MÔ HÌNH GIỚI THIỆU KỸ NĂNG CHUYÊN MÔN, CÔNG CỤ, KỸ NĂNG MỀM CỦA CÔNG VIỆC TRONG NGÀNH CÔNG NGHỆ THÔNG TIN

MODEL RECOMMEND TECHNICAL SKILLS, SOFT SKILLS, TOOLS OF JOBS FOR INFORMATION TECHNOLOGY

Trần Hữu Ân   
Khoa công nghệ thông tin  
Trường Đại học Giao Thông Vân Tải Tp HCMtranhuuan0202@gmail.com

*Abstract*—Đây là bài báo nghiên cứu mô tả các công việc về ngành công nghệ thông tin được thực hiện thu thập dữ liệu dựa trên trang web Indeed.com là một trang web tuyển dụng được thu thập vào ngày 25 tháng 01 năm 2023. Kiểm định giả thuyết Chi-square được sử dụng để đánh giá mức độ tương quan với nhau. Nhằm thống kê các loại hình vị trí công việc trong ngành công nghệ thông tin, các kỹ năng hiện đang được các công ty yêu cầu, các công nghệ sử dụng. Dựa vào đó ứng viên có thể tìm ra được các loại kỹ năng và công nghệ cho bản thân trước khi ứng tuyển.

Nghiên cứu đã ứng dụng mô hình BERTopic để xây dựng mô hình đề xuất các công việc theo các loại kỹ năng và công nghệ sử dụng của ứng viên trong ngành công nghệ thông tin. Với đầu vào là một đoạn văn bản nói về kỹ năng và công nghệ có thể sử dụng của ứng viên.

Keywords— information technology, skill, tool, degree, location, level.

# Đặt vấn đề.

Công nghệ thông tin là một trong những ngành có chuyển biến tích cực nhất trong những năm qua khi đã triển khai ứng dụng công nghệ thông tin trong đời sống, xã hội và doanh nghiệp. Việt Nam là một trong những nước có số lượng người tham gia internet thuộc top cao nhất trên thế giới hiện nay. Việt Nam có hơn 61,37 triệu người dùng, tương đương tỉ lệ 64% dân số đang sở hữu smartphone nằm trong top 10 quốc gia có lượng người dùng smartphone lớn nhất thế giới [1]. Nguồn lực công nghệ thông tin cũng được chú trọng để đáp ứng xu thế phát triển của xã hội. Số lượng các công việc trong các ngành nghề vô cùng đa dạng. Theo quan sát các chỉ số về tỉ lệ doanh nghiệp và nhân lực ngành CNTT trong báo cáo VNICT Index 2016, số lượng nhu cầu nhân lực trong ngành vẫn tiếp tục tăng trưởng ở mức 8%. Lượng nhân sự thiếu hụt khoảng 78.000 nhân lực CNTT mỗi năm và đến năm 2020 sẽ thiếu hơn 500.00 nhân lực CNTT, chiếm hơn 78% tổng số nhân lực mà thị trường này cần [2].

Theo báo cáo thị trường công nghệ thông tin Việt Nam 2021 Developers Recruitment State, chất lượng sinh viên sau khi tốt nghiệp đáp ứng yêu cầu tuyển dụng của doanh nghiệp còn thấp. Cụ thể, trong số hơn 55.000 sinh viên công nghệ thông tin (CNTT) tốt nghiệp mỗi năm chỉ có khoảng 16.500 sinh viên (30%) đáp ứng được những kỹ năng và chuyên môn mà doanh nghiệp cần. Tuy nhiên nhu cầu cho ngành công nghệ thông tin tại Việt Nam vẫn đang tăng cao liên tục. Năm 2021 Việt Nam cần đến 450.000 nhân lực trong ngành công nghệ thông tin [3].

Với nhu cầu tuyển dụng trong lĩnh vực công nghệ thông tin (CNTT) luôn trên đà tăng trưởng mạnh. Việc thiếu hụt nhân sự trong ngành công nghệ thông tin tại Việt Nam là một vấn đề nghiêm trọng. Đối tượng sinh viên là một trong những đối tượng cần được quan tâm trong việc giải quyết vấn đề này. Nghiên cứu này mong muốn giải quyết được vấn đề thiếu hụt nhân sự ở các công ty và tình trạng thất nghiệp của ứng viên.

