

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/357150947>

# PHÁT TRIỂN HỆ KHUYẾN NGHỊ VIỆC LÀM CHO SINH VIÊN CÔNG NGHỆ THÔNG TIN THEO NĂNG LỰC HỌC TẬP ĐÁP ỨNG NHU CẦU DOANH NGHIỆP

Conference Paper · December 2021

DOI: 10.15625/vap.2021.0072

CITATIONS

0

READS

1,202

1 author:



Trương Ho-Viet Phan

Van Lang University

8 PUBLICATIONS 46 CITATIONS

SEE PROFILE

# PHÁT TRIỂN HỆ KHUYẾN NGHỊ VIỆC LÀM CHO SINH VIÊN CÔNG NGHỆ THÔNG TIN THEO NĂNG LỰC HỌC TẬP ĐÁP ỨNG NHU CẦU DOANH NGHIỆP

Nguyễn Thái Anh, Phan Hồ Viết Trường

Khoa Công nghệ thông tin, Đại học Văn Lang, TP. Hồ Chí Minh  
anh.187it14044@vanlanguni.vn, truong.phv@vlu.edu.vn

**TÓM TẮT:** Hệ khuyến nghị là một hệ thống gợi ý sản phẩm, dịch vụ cho người dùng dựa vào lịch sử mua hàng của họ. Hệ khuyến nghị được ứng dụng trong nhiều lĩnh vực khác nhau như du lịch, giáo dục, tin tức, âm nhạc... Tuy nhiên, rất ít nghiên cứu dùng hệ khuyến nghị giới thiệu việc làm cho sinh viên khi chưa tốt nghiệp. Bài báo này trình bày một giải pháp xây dựng hệ khuyến nghị việc làm cho sinh viên ngành Công nghệ thông tin đang theo học các chuyên ngành Khoa học dữ liệu, Kỹ thuật phần mềm và Mạng máy tính tại Trường Đại học Văn Lang. Nghiên cứu này sử dụng kỹ thuật lọc nội dung để đo lường độ tương đồng giữa các môn học mà sinh viên đã đạt điểm cao với nhu cầu của doanh nghiệp công nghệ thông tin. Sau đó, hệ thống sẽ chọn 5 công ty phù hợp nhất để gửi thông tin tuyển dụng đến hộp thư cá nhân của sinh viên. Dữ liệu thử nghiệm được thu thập từ bảng điểm của khoảng 500 sinh viên năm ba và năm cuối đang theo học các chuyên ngành trên tại Đại học Văn Lang. Kết quả thử nghiệm đánh giá giải pháp đề xuất đạt hiệu quả tốt hơn so với các độ đo phổ biến như Jaccard, Cosine, TF-IDF + cosine, Word2Vec + cosine. Giải pháp này có thể được áp dụng cho các khoa khác tại Trường Đại học Văn Lang.

**Từ khóa:** Hệ khuyến nghị, độ đo tương đồng, rút trích thông tin, hệ khuyến nghị việc làm.

## I. GIỚI THIỆU

Hệ khuyến nghị là một hệ thống gợi ý những sản phẩm, dịch vụ phù hợp với nhu cầu của người dùng. Điều này giúp cho người dùng có nhiều lựa chọn hơn trong việc quyết định sản phẩm cần mua. Đối với doanh nghiệp, họ có cơ hội quảng bá nhiều sản phẩm, dịch vụ đến người dùng nhằm tăng doanh số bán hàng. Hệ khuyến nghị được ứng dụng trong các lĩnh vực khác nhau như âm nhạc, tin tức, giáo dục, thương mại điện tử. Trong tin tức, hệ khuyến nghị giúp người dùng có thể đọc những tin tức có liên quan đến bản tin mà người dùng quan tâm [1]. Trong âm nhạc, người dùng có thể được gợi ý những bài hát mà họ có thể nghe dựa vào phản hồi tiềm ẩn [2]. Trong giáo dục, hệ khuyến nghị dùng để dự báo kết quả học tập, lựa chọn môn học, tra cứu tài liệu [3]. Trong thương mại điện tử, người dùng được gợi ý mua sản phẩm dựa theo số lần duyệt, số lần mua hàng [4]. Trong du lịch, người dùng được gợi ý các địa điểm du lịch phù hợp theo ngữ cảnh [5].

Tuy nhiên, có ít nghiên cứu sử dụng hệ khuyến nghị cho tác vụ giới thiệu việc làm. Các sinh viên khi đăng ký theo học một ngành đều mong muốn sau khi tốt nghiệp sẽ tìm được một công việc phù hợp với chuyên môn đã chọn. Tuy nhiên, nếu đợi sau khi tốt nghiệp rồi mới tìm việc làm thì sinh viên khó đáp ứng được một số yêu cầu từ phía doanh nghiệp. Ví dụ, doanh nghiệp cần tuyển dụng vị trí lập trình viên với yêu cầu ứng viên có kinh nghiệm sử dụng HBase, Redis, Distributed Computing, trong khi sinh viên chỉ được học các môn như MS SQL Server, C#, Java, Python, lập trình Web, lập trình di động. Điều này dẫn đến sinh viên gặp khó khăn khi nộp hồ sơ xin việc. Hệ thống giới thiệu việc làm cho sinh viên từ năm thứ 2 tuy chưa thể giúp các em đi làm ngay được nhưng sẽ giúp các em nhận thấy những kỹ năng chuyên môn còn thiếu sót để có sự chuẩn bị trước khi các em ra trường. Đối với những em năm cuối, hệ thống gợi ý những công ty có yêu cầu phù hợp với những kiến thức các em đã được đào tạo.

Bài báo này đề xuất một giải pháp xây dựng hệ khuyến nghị việc làm cho sinh viên năm 3 và năm cuối theo năng lực học tập của mỗi cá nhân. Cụ thể, bài viết sử dụng phương pháp lọc nội dung để đo độ tương đồng giữa những môn học chuyên ngành mà các em đã đạt điểm số cao với yêu cầu từ phía doanh nghiệp. Sau đó, hệ thống chọn 5 doanh nghiệp hàng có điểm số gần với năng lực học tập của mỗi em và gửi email thông báo cho các em biết khả năng của các em so với nhu cầu của doanh nghiệp. Hệ thống vận hành trên môi trường Web và tích hợp với phần mềm quản lý của phòng đào tạo nhằm mang lại một kênh thông tin việc làm bổ ích cho sinh viên từ khi còn ngồi trên ghế nhà trường.

