**HỆ THỐNG ĐIỀN BIỂU MẪU TỰ ĐỘNG HỖ TRỢ AI DỰA TRÊN LỊCH SỬ NGƯỜI DÙNG: MỘT CÁCH TIẾP CẬN MỚI ĐỂ NÂNG CAO HIỆU QUẢ HÀNH CHÍNH TRONG GIÁO DỤC**

AI-Assisted Automatic Form Filling System Based on User History: A Novel Approach to Enhancing Administrative Efficiency in Education

***Nguyễn Tuấn Đạt1****,* ***Nguyễn Anh Duy2, ThS. Trần Thanh Nam2, ThS. Trần Văn Thiện1***

*1Trường Đại học Nam Cần Thơ, Việt Nam*

*2Adhightech Ltd, Viet Nam*

***Nguyễn Tuấn Đạt1, Nguyễn Anh Duy2, ThS. Trần Thanh Nam2, ThS. Trần Văn Thiện1***

*1Nam Can Tho University, Việt Nam*

*2Adhightech Ltd, Viet Nam; Email: ntdat1610c@gmail.com*

**Tóm tắt -** Trong bối cảnh số hóa mạnh mẽ lĩnh vực giáo dục, nhu cầu tự động hóa các tác vụ hành chính đang ngày càng trở nên cấp thiết. Nghiên cứu này đề xuất và phát triển một hệ thống điền biểu mẫu tự động thông minh dựa trên lịch sử người dùng, nhằm giảm thiểu thao tác thủ công lặp lại và hạn chế sai sót trong các biểu mẫu hành chính như đơn xin nghỉ học, đăng ký học phần, hồ sơ học sinh – sinh viên, và phiếu khảo sát. Hệ thống tập trung phục vụ các đối tượng trong môi trường giáo dục như học sinh, sinh viên, giảng viên và cán bộ quản lý. Phương pháp tiếp cận kết hợp giữa các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) từ OpenAI để gợi ý và hoàn thiện nội dung theo ngữ cảnh cá nhân hóa. Quá trình phát triển hệ thống bao gồm thu thập dữ liệu biểu mẫu thực tế, chuẩn hóa thông tin, xây dựng thuật toán so khớp trường dữ liệu, và huấn luyện mô hình dựa trên hành vi người dùng. Tiềm năng mở rộng bao gồm tích hợp với các nền tảng quản lý giáo dục (LMS, SIS), hỗ trợ các định dạng biểu mẫu khác và phát triển khả năng học liên tục để nâng cao chất lượng đề xuất theo thời gian. Qua đó, nghiên cứu góp phần khẳng định vai trò thiết yếu của AI trong tối ưu hóa công tác hành chính và nâng cao hiệu quả quản lý giáo dục trong thời đại số.

**Từ khóa -** Trí tuệ nhân tạo, Hệ thống điền biểu mẫu tự động, Quản lý giáo dục, Xử lý ngôn ngữ tự nhiên, Lịch sử người dùng, Mô hình ngôn ngữ lớn, Tự động hóa hành chính giáo dục. **Abstract -** In the context of rapid digitalization in the education sector, the need to automate administrative tasks has become increasingly urgent. This study proposes and develops an intelligent automatic form-filling system based on user history to minimize repetitive manual input and reduce errors in administrative forms such as leave applications, course registration forms, student records, and surveys. The system is designed to serve key stakeholders in the educational environment, including students, lecturers, and administrative staff. The approach integrates natural language processing (NLP) techniques from OpenAI to suggest and complete content in a personalized and context-aware manner. The development process involves collecting real-world form data, standardizing information, constructing field-matching algorithms, and training models based on user behavior. Potential for system expansion includes integration with educational management platforms (LMS, SIS), support for various form formats, and the development of continual learning capabilities to improve suggestion quality over time. This study highlights the essential role of AI in optimizing administrative processes and enhancing educational management efficiency in the digital age.

**Keywords -** Artificial Intelligence, Automatic Form Filling System, Educational Management, Natural Language Processing, User History, Large Language Models, Educational Administrative Automation.

# Giới thiệu

Trong bối cảnh chuyển đổi số đang được đẩy mạnh tại Việt Nam và các quốc gia đang phát triển, việc tối ưu hóa quy trình hành chính trong giáo dục là một yêu cầu cấp thiết. Hiện nay, tại nhiều trường đại học, quy trình xử lý các biểu mẫu như đơn xin nghỉ học, đơn đăng ký môn học, phiếu khảo sát... vẫn chủ yếu được thực hiện thủ công. Theo khảo sát nội bộ tại một số cơ sở đào tạo quy mô trung bình, mỗi học kỳ có thể phát sinh từ 5.000 đến 10.000 biểu mẫu cần xử lý, và thời gian trung bình để một sinh viên hoặc cán bộ hoàn thành một biểu mẫu thủ công dao động từ 5 đến 10 phút. Với con số hàng nghìn biểu mẫu, lượng thời gian tiêu tốn là rất lớn và dễ dẫn đến sai sót, quá tải hệ thống, hoặc làm giảm trải nghiệm người dùng.