Nghiên cứu này sẽ đưa ra các thống kê về các vị trí, các kỹ năng cũng như các công cụ trong ngành công nghệ thông tin. Sau đó các ứng viên có thể dựa vào các công nghệ, kỹ năng thống kê của nghiên cứu để chọn ra các kỹ năng, công nghệ phù hợp với vị trí tuyển chọn cho bản thân. Nghiên cứu đã xây dựng một mô hình gợi ý vị trí công việc theo kỹ năng và công nghệ của ứng viên để đề xuất ra các vị trí công việc phù hợp.

# phương pháp nghiên cứu.

## Đối tượng nghiên cứu.

Đối tượng nghiên cứu được lấy từ các bài viết tuyển dụng nhân sự của các công ty tại Việt Nam được lấy từ website Indeed. Từ ngày 25 tháng 01 năm 2023.

## Thời gian và địa điểm nghiên cứu.

### Thời gian nghiên cứu.

Trong tháng 1 năm 2023.

### Địa điểm nghiên cứu.

Trường Đại học Giao Thông Vận Tải TP HCM.

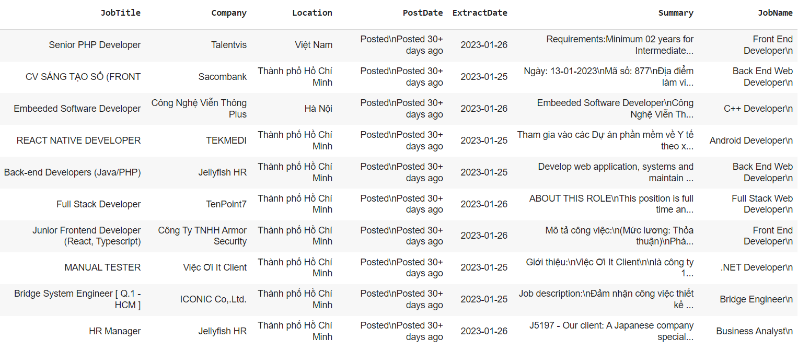
## Thiết kế nghiên cứu.

Đây là một nghiên cứu mô tả được thực hiện thu thập từ ngày 25 tháng 01 năm 2023 đến ngày 26 tháng 01 năm 2023. Bộ dữ liệu được thực hiện thu thập dựa trên các thông tin được công khai từ các bài tuyển dụng trên website. Dữ liệu được lấy từ 4379 bài tuyển dụng khác nhau về các vị trí được tuyển dụng trong ngành công nghệ thông tin.

## Thu thập dữ liệu.

### Quy trình thu thập dữ liệu.

Thu thập dữ liệu sẽ tiếp cận website Indeed vào ngày 25 tháng 01 năm 2023 thông qua sử dụng thư viện Selenium, pandas, datetime của ngôn ngữ Python để thu thập dữ liệu thô. Dữ liệu được thu thập bao gồm: Tên vị trí công việc, địa điểm, công ty, thời gian đã đăng, ngày thu thập dữ liệu, mô tả công việc, loại công việc.



bảng 1: Dữ liệu thu thập ở website tuyển dụng

Môi trường Jupyter notebook được sử dụng để thu thập dữ liệu bằng ngôn ngữ python. Thông tin dữ liệu thu thập sẽ được lấy từ các thẻ của bài tuyển dụng trên website. Bộ dữ liệu sẽ được làm sạch bằng cách xóa bỏ các kí tự đặc biệt. Sau đó sẽ lọc các từ khóa quan trọng có trong đoạn văn bằng cách sử dụng TF-IDF và thư viện nltk.

Quá trình sau khi đã tiền xử lý dữ liệu sẽ thu được các thông tin như: tên vị trí công việc, địa điểm, công ty, thời gian đã đăng, ngày thu thập dữ liệu, kỹ năng, công nghệ sử dụng, trình độ kinh nghiệm thâm niên theo cấp bậc, trình độ kinh nghiệm theo năm, trình độ học vấn, loại công việc.



bảng 2: dữ liệu mẫu của công việc sau khi đã được tiền xử lý.

### Phân tích dữ liệu.

Các biến định tính như: Kỹ năng, công nghệ sử dụng, kinh nghiệm trình độ thâm niên… Sẽ được mô tả dưới dạng biểu đồ tỉ lệ phần trăm. Kiểm định giả thuyết tương quan sẽ được sử dụng để đánh giá tính đúng đắn của kết quả. Với giá trị mức ý nghĩa là 0.05.

# Kết quả nghiên cứu.

Nghiên cứu đã tiến hành thu thập dữ liệu trên website Indeed.com được 4379 các bài tuyển dụng và có 3760 dữ liệu không bị trùng lặp với nhau. Tỉ lệ trùng lặp là 14.5%. Trong tổng số 3760 bài đăng tuyển dụng.

Nghiên cứu đã áp dụng phương pháp kiểm định giả thuyết Chi-Square là một kiểm định về giả thuyết tương quan. Công thức kiểm định giả thuyết tương quan.

* : Là giá trị thống kê Chi-Square.
* O: Là tần số quan sát trong số liệu.
* E: là tần số giả thuyết nếu giả thuyết không chắc chắn.

Nghiên cứu đã sử dụng thuật toán TF-IDF để phân tích và tìm ra độ quan trọng các từ trong câu dựa theo trong trọng số của chúng. Tầm quan trọng tỉ lệ thuận với số lần các từ xuất hiện trong tài liệu và tỉ lệ nghịch với tần số của các từ xuất hiện trong kho văn bản. Theo bài báo LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model thuật toán TF-IDF được biểu diễn dưới dạng toán học như sau [4]:

Trong đó:

* : Là tần số xuất hiện của một từ trong văn bản.
* : Tần số nghịch của 1 từ trong văn bản.

## Thông tin về trình độ học vấn.

Trong tổng số 3760 bài tuyển dụng về nhân sự đã được chắt lọc từ 4379 bài tuyển dụng cho công việc tại các công ty công nghệ thông tin. Có khoảng gần 68.7% các công ty công nghệ yêu cầu về trình độ bằng cấp và 31.1% công ty công nghệ không yêu cầu điều này.

Biểu đồ 1: Phần trăm các công ty công nghệ đề cập đến bằng cấp.

Với kiểm định giả thuyết Chi-Square P-value < -0.2 theo mức ý nghĩa là 0.05. Nghiên cứu đã cho rằng yêu cầu về trình độ bằng cấp có liên quan đến trình độ kinh nghiệm thâm niên trong nghề. Phần lớn các công ty công nghệ khi tuyển dụng không đề cập đến bằng cấp nằm ở kinh nghiệm thâm niên là Internship/fresher và kinh nghiệm thâm niên Senior ít đề cập về bằng cấp nhất. Chi tiết được thể hiện ở biểu đồ 2.

Biểu đồ 2: Tỉ lệ phần trăm yêu cầu về trình độ bằng cấp của các công ty tuyển dụng theo kinh nghiệm thâm niên.

## Nguồn công việc theo vị trí địa lý.

Các thành phố lớn ở Việt Nam như Hà Nội, TP. HCM luôn là những điểm có nhiều công việc. Điều này có thể được giải thích bằng một số lí do. Các thành phố lớn là trung tâm của kinh tế và chính trị của đất nước. Các thành phố lớn có nhiều trường đại học, cao đẳng và trung cấp chuyên nghiệp. Các thành phố lớn có sự phát triển về công nghệ, dịch vụ và du lịch. Theo nghiên cứu phần lớn các công việc tuyển dụng nằm ở TP. HCM chiếm khoảng 63.8% lượng công việc của cả nước. Chi tiết xem biểu đồ 3.

Biểu đồ 3: Tỉ lệ phần trăm công việc tuyển dụng công nghệ thông tin ở các địa điểm của Việt Nam.

## Vị trí trong công việc công nghệ thông tin.

Hầu hết các vị trí Business Analyst chiếm 16.74% và .Net Developer chiếm 14.92% là các vị trí được tuyển dụng nhiều nhất trong 13 vị trí trong ngành công nghệ thông tin thông qua các bài tuyển dụng.

Biểu đồ 4: Tỉ lệ phần trăm vị trí công việc có tại Việt Nam.

## Tỉ lệ công nghệ được sử dụng trong nhiều trong ngành công nghệ thông tin.

Trong ngành công nghệ thông tin ứng dụng các công nghệ đóng vai trò rất quan trọng trong việc giải quyết các vấn đề về kỹ thuật, tối ưu hóa quy trình làm việc. Hầu hết Excel chiếm 23.7% trong tổng số các công cụ là một trong những công cụ được sử dụng phổ biến nhất trong các vị trí công nghệ thông tin.

Biểu đồ 5: công nghệ có tỉ lệ sử dụng cao nhất trong ngành công nghệ thông tin tại Việt Nam.

## Tỉ lệ ngôn ngữ được sử dụng nhiều trong ngành công nghệ thông tin.

Ngôn ngữ lập trình là một công cụ cho phép lập trình viên tạo ra các chương trình máy tính và ứng dụng. Theo nghiên cứu Javascript có đến 30.9% chiếm nhiều nhất trong số các ngôn ngữ lập trình được sử dụng.

Biểu đồ 6: Tỉ lệ phần trăm các ngôn ngữ phổ biến trong ngành công nghệ thông tin hiện nay.

## Các kỹ năng được tuyển dụng trong ngành công nghệ thông tin.

Kỹ năng mềm (soft skill) là một trong những yếu tố quan trọng trong công việc công nghệ thông tin. Các kỹ năng này sẽ hỗ trợ nhân viên làm việc một cách hiệu quả. Kỹ năng mềm bao gồm kỹ năng về ngoại ngữ, giải quyết vấn đề, giao tiếp, làm việc nhóm, …. Theo nghiên cứu ngoại ngữ, quản lí, giao tiếp là những kỹ năng quan trọng hầu hết các nhà tuyển dụng đều yêu cầu.

Biểu đồ 7: Kỹ năng mềm được yêu cầu tuyển dụng trong ngành công nghệ thông tin hiện nay.

Technical skill là kỹ năng quan trọng trong ngành công nghệ thông tin. Kỹ năng bao gồm về kỹ năng lập trình, thiết kế, phân tích, databases. Theo thống kê trong nghiên cứu có kỹ năng về design chiếm 23.56% là một kỹ năng phổ biến trong lập trình ngày nay.

Biểu đồ 8: Tỉ lệ phần trăm kỹ năng chuyên môn trong ngành công nghệ thông tin ở Việt Nam

# thảo luận.

Tuyển dụng vị trí ngành công nghệ thông tin đang là một vấn đề gây tranh cãi. Với sự phát triển của kinh tế số, nhu cầu nhân lực trong lĩnh vực này ngày càng tăng. Tuy nhiên, việc tuyển dụng nhân lực chất lượng cao trong lĩnh vực công nghệ thông tin lại không phải là một điều dễ dàng. Theo một cuộc nghiên cứu của lực lượng lao động khoa học và kỹ thuật của Hoa Kỳ vào tháng 6 năm 2013, đã cho rằng vào năm 2011 tỉ lệ thất nghiệp ngành công nghệ thông tin lên đến 4.1% khoảng 3.293.020 người thất nghiệp đứng thứ hai trong khối ngành S&E (Science & Engineer) sau Mathematical [5].

Một trong những thách thức đối với tuyển dụng công nghệ thông tin ở Việt Nam hiện nay là thiếu hụt nhân lực chất lượng. Nhiều doanh nghiệp gặp khó khăn trong việc tìm kiếm và tuyển dụng các ứng viên có kinh nghiệm và kỹ năng đáp ứng được yêu cầu của công việc

Các kỹ năng, công cụ và ngôn ngữ lập trình đóng vai trò rất quan trọng trong việc xây dựng và phát triển các ứng dụng và hệ thống. Đầu tiên, các kỹ năng về ngôn ngữ lập trình, quản lý dữ liệu, bảo mật thông tin, mạng. Ngoài ra, Kỹ năng giải quyết vấn đề, tư duy logic…cũng rất quan trọng trong việc xử lý các tình huống phát sinh và tối ưu hiệu xuất của ứng dụng. Các công cụ phát triển phần mềm các trình quản lý mã nguồn và công cụ lưu trữ dữ liệu là cần thiết để tạo ra một hệ thống chất lượng, đảm bảo tính ổn định. Kiểm định giả thuyết tỉ lệ đã đánh giá, Theo thống kê nghiên cứu của VietNam IT market report ngôn ngữ lập trình đang phổ biến là Javascript lên đến 70% vào năm 2021 so sánh tỉ lệ trong nghiên cứu này vào năm 2023 tỉ lệ đã giảm xuống [3].

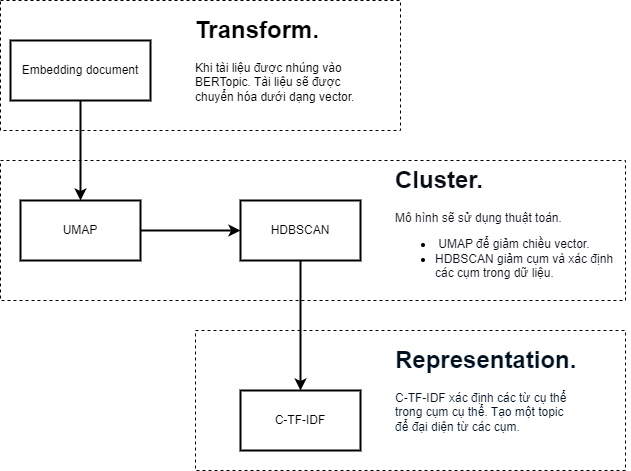
So sánh về mối liên quan giữa kinh nghiệm thâm niên và ngôn ngữ lập trình được sử dụng. Nhìn chung ngôn ngữ C, Python và C++ là ngôn được kinh nghiệm Internship / fresher sử dụng nhiều nhất. Java là ngôn được nhưng nhân viên có kinh nghiệm Senior được sử dụng nhiều. Nhưng ngôn ngữ được các nhân viên Internship hoặc fresher sử dụng phần lớn đây là những ngôn ngữ dễ học hoặc là những ngôn ngữ này được các trường đại học dùng để giảng dạy.

Lưu ý: Bài báo này vẫn chưa giải quyết được vấn đề trong quá trình thu thập dữ liệu thô từ các bài tuyển dụng của các công ty. Có một số bài tuyển dụng thuộc các công ty vừa và nhỏ yêu cầu làm nhiều công việc trong bài tuyển dụng của mình. Hiện tại nhóm nghiên cứu vẫn chưa giải quyết được vấn đề đó. Vì thế quá trình xây dựng mô hình dự đoán có thể gặp sai sót.

# mô hình gợi ý công việc theo phương pháp thống kê.

## Giới thiệu mô hình BERTopic.

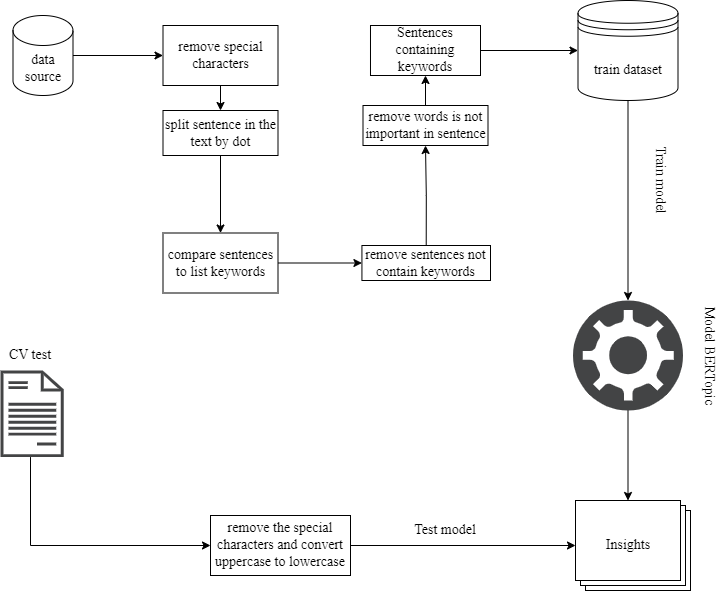
BERTopic là một mô hình phân cụm văn bản dựa trên BERT (Bidirectional Encoder Representations from Transformers), một kiến trúc mạng Neural sử dụng trong xử lý ngôn ngữ tự nhiên (NLP). Dùng để phân loại và trích xuất các chủ để từ tập dữ liệu văn bản. Chi tiết về BERTopic được mô tả ở sơ đồ 1.



Sơ đồ 1: Biểu diễn quy trình xử lý của thuật toán BERTopic.

## Quy trình xử lý công việc và xây dựng mô hình gợi ý vị trí công việc.

Nghiên cứu đã xây dựng mô hình gợi ý theo phương pháp thống kê của một hồ sơ ứng viên như sau. Đầu vào là đoạn mô tả các kỹ năng và công nghệ của ứng viên. Các kỹ năng và công nghệ mà ứng viên có sẽ được so sánh với các kỹ năng tương ứng mà vị trí công việc có trong bộ dữ liệu.



Sơ đồ 2: Mô hình gợi ý dựa theo mô hình BERTopic kết hợp phương pháp thống kê.

Quy trình xử lý như sau: Nguồn dữ liệu thu thập được làm sạch thông qua các bước như loại bỏ các kí tự đặc biệt, phân tách đoạn văn bản thành các câu, chọn lọc các câu có chứa các từ khóa và các câu không chứa các từ khóa sẽ bị loại bỏ đi. Các câu có chứa từ khóa sẽ được tiếp tục loại bỏ các từ không quan trọng trong câu. Các câu có chứa từ khóa sẽ được gắn một label về tên vị trí công việc. Mô hình BERTopic (mô hình phân cụm) sẽ được ứng dụng để train bộ dữ liệu. Kết quả đầu ra là những danh sách vị trí và mức độ phù hợp với ứng viên.

# kết luận.

Hiện nay, ngành công nghệ thông tin đang phát triển với tốc độ vượt bậc và đã góp phần tạo nên sự thay đổi to lớn trong cuộc sống của chúng ta. Tuy nhiên, những thách thức của ngành công nghệ đang phải gặp là thiếu hụt nhân sự chuyên môn. Với tốc độ phát triển nhanh chóng về ngành công nghệ thông tin này. Các công ty công nghệ đòi hỏi các nhân viên trong lĩnh vực này phải liên tục cập nhật và thay đổi và trau dồi các kỹ năng.

Nghiên cứu xây dựng mô hình theo phương pháp thống kê để đưa ra các gợi ý công việc phù hợp cho ứng viên theo công nghệ sử dụng và kỹ năng của ứng viên. Các kỹ năng và công nghệ của ứng viên sẽ được so sánh với mặt bằng chung của các vị trí ngành công nghệ trên thị trường hiện nay. Để góp phần giảm thiểu về tình trạng thiếu hụt nhân sự hiện nay.

# Tài liệu trích dẫn.

|  |  |
| --- | --- |
| [1] | T. L. T. ANH, "Tiềm năng phát triển nền kinh tế số ở Việt Nam và một số hàm ý chính sách," *TẠP CHÍ CÔNG THƯƠNG - CƠ QUAN THÔNG TIN LÝ LUẬN CỦA BỘ CÔNG THƯƠNG,* 2022. |
| [2] | B. T. T. V. T. THÔNG, "Báo cáo chỉ số sẵn sàng cho phát triển và ứng dụng công nghệ thông tin và truyền thông Việt Nam 2016 -VN Index 2016," THÔNG TIN VÀ TRUYỀN THÔNG, 2017. |
| [3] | TopDev, "VIET NAM IT REPORT MARKET Developers Recruitment State," Viet Nam, 2021. |
| [4] | M. C. Y. C. O. B. Mohamed Chiny, "LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model," *(IJACSA) International Journal of Advanced Computer Science and Applications,* 2021. |
| [5] | J. F. S. Jr., "The U.S. Science and Engineering Workforce: Recent, Current, and Projected Employment, Wages, and Unemployment," Congressional Research Service, 2013. |
| [6] | V. InTECH, "báo cáo thị trường nhân lực ngành CNTT thập niên 2010 & năm 2020," Viet Nam, 02/06/2020. |
| [7] | J. S. G. G. Karen Kaminski, "Workforce readiness: A study of university students’ fluency with information technology," September 2009. |
| [8] | D. G. S. Dr. Janet Bailey, "Preparing the Information Technology Workforce," *ACM Journals,* p. https://dl.acm.org/doi/abs/10.1145/571475.571476, 01 August 2002. |
| [9] | M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 11 03 2022. [Online]. Available: chrome-extension://gphandlahdpffmccakmbngmbjnjiiahp/https://arxiv.org/pdf/2203.05794.pdf. |