp và đưa vào OCR.
3. Vì CV được tải lên có thể ở bất kỳ định dạng nào, chẳng hạn như '.txt', '.pdf', '.doc',

'.docx', '.odt'...Chúng tôi sẽ sử dụng OCR để chuyển đổi CV thành 1 dạng văn bản đơn giản.

4. CV được chuyển đổi sau đó được đưa vào mô-đun NLP, nó lấy văn bản đơn giản làm đầu vào và chuyển đổi nó thành dữ liệu có ý nghĩa. Sử dụng NLP, chúng tôi sẽ phân tích CV, NLP yêu cầu những điều sau để phân tích cú pháp:

- . Phân tích từ vựng: Phân tích từ vựng là giai đoạn đầu tiên của phân tích cú pháp NLP, đầu vào văn bản đơn giản được phân đoạn thành các từ và đoạn văn, sau đó các mã thông báo (token) sẽ được tạo.

- . Phân tích cú pháp: Trong phân tích cú pháp, việc phân tích ngữ pháp và sắp xếp từ ngữ một cách có ý nghĩa được kiểm tra, những câu vô nghĩa sẽ bị loại bỏ.

- . Phân tích ngữ nghĩa: Phân tích ngữ nghĩa kiểm tra ý nghĩa chính xác của văn bản.

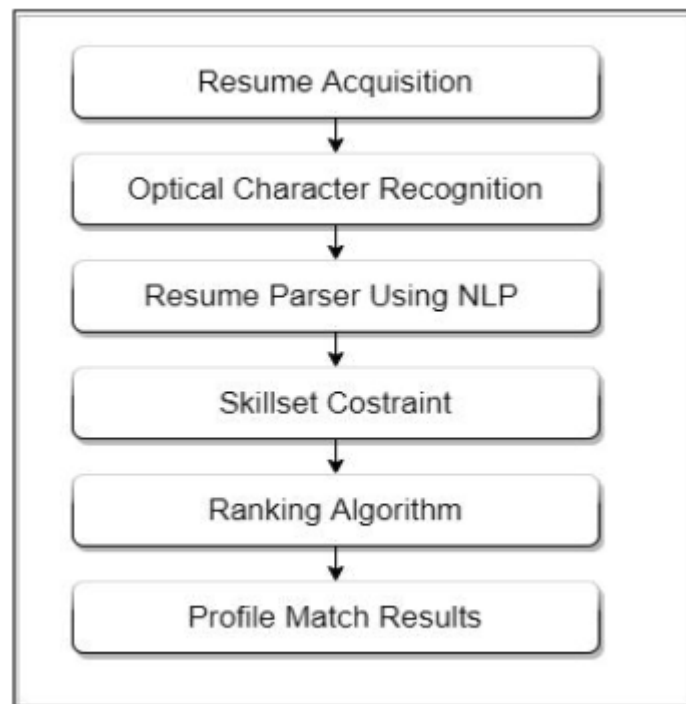
- . Nhận dạng thực thể được đặt tên (Named Entity Recognition hay NER): Một trong những vấn đề khi sử dụng cùng một mô-đun NLP cho tất cả các công ty là các thuật ngữ và từ có ý nghĩa gì đó đối với miền của công ty đó và có thể có ý nghĩa khác nói chung. Trở ngại này được khắc phục trong hệ thống của chúng tôi với sự trợ giúp của NER. Một thực thể được đặt tên là một đối tượng tồn tại trong từ thực. Với NER, chúng tôi có thể tinh chỉnh mô-đun NLP của mình để hiểu các đối tượng từ thực từ một miền. Ví dụ, nếu một công ty muốn thuê lập trình viên, họ có thể sử dụng hệ thống của chúng tôi để phân biệt giữa những người yêu thích “Python”, ngôn ngữ lập trình và những người yêu thích “python”, con rắn, dựa trên ngữ cảnh của từ đó đã sử dụng.

5. Dữ liệu thu được sẽ được sử dụng cho một bảng thông tin (dashboard) chứa các biểu đồ dựa trên dữ liệu trong CV. HR có thể sử dụng bảng thông tin này để chuẩn bị các truy vấn của mình dựa trên yêu cầu tương ứng mà công ty đưa ra. Để tạo trang tổng quan này, tất cả dữ liệu sẽ được đưa vào Elasticsearch. Để làm cho hệ thống không phụ thuộc vào nhiều định dạng, dữ liệu sẽ được cung cấp bằng cách sử dụng LogStash.

6. Sử dụng các truy vấn có sẵn trong Elasticsearch, các CV sẽ được tính điểm và sau đó chúng sẽ được sắp xếp theo các ràng buộc do bộ phận HR cung cấp.

7. Các đặc điểm cá nhân của ứng viên sẽ được tăng cường tương ứng dựa trên mức độ ưu tiên của đặc điểm đó.

8. Cuối cùng, điểm tích lũy sẽ được sử dụng để phân loại ứng viên. Danh sách ứng viên được sắp xếp cuối cùng sẽ được hiển thị cho bộ phận HR.



KẾT QUẢ MONG ĐỢI

Việc ứng dụng NLP vào việc xử lý CV mang lại các kết quả mong đợi như sau:

- Việc xử lý CV trở nên tự động, nhanh chóng, và hiệu quả.
- Tiết kiệm chi phí trong việc sàng lọc CV.
- Giảm thiểu sai sót, bất công và phân biệt gây ra bởi người tuyển dụng.

TÀI LIỆU THAM KHẢO (Định dạng DBLP)

- [1]. Vinaya R. Kudatarkar, ManjulaRamannavar, Dr.Nandini S. Sidnal “An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes”, 2015, IJIRAE.
- [2]. Satyaki Sanyal, Neelanjan Ghosh, SouvikHazra, Soumyashree Adhikary,

“Resume Parser with Natural language Processing”, 2007,IJESC.

[3]. Papiya Das, Manjusha Pandey, Siddharth Swarup Rautaray, “A CV parser Model using Entity Extraction Process and Big Data Tools ”, 2018,IJTCS.

[4]. Ayishathahira and Sreejith , “Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing”,International CET Conference on Control, Communication and Computing(IC4),2018..

[5]. Dr.Parkavi A,Pooja Pandey,Poornima J,Vaibhavi G S,Kaveri BW, “E-Recruitment System Through Resume Parsing,Psychometric Test and Social Media Analysis”, 2019,IJARBEST.

[6]. Nirali Bhaliya, Jay Gandhi, Dheeraj Kumar Singh, “NLP based Extraction of Relevant Resume using Machine Learning ”,2020,IJITEE.