Các giải pháp truyền thống hiện nay như Autofill của trình duyệt (Chrome, Safari), các phần mềm quản lý hành chính hoặc các nền tảng tự động hóa như Microsoft Power Automate tuy đã góp phần giảm tải một phần công việc, nhưng vẫn còn nhiều điểm hạn chế. Những công cụ này không có khả năng hiểu rõ nội dung biểu mẫu, không đưa ra gợi ý thông minh dựa trên lịch sử người dùng, và thường chỉ áp dụng được với các biểu mẫu có cấu trúc cố định. Việc cập nhật hoặc cá nhân hóa thông tin vẫn phải thực hiện thủ công trong hầu hết các trường hợp.

Trong khi đó, hệ thống được đề xuất trong nghiên cứu này kết hợp giữa trí tuệ nhân tạo (AI) và xử lý ngôn ngữ tự nhiên (NLP) nhằm phân tích trực tiếp các biểu mẫu Word (.docx) do người dùng tải lên từ web. Hệ thống không chỉ nhận diện các trường thông tin mà còn hiểu ngữ cảnh, gợi ý nội dung dựa trên lịch sử sử dụng trước đó của từng cá nhân, và cho phép tải xuống biểu mẫu đã hoàn chỉnh. Nhờ đó, thời gian xử lý mỗi biểu mẫu có thể rút ngắn xuống chỉ còn 1 – 2 phút, đồng thời giảm đáng kể nguy cơ sai sót và tăng tính nhất quán trong dữ liệu hành chính.

Điểm đặc biệt là hệ thống có khả năng học hỏi liên tục từ hành vi người dùng thực tế, ngày càng nâng cao độ chính xác và chất lượng đề xuất. Đây là một điểm mạnh vượt trội so với các phương pháp thủ công hoặc các công cụ không dùng AI. Hệ thống cũng có thể dễ dàng tích hợp với các nền tảng giáo dục hiện có như LMS hoặc SIS, giúp tăng hiệu quả quản lý tổng thể.

Tóm lại, nghiên cứu này không chỉ có ý nghĩa trong việc giảm tải công việc hành chính lặp lại trong giáo dục, mà còn thể hiện một bước tiến rõ rệt trong việc ứng dụng AI vào các quy trình hành chính thông minh và cá nhân hóa. Tính khả thi, tính thực tiễn và tiềm năng mở rộng của hệ thống có thể tạo ra tác động tích cực lâu dài cho ngành giáo dục tại Việt Nam và các quốc gia đang phát triển.

# Nghiên cứu hệ thống điền biểu mẫu tự động

Việc điền biểu mẫu tự động là một trong những lĩnh vực quan trọng trong nghiên cứu về tự động hóa hành chính và xử lý ngôn ngữ tự nhiên. Trải qua nhiều giai đoạn phát triển, các phương pháp tiếp cận trong lĩnh vực này ngày càng được cải tiến nhằm nâng cao tính linh hoạt, độ chính xác và khả năng cá nhân hóa cho người dùng.

Giai đoạn đầu tiên của việc tự động điền biểu mẫu chủ yếu dựa trên hệ thống rule-based – tức là sử dụng các luật cứng để xác định và điền thông tin vào các trường dữ liệu. Các hệ thống này thường dựa vào cấu trúc DOM của trang web để nhận dạng tên trường và nội dung cần điền thông qua các nhãn như “Tên”, “Email”, “Địa chỉ”,... Các công cụ tiêu biểu trong giai đoạn này như hệ thống đề xuất bởi De Luca et al. (2007) đã ứng dụng proxy cục bộ trên thiết bị di động để phân tích biểu mẫu đầu vào và áp dụng các luật logic cố định nhằm xác định vị trí và ý nghĩa các trường dữ liệu. Công cụ Dhvani, chẳng hạn, sử dụng bố cục biểu mẫu và các quy tắc ngữ cảnh để tự động gán giá trị cho trường nhập liệu. Tuy nhiên, điểm hạn chế lớn của các phương pháp rule-based là thiếu khả năng thích ứng: khi gặp biểu mẫu có bố cục hoặc ngữ nghĩa mới lạ, hệ thống dễ bị lỗi và không thể điền chính xác. Việc cập nhật hệ thống cũng tốn kém và đòi hỏi lập trình thủ công mỗi khi có thay đổi nhỏ trong biểu mẫu.

Từ năm 2010, học máy (machine learning) bắt đầu được áp dụng vào