## II. HỆ KHUYẾN NGHỊ VIỆC LÀM CHO SINH VIÊN CÔNG NGHỆ THÔNG TIN THEO NĂNG LỰC HỌC TẬP ĐÁP ỨNG NHU CẦU DOANH NGHIỆP

Hiện nay, hệ khuyến nghị có ba tiếp cận phổ biến: Lọc cộng tác, lọc nội dung và hệ lai ghép [6].

### A. Lọc cộng tác

Phương pháp lọc cộng tác dựa trên việc thu thập và phân tích một lượng lớn các hành vi, hoạt động của người dùng và dự đoán xem họ có nét tương đồng như thế nào đối với người khác [7], [8]. Tiếp cận này không sử dụng học máy nhưng có thể đề xuất chính xác những sản phẩm khác mà người dùng đang quan tâm. Ví dụ: một khách hàng khi mua kem đánh răng hiệu PS, hệ khuyến nghị sẽ đưa ra gợi ý thêm những hiệu khác như Colgate. Hệ thống này có ưu, khuyết điểm như sau:

1. *Ưu điểm. Đề xuất được những sản phẩm người dùng quan tâm. Người dùng đánh giá trực tiếp đối với các sản phẩm, dịch vụ, nội dung.*
2. *Nhược điểm. Ma trận đánh giá rất thưa. Khi ma trận đánh giá lớn, nghĩa là số người dùng lẫn sản phẩm lớn thì thời gian tính toán sẽ tăng cao, khó đáp ứng tư vấn thời gian thực hoặc gần thời gian thực. Các từ đồng nghĩa trong đánh giá sản phẩm gây cản trở cho việc tính toán độ tương tự. Đặc biệt, khi người dùng có đánh giá tương phản. Ví dụ như người dùng thích sản phẩm này nhưng lại đánh giá không thích nên không thể khuyến nghị chính xác cho những người này.*

**B. Lọc dựa trên nội dung**

Hệ thống này dựa trên việc rút nội dung mô tả mặt hàng và hồ sơ các lựa chọn của người dùng trong quá khứ [9]. Từ đó, hệ thống thực hiện so sánh sự tương đồng nội dung để chọn những điểm chung nhất phục vụ cho việc gợi ý. Ví dụ: Sử dụng phương pháp lọc nội dung trong giáo dục để chấm bài thi lập trình. Nếu 2 bài có độ tương đồng nhiều thì ta kết luận là sinh viên sao chép bài bạn khác.

1. *Ưu điểm. Đảm bảo tính độc lập giữa các người dùng và giải quyết phần lớn vấn đề khởi động nguội dành cho người dùng mới hoặc sản phẩm mới.*
2. *Nhược điểm. Hệ thống phải phân tích và trích chọn nội dung sản phẩm. Điều này dẫn đến hệ thống tư vấn các sản phẩm quen thuộc. Đồng thời, hệ thống gặp khó khăn khi gợi ý các sản phẩm mới cho người dùng mới.*

**C. Hệ thống gợi ý lai**

Là sự kết hợp cả hai phương pháp lọc dựa trên nội dung và lọc cộng tác. Phương pháp này nhằm giải quyết vấn đề khởi động lạnh và ma trận đánh giá thưa.

**D. Ứng dụng lọc nội dung cho hệ khuyến nghị việc làm cho sinh viên chưa tốt nghiệp**

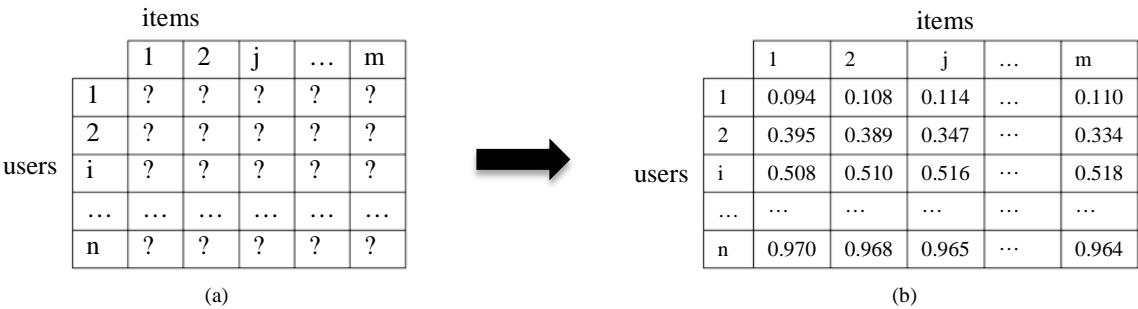
Bài báo này sử dụng tiếp cận lọc dựa trên nội dung để giải quyết bài toán khuyến nghị việc làm cho sinh viên chưa tốt nghiệp. Mô hình lọc nội dung cần 3 tập dữ liệu về users, items và feedback như trong Hình 1. Trong đó, users là danh sách người dùng, items là danh sách sản phẩm, đối tượng của hệ thống. Ví dụ như các bài viết trên trang viblo, các video trên youtube. Mỗi item có thể kèm theo thông tin mô tả. Tập feedback là lịch sử tương tác của user với mỗi item, có thể là đánh giá của mỗi user với một item, số ratings, hoặc comment, việc user click, view hoặc mua sản phẩm. Từ mô hình lọc nội dung, bài báo này ứng dụng vào hệ khuyến nghị việc làm gồm các tập dữ liệu users, items, và feedback. Tập user chứa danh sách sinh viên, tập items chứa nội dung các tin tuyển dụng việc làm và tập feedback chứa mô tả các học phần so với nội dung yêu cầu trong tập items.

**1. Xây dựng hồ sơ items**

Hồ sơ items là những mô tả nhu cầu tuyển dụng của doanh nghiệp. Hồ sơ này được biểu diễn dưới dạng toán học là một vector đặc trưng  $n$  chiều. Từ đó, hệ thống có thể xác định các item có nội dung tương tự bằng cách tính độ tương đồng giữa các feature vector của chúng. Một số phương pháp thường được sử dụng để xây dựng vector đặc trưng là biểu diễn nhị phân, TF-IDF và nhúng từ. Trong bài báo này, nghiên cứu sử dụng kỹ thuật nhúng từ Fasttext để rút đặc trưng cho mô tả nhu cầu doanh nghiệp và các học phần của sinh viên.

**2. Mô hình biểu diễn của tập users**

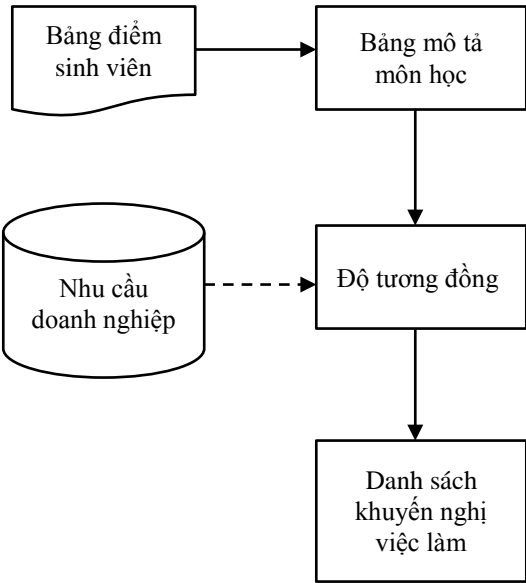
Gọi  $N$  là số sinh viên trong tập users,  $M$  là số mẫu tin tuyển dụng trong tập items,  $Y$  là ma trận user-item. Trong đó,  $y(i, j)$  là mức độ quan tâm của sinh viên thứ  $i$  với việc làm thứ  $j$  như Hình 1. Mức độ quan tâm của sinh viên được hệ thống tính tự động giữa các môn học với nhu cầu tuyển dụng. Nghiên cứu này sẽ giải quyết bài toán tính mức độ quan tâm  $y(i, j)$ .



Hình 1. Ma trận user-item biểu diễn mối quan tâm của sinh viên với việc làm. Hình (a) ma trận user-item khi chưa tính mức độ quan tâm. Hình (b) biểu diễn ma trận user-item đã được tính mức độ quan tâm sinh viên với việc làm

Nhóm sẽ thu thập nội dung của của các môn học chuyên ngành (Mô tả nội dung của học phần) và nội dung của trang tìm kiếm việc làm trong lĩnh vực công nghệ thông tin để tìm sự tương đồng rồi sau đó gợi ý ra danh sách các công ty đang tuyển dụng cho các sinh viên. Ngoài ra, sinh viên có thể sử dụng hệ thống này để biết những kiến thức bản thân còn thiếu so với nhu cầu doanh nghiệp để sinh viên kịp thời bổ sung trước khi ra trường.

Hệ thống tổng quát của bài báo được trình bày như trong Hình 2. Đầu tiên, hệ thống cần bảng điểm toàn khóa của sinh viên. Tiếp theo, hệ thống chọn những môn chuyên ngành có điểm thỏa một ngưỡng cho trước để từ đó rút ra nội dung của những môn học này. Sau đó, hệ thống sẽ đo độ tương đồng với những mô tả yêu cầu của doanh nghiệp đã được lưu trữ trong cơ sở dữ liệu. Cuối cùng, từ điểm số tương đồng tính được, hệ thống sẽ gợi ý những doanh nghiệp phù hợp với năng lực học tập của sinh viên. Các thành phần trong Hình 1 sẽ được trình bày chi tiết trong phần sau.

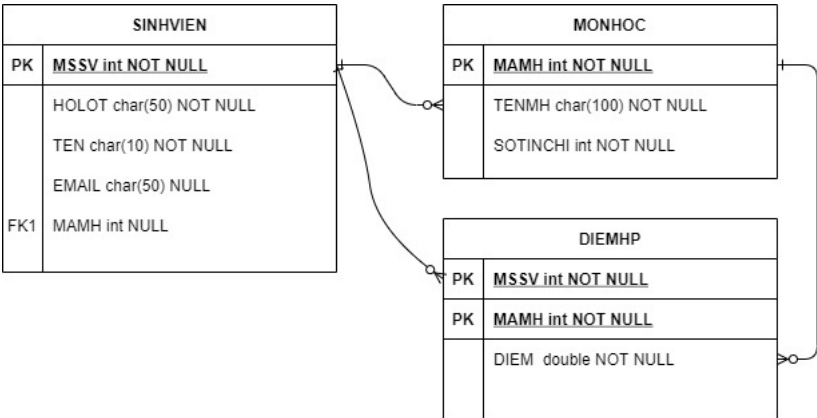


Hình 2. Tổng quát hệ thống khuyến nghị việc làm cho sinh viên công nghệ thông tin

E. Bảng điểm sinh viên

Bảng điểm dùng để đánh giá quá trình và năng lực học tập của sinh viên. Hệ thống sử dụng bảng điểm của sinh viên để làm tiêu chí so sánh với nhu cầu của doanh nghiệp. Hệ thống thu thập khoảng 500 bảng điểm của sinh viên năm hai và năm ba đang theo học các chuyên ngành Khoa học dữ liệu, Kỹ thuật phần mềm và Mạng máy tính tại Trường Đại học Văn Lang. Những bảng điểm này được lưu vào cơ sở dữ liệu. Do bảng điểm có dạng tập tin PDF nên hệ thống phải thuê các bạn sinh viên nhập điểm thủ công. Trong cơ sở dữ liệu, hệ thống tạo 3 bảng để mô tả thông tin và quá trình học tập của mỗi sinh viên như trong Hình 3. Các bảng được mô tả như sau:

- **SINHVIEN**: Bảng này mô tả thông tin sinh viên gồm mã số sinh viên, họ, tên, email, mã môn học.
- **MONHOC**: Bảng này mô tả thông tin môn học gồm mã môn học, tên môn học, số tín chỉ.
- **DIEMHP**: Bảng này mô tả thông tin điểm học phần gồm mã sinh viên, mã môn học, điểm trung bình học phần.



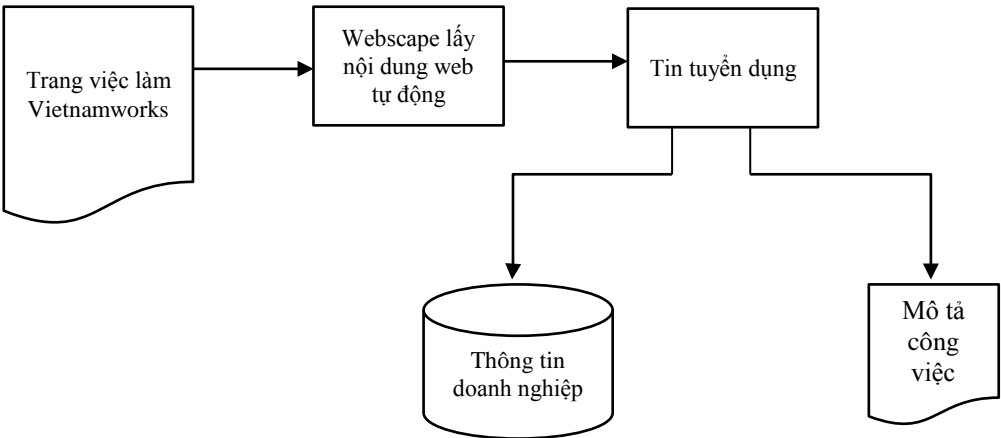
Hình 3. Lược đồ quan hệ giữa bảng SINHVIEN, MONHOC và DIEMHP

F. Bảng mô tả môn học

Bảng mô tả môn học được lấy từ đề cương chi tiết của mỗi môn học. Trong đề cương chi tiết cung cấp các thành phần mà hệ thống quan tâm đó là mục tiêu học phần, kiến thức, kỹ năng và mô tả vắn tắt nội dung học phần. Đây là những nội dung mà sinh viên đạt được. Điểm số là kết quả đánh giá hữu hiệu nhất mức độ lĩnh hội của sinh viên. Hệ thống chỉ thu thập tất cả 65 môn chuyên ngành Khoa học dữ liệu, Kỹ thuật phần mềm và Mạng máy tính. Các nội dung mô tả môn học được lưu dưới dạng tập tin văn bản. Khi đó độ tương đồng với nhu cầu doanh nghiệp, hệ thống sẽ lọc những môn mà sinh viên đạt một ngưỡng điểm cho trước. Mặc định, hệ thống thiết lập sẵn điểm 7.0 là ngưỡng ban đầu. Tuy nhiên, người dùng có thể điều chỉnh mức điểm cho phù hợp nhu cầu các nhân. Sau khi đã chọn các môn chuyên ngành thỏa ngưỡng điểm, hệ thống sẽ nối mô tả của những môn này lại để tạo thành một văn bản. Văn bản này sẽ được so sánh độ tương đồng với nhu cầu doanh nghiệp.

G. Nhu cầu doanh nghiệp

Một mẫu tin tuyển dụng thường gồm thông tin công ty, mô tả công việc, yêu cầu công việc, mức lương, phúc lợi... Trong đó, mô tả và yêu cầu công việc là nội dung mà nghiên cứu này quan tâm vì chúng cho biết nhu cầu về chuyên môn của doanh nghiệp đối với các ứng viên. Hệ thống thu thập khoảng 1.000 công việc về công nghệ thông tin từ trang web Vietnamworks. Vietnamworks là một trang web chuyên về tuyển dụng việc làm. Trang này cũng cho phép người dùng tham khảo thông tin công ty tuyển dụng và mô tả chi tiết về công việc. Hệ thống sử dụng mã nguồn mở Webscape để tự động lấy thông tin tuyển dụng của doanh nghiệp. Sau đó, phần thông tin sẽ được lưu vào cơ sở dữ liệu, còn phần mô tả và yêu cầu công việc sẽ được lưu thành tập tin văn bản. Phần thông tin được hiển thị trong danh sách khuyến nghị việc làm cho sinh viên, phần mô tả được dùng để so sánh với các môn học mà sinh viên đã đạt. Quá trình thu thập thông tin doanh nghiệp được thể hiện trong Hình 4.



Hình 4. Minh họa quá trình thu thập thông tin và nhu cầu tuyển dụng của doanh nghiệp

H. Độ đo tương đồng

Độ tương đồng văn bản là xác định sự giống nhau giữa hai văn bản về mặt từ vựng và ngữ nghĩa [10]. Bài báo này dùng phép nhúng từ để biểu diễn mô tả môn học và việc làm. Sau đó, hệ thống dùng các độ đo phổ biến để tính điểm số.

Phép nhúng từ là phương pháp ánh xạ những từ ngữ vào các vector hoặc số thực, còn được gọi là phương pháp mô hình hóa ngôn ngữ, dữ liệu theo đặc trưng trong xử lý ngôn ngữ tự nhiên. Phép nhúng từ là công cụ đóng vai trò quan trọng trong nhiều thuật toán, Machine Learning, Deep Learning trong việc xử lý Input ở dạng câu chữ, máy tính chỉ có thể hiểu được Input ở dạng là số nên phải chuyển đổi câu chữ thành các vector, ma trận. từ đó mới thực hiện các công việc phân loại, hồi quy. Ví dụ, cho 2 câu “Hôm nay đi chơi” và “Hôm nay đi làm”, khi thực hiện nhúng từ thì “chơi” và “làm” sẽ có vị trí gần nhau trong không gian vector do chúng có vị trí giống nhau trong một câu. Có 2 loại phép nhúng từ phổ biến là nhúng sử dụng mạng neuron như Word2vec, GloVe, BERT, XLNet và nhúng từ sử dụng thống kê như BoW (Bag of Words), TF-IDF (Term Frequency, Inverse Document Frequency).

Phương pháp nhúng từ hiện này trở nên rất phổ biến khi các mô hình được huấn luyện sẵn trên tập dữ liệu rất lớn. Đại diện cho phép nhúng từ này là word2vec và fasttext.

1. Word2Vec

Do Mikalov đề xuất vào năm 2013 [11]. Mô hình này tính toán sự tương đồng nghĩa giữa các từ để dự đoán từ tiếp theo bằng cách đưa qua một mạng neuron trong phạm vi các từ xung quanh. Word2vec sử dụng 2 phương pháp phổ biến chính: Continuous Bag-of-Words (CBOW) và Skip-gram. CBOW lấy đầu vào là một hoặc nhiều từ context word và dự đoán từ đầu ra thông qua một tầng neuron đơn giản. Skip-gram sử dụng đầu vào là từ đích để dự đoán những từ gần nó. CBOW sẽ có thời gian huấn luyện nhanh hơn Skip-gram.

2. FastText

FastText là thư viện mã nguồn mở do Facebook tạo ra năm 2016 [12]. Mô hình này hỗ trợ việc huấn luyện phép nhúng từ và phân loại văn bản. Đây là một dạng mở rộng của Word2Vec, nhưng khác ở điểm nó tạo thêm các n-gram để học.

Ví dụ:

Chữ “<thanh>“, với n-gram=3 sẽ được tách thành [“<thanh>“, “<th”, “tha”, “han”, “anh”, “nh>“]

Sau khi được huấn luyện, vector nhúng từ của chữ “thanh” sẽ là tổng của các vector nhúng từ [“<thanh>“, “<th”, “tha”, “han”, “anh”, “nh>“], nhân với vector của “thanh”.

Bài báo này sử dụng phương pháp nhúng từ để biểu diễn văn bản là FastText do mô hình này hỗ trợ tiếng Việt.

1. Danh sách khuyến nghị việc làm doanh nghiệp

Sau khi đã thực hiện độ đo tương đồng giữa mô tả các môn học với nhu cầu công việc của doanh nghiệp. Hệ thống sẽ chọn 5 doanh nghiệp của điểm số cao nhất và hiển thị thông tin việc làm, mức lương cho sinh viên tham khảo hoặc lựa chọn nộp hồ sơ ứng tuyển. Đồng thời, hệ thống cũng gửi thông tin việc làm mà doanh nghiệp đang tuyển qua email của sinh viên.

III. THỰC NGHIỆM

Hệ thống được cài đặt trên máy tính có cấu hình CPU Intel i7 1.8GHz, RAM 12GB. Máy tính cài phần mềm MS SQL Server 2014, Sublime Text, python 3.9.0.

A. Dữ liệu

Nghiên cứu này thu thập bảng điểm của gần 500 sinh viên ngành Công nghệ thông tin gồm 3 chuyên ngành khoa học dữ liệu, kỹ thuật phần mềm và mạng máy tính. Ngoài ra, còn có 65 môn học chuyên ngành, gần 100 việc làm trong lĩnh vực công nghệ thông tin. Mỗi mẫu tin tuyển dụng được chia 2 thành phần: Nhu cầu doanh nghiệp và thông tin doanh nghiệp. Trong thành phần nhu cầu doanh nghiệp, hệ thống lưu phần mô tả, yêu cầu của nhà tuyển dụng. Trong thành phần thông tin doanh nghiệp, hệ thống lưu các thông tin như tên doanh nghiệp, địa chỉ, mức lương, vị trí việc làm. Nghiên cứu sử dụng mã nguồn mở Webscape để lấy các mẫu tin tuyển dụng từ website việc làm trong lĩnh vực công nghệ thông tin.

B. Phương pháp đánh giá

**Độ đo Jaccard:** Là phương pháp thống kê dùng để đo độ tương đồng của các tập mẫu khác nhau. Độ đo Jaccard được tính bằng cách đếm các từ giống nhau giữa 2 văn bản chia cho tổng số các từ duy nhất trong cả 2 văn bản [13].

**Độ đo Cosine:** Là phương pháp đo mức tương tự giữa hai vector không gian [14]. Công thức tính tích vô hướng của hai vector theo độ đo cosine như sau.

$$\vec{a} \cdot \vec{b} = |\vec{a}| \cdot |\vec{b}| \cos(\vec{a}, \vec{b})$$
 (1)

Ta sẽ rút ra được CT tính độ tương đồng của vector A và vector B

$$\cos(\vec{a}, \vec{b}) = \frac{a_1b_1+a_2b_2+a_3b_3}{\sqrt{a_1^2+a_2^2+a_3^2}\sqrt{b_1^2+b_2^2+b_3^2}}$$
 (2)

Cos góc alpha được gọi là cosine similarity, đó chính là độ tương đồng của vector A và B. Giá trị của cosine luôn nằm trong đoạn [-1, 1]. Nếu giá trị cosine bằng 1, nghĩa là hai vector trùng nhau. Ngược lại, giá trị cosine càng tiến dần về -1 thì A, B là khác nhau. Cosine được áp dụng rất phổ biến trong các bài toán khai phá văn bản, hệ thống gợi ý.

TF-IDF là một phiên bản phát triển từ BoW [15]. Phương pháp này giúp ta tìm ra được những từ hiếm cũng như sàng lọc ra những từ xuất hiện quá nhiều và ít mang nghĩa như là, mà, thì, ở . . . TF (term frequency) là tần suất xuất hiện của một từ trong dữ liệu văn bản. IDF là một hệ số giúp làm giảm trọng số của những từ hay xuất hiện trong dữ liệu văn bản. TF-IDF dùng để biểu diễn văn bản thành vector đặc trưng.

C. So sánh các độ đo phổ biến với độ đo nghiên cứu đề xuất có sử dụng word embedding

Phần này sẽ thực hiện so sánh độ tương đồng của phương pháp có sử dụng word embedding với các tiếp cận phổ biến đã nêu Jaccard, cosine, tf-idf + cosine, word2vec + cosine. Nghiên cứu đặt ngưỡng điểm đạt yêu cầu là 7.0 và tính độ tương đồng nhu cầu tuyển dụng của doanh nghiệp với các môn học chuyên ngành mà sinh viên đã vượt qua. Nghiên cứu sử dụng FastText để thực hiện chuyển mỗi từ trong nhu cầu của doanh nghiệp và các mô tả môn học thành các vector 300 chiều. Mô hình fasttext được huấn luyện sẵn có 2 dạng hỗ trợ tiếng Anh và đa ngôn ngữ. Hệ thống chọn dạng đa ngôn ngữ vì nghiên cứu này thực hiện trên tiếng Việt.

Nghiên cứu này chọn ngẫu nhiên 3 doanh nghiệp có nhu cầu tuyển dụng đại diện cho 3 chuyên ngành khoa học dữ liệu, kỹ thuật phần mềm và mạng máy tính và thực nghiệm như sau:

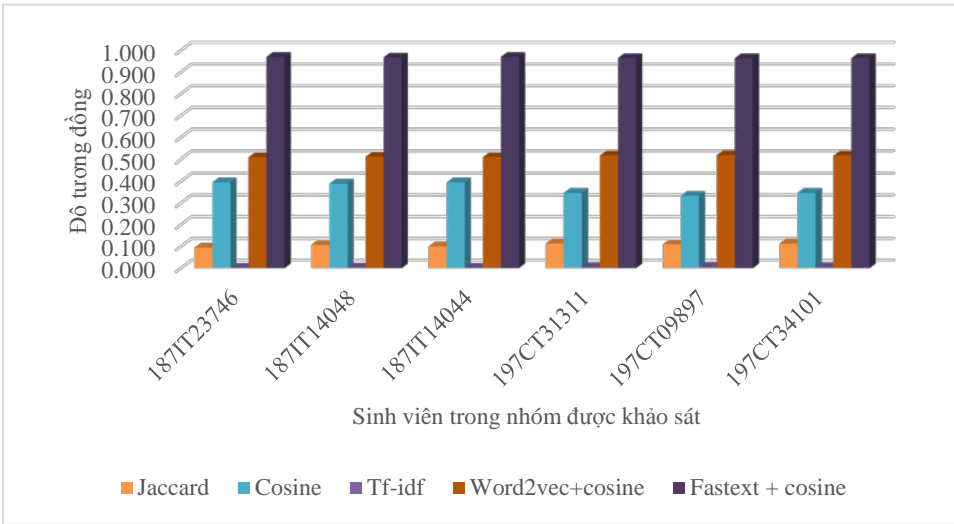
1. Thử nghiệm 1 - Doanh nghiệp 1 chuyên về công nghệ dữ liệu có nhu cầu tuyển dụng như sau:

- Yêu thích và đam mê nghiên cứu, xử lý dữ liệu tốt nghiệp cao đẳng/đại học chuyên ngành toán học, kinh tế, khoa học máy tính, quản trị thông tin hoặc thống kê.
- Tối thiểu 1,5 năm kinh nghiệm làm việc trong lĩnh vực phân tích/khoa học dữ liệu.
- Sử dụng tốt Power BI.
- Kỹ năng phân tích, tư duy logic, lý luận chặt chẽ.
- Kỹ năng tổ chức công việc, làm việc độc lập.
- Giao tiếp tự tin, nhanh nhẹn, có khả năng ngoại giao, đối ngoại.
- Cẩn thận, tỉ mỉ, khéo léo, ý thức mạnh mẽ về trách nhiệm.
- Tương tác với các bộ phận thường xuyên, chủ động.

Bảng 1 thể hiện kết quả đo mức độ năng lực học tập của 6 sinh viên ngẫu nhiên trong mẫu khảo sát phù hợp với nhu cầu tuyển dụng doanh nghiệp 1 chuyên về khoa học dữ liệu.

Bảng 1. So sánh độ tương đồng giữa năng lực học tập của sinh viên với nhu cầu tuyển dụng ngành khoa học dữ liệu

Độ đo	187IT23746	187IT14048	187IT14044	197CT31311	197CT09897	197CT34101
Jaccard	0,094	0,108	0,100	0,114	0,110	0,114
Cosine	0,395	0,389	0,395	0,347	0,334	0,347
Tf-idf +cosine	0,002	0,003	0,002	0,005	0,006	0,005
Word2vec+cosine	0,508	0,510	0,508	0,516	0,518	0,516
Fastext + cosine	0,970	0,968	0,970	0,965	0,964	0,965



Hình 5. Độ tương đồng môn học của sinh viên trong nhóm khảo sát với nhu cầu tuyển dụng việc làm về khoa học dữ liệu của doanh nghiệp 1

Trong Bảng 1 và Hình 5, xét điểm số dòng Fasttext + cosine, hai sinh viên có mã số 187IT23746 và 187IT14044 có mức độ phù hợp công việc mà nhà tuyển dụng đưa ra cao nhất đều đạt 0,97. Sinh viên có mã số 187IT14048 tuy học cùng khóa với 2 sinh viên trên nhưng mức độ phù hợp thấp hơn. Lý giải điều này như sau, do hiện nay Đại học Văn Lang đang đào tạo theo hệ thống tín chỉ. Sinh viên đăng ký môn học chuyên ngành yêu thích đầu mỗi học kỳ. Điều này dẫn đến sinh viên tuy cùng khóa nhưng số môn đạt được khác nhau trong đó có môn không nằm trong chuyên ngành Khoa học dữ liệu.

2. Thử nghiệm 2 - Nhu cầu doanh nghiệp 2 chuyên về kỹ thuật phần mềm như sau:

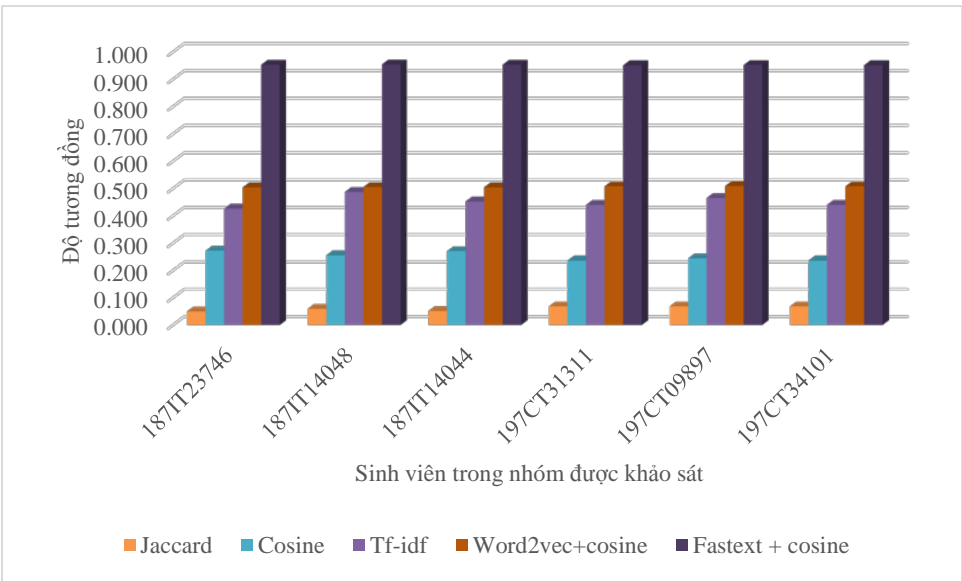
- Hơn 2 năm kinh nghiệm làm việc trong các công nghệ .NET / C # liên quan.
- Thực hành với ASP.NET MVC/Web Forms, Entity Framework hoặc NHibernate, SQL Server hoặc T-SQL, HTML/CSS, JavaScript OOP và jQuery.
- Kinh nghiệm là một lợi thế trong ASP.NET Core 3.1, Entity Framework Core, Microservice.
- Khả năng làm việc độc lập và làm việc nhóm tốt.

- Kỹ năng giao tiếp tiếng Anh tốt.
- Xử lý nhiều công việc.

Bảng 2 thể hiện kết quả đo mức độ năng lực học tập của 6 sinh viên trong Bảng 1 trong mẫu khảo sát phù hợp với nhu cầu tuyển dụng doanh nghiệp 2 chuyên về kỹ thuật phần mềm.

Bảng 2. So sánh độ tương đồng giữa năng lực học tập của sinh viên với nhu cầu tuyển dụng ngành kỹ thuật phần mềm

Độ đo	187IT23746	187IT14048	187IT14044	197CT31311	197CT09897	197CT34101
Jaccard	0,049	0,059	0,051	0,068	0,068	0,068
Cosine	0,272	0,255	0,270	0,236	0,244	0,236
Tf-idf + cosine	0,426	0,487	0,452	0,440	0,465	0,440
Word2vec+cosine	0,504	0,505	0,504	0,508	0,509	0,508
Fasttext + cosine	0,953	0,954	0,953	0,951	0,952	0,951



Hình 6. Độ tương đồng môn học của sinh viên trong nhóm khảo sát với nhu cầu tuyển dụng việc làm về kỹ thuật phần mềm của doanh nghiệp 2

Trong Bảng 2 và Hình 6, xét điểm số dòng Fasttext + cosine, hai sinh viên có mã số 187IT23746 và 187IT14044 có mức độ phù hợp công việc mà nhà tuyển dụng đưa ra đều đạt 0,953. Sinh viên có mã số 187IT14048 tuy học cùng khóa với 2 sinh viên trên nhưng mức độ phù hợp cao nhất là 0,954. Như đã trình bày, do sinh viên được môn học theo tín chỉ nên có thể thấy sinh viên 187IT14048 phù hợp với ngành kỹ thuật phần mềm.

3. Thử nghiệm 3 - Doanh nghiệp 3 chuyên về mạng máy tính có nhu cầu tuyển dụng như sau:

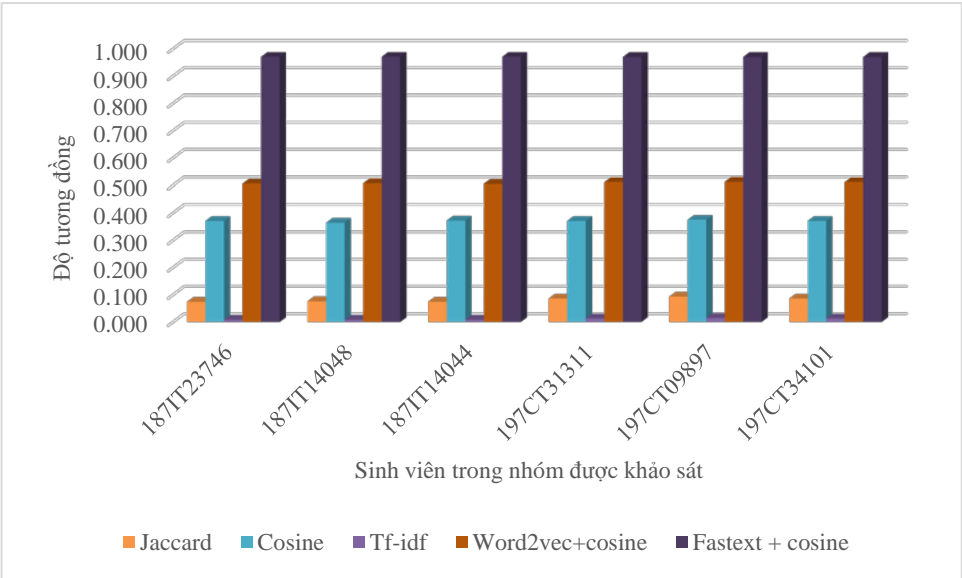
- Có ít nhất 2 năm làm kinh nghiệm trong công tác quản trị mạng, quản trị server, phần cứng mạng.
- Tốt nghiệp cao đẳng trở lên ưu tiên các ngành công nghệ thông tin.
- Thành thạo các trang thiết bị, phần mềm văn phòng.
- Kỹ năng sử dụng thành thạo các trang thiết bị văn phòng: Máy in, máy vi tính, máy scan....
- Trung thực, có tinh thần trách nhiệm cao trong công việc. Chủ động, độc lập trong công việc được giao, phối hợp nhóm tốt.

Bảng 3 thể hiện kết quả đo mức độ năng lực học tập của 6 sinh viên trong Bảng 1 trong mẫu khảo sát phù hợp với nhu cầu tuyển dụng doanh nghiệp 3 chuyên về mạng máy tính.

Bảng 3. So sánh độ tương đồng giữa năng lực học tập của sinh viên với nhu cầu tuyển dụng ngành mạng máy tính

Độ đo	187IT23746	187IT14048	187IT14044	197CT31311	197CT09897	197CT34101
Jaccard	0,074	0,076	0,074	0,086	0,093	0,086
Cosine	0,371	0,364	0,372	0,371	0,375	0,371
Tf-idf + cosine	0,008	0,008	0,008	0,013	0,015	0,013
Word2vec+cosine	0,506	0,507	0,505	0,511	0,512	0,511
Fasttext + cosine	0,971	0,971	0,971	0,970	0,970	0,970





Hình 7. Độ tương đồng môn học của sinh viên trong nhóm khảo sát với nhu cầu tuyển dụng việc làm về mạng máy tính của doanh nghiệp 3

Trong Bảng 3 và Hình 7, xét điểm số dòng Fasttext + cosine, ba sinh viên có mã số 187IT23746, 187IT14044, và 187IT14048 có mức độ phù hợp công việc mà nhà tuyển dụng đưa ra đều đạt như nhau 0,971. Như vậy, ba sinh viên này phù hợp với ngành mạng máy tính.

Qua thử nghiệm trình bày trong Bảng 1, Bảng 2 và Bảng 3, kỹ thuật nhúng từ FastText + cosine cho kết quả khá cao trên 90% so với các độ đo còn lại.

IV. KẾT LUẬN

Hệ khuyến nghị đem lại nhiều ứng dụng trong nhiều lĩnh vực và giúp người dùng dễ dàng tiếp cận những sản phẩm, dịch vụ đang quan tâm. Trong bài báo này, tác giả áp dụng mô hình lọc nội dung trong lĩnh vực giáo dục thông qua đề xuất hệ khuyến nghị việc làm cho sinh viên khi còn ngồi trên ghế nhà trường nhằm giúp sinh viên chuẩn bị kiến thức đầy đủ và chắc chắn trước khi chọn một công việc phù hợp. Nghiên cứu này sử dụng phương pháp nhúng từ để biểu diễn mô tả môn học chuyên ngành mà sinh viên đạt được và nhu cầu tuyển dụng của doanh nghiệp thành các vector. Sau đó, hệ thống tiến hành đo độ tương đồng nhằm gợi ý doanh nghiệp phù hợp với năng lực học tập của mỗi sinh viên. Nghiên cứu này nhận thấy việc dùng kỹ thuật nhúng từ sẽ cho kết quả cao hơn các kỹ thuật đo độ tương đồng phổ biến không dùng nhúng từ như cosine, Jaccard. Kết quả thử nghiệm ngoài việc khuyến nghị việc làm phù hợp còn phát hiện sở thích của sinh viên khi đăng ký môn học. Tuy nhiên, hệ thống còn những hạn chế như chỉ thử nghiệm trên tập dữ liệu sinh viên khá nhỏ chưa xử lý đối với dữ liệu lớn. Hệ thống chưa khuyến nghị những môn sinh viên cần phải học để trang bị thêm kiến thức so với nhu cầu doanh nghiệp. Trong tương lai, tác giả tiếp tục nghiên cứu giải pháp xử lý dữ liệu lớn và sử dụng các mô hình học máy để cải tiến hiệu suất của hệ thống.

TÀI LIỆU THAM KHẢO

[1] Đỗ Thành Nhân, Trần Nguyễn Minh Thư, “Hệ thống gợi ý áp dụng cho trang web tổng hợp tin tức tự động”, *Tạp chí Khoa học*, Trường Đại học Cần Thơ, pp. 190-197, 2013.

[2] Nguyễn Thái Nghe, Nguyễn Tấn Phong, “Xây dựng hệ thống gợi ý bài hát dựa trên phản hồi tiềm ẩn”, *Tạp chí Khoa học*, Trường Đại học Cần Thơ, pp. 81-91, 2014.

[3] Trần Nguyễn Minh Thư, Huỳnh Quang Nghi, “Hệ thống hỗ trợ tra cứu tài liệu”, *Tạp chí Khoa học*, Trường Đại học Cần Thơ, pp. 126-134, 2016.

[4] Farah Tawfiq Abdul Hussien, Abdul Monem S. Rahma, Hala Bahjat Abdul Wahab, “Recommendation systems for E-commerce systems an overview”, *Journal of Physics: Conference Series*, vol. 1897, 2021.

[5] Zihao Pu, Hongyu Du, Sizhe Yu, Duanyu Feng, “Improved tourism recommendation system”, *ICMLC, 12th International Conference on Machine Learning and Computing*, Shenzhen, China, 2020.

[6] Zeshan Fayyaz, Mahsa Ebrahimiyan, Dina Nawara, Ahmed Ibrahim, Rasha Kashef, “Recommendation systems: Algorithms, challenges, metrics, and business opportunities”, *Applied Sciences*, vol. 10, No. 7748, 2020.

[7] Trần Đình Khang, Võ Đức Quang, Nguyễn Đăng Tuấn Anh, “Hệ khuyến nghị cộng tác đồng tác giả”, *Tạp chí Khoa học*, Trường Đại học Huế: Kỹ thuật và Công nghệ, vol. 127, No. 2A, pp. 109-120, 2018.

[8] Fernando Ortega, Ángel González-Prieto, “Recommender systems and collaborative filtering”, *Applied Sciences*, vol. 10, No. 7050, 2020.

[9] Aleksandra Pawlicka, Marek Pawlicki, , Rafał Kozik, Ryszard S. Chora’s, “A systematic review of recommender systems and their applications in cybersecurity”, *Sensors*, vol. 21, No. 5248, 2021.

- [10] Didik Dwi Prasetya, Aji Prasetya Wibawa, Tsukasa Hirashima, “The performance of text similarity algorithms”, *International Journal of Advances in Intelligent Informatics*, vol. 4, No. 1, pp. 63-69, 2018.
- [11] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, “Efficient estimation of word representations in vector space”, *arXiv:1301.3781v3*, 2013.
- [12] Pritom Mojumder, Mahmudul Hasan, Md. Faruque Hossain, K. M. Azharul Hasan, “A study of fasttext word embedding effects in document classification in bangla language”, *Cyber Security and Computer Science. ICONCS*, Springer, Cham, pp. 441-453, 2020.
- [13] Maciej Besta, Raghavendra Kanakagiri, Harun Mustafa, Mikhail Karasikov, Gunnar Ratsch, Torsten Hoefer, Edgar Solomonik, “Communication-efficient jaccard similarity for high-performance distributed genome comparisons”, *arXiv:1911.04200v3*, 2019.
- [14] Tan Thongtan, Tanasanee Phienthrakul, “Sentiment classification using document embeddings trained with cosine similarity”, *Proceedings of the 57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, Florence, Italy, 2019.
- [15] Odunayo Esther Oduntan, Ibrahim Adeyanju, Falohun A.s, Olumide Obe, “A comparative analysis of euclidean distance and cosine similarity measure for automated Essay-Type Grading”, *Journal of Engineering and Applied Sciences*, vol. 13, No. 11, pp. 4198-4204, 2018.

## DEVELOPING A JOB RECOMMENDED SYSTEM FOR STUDENTS IN INFORMATION TECHNOLOGY BASED ON LEARNING CAPABILITY MEETING THE NEEDS OF EMPLOYERS

Nguyen Thai Anh, Phan Ho Viet Truong

**ABSTRACT:** A recommended system is a system that suggests products, services for users according to their purchase history. The recommended system has been applied in different fields such as tourism, education, news, music, so on. However, little studies have used the recommended system to introduce jobs to undergraduate students. This paper proposed a method to build a job recommended system for Van Lang University's students in information technology that includes majors, i.e., data science, software engineering, and computer network. This study used the content-filtered technique to measure the similarity between subjects that students passed with high scores and the needs of information technology companies. The system then selected top 5 the most appropriated companies and sent recruitment information to student's email. The experimented data were collected from the transcripts of about 500 third and final-year students who have followed those majors in Van Lang University. As a result, our proposed method got better performance than the common methods such as Jaccard, Cosine, TF-IDF+cosine, word2vec + cosine. Our method may be applied for other departments at Van Lang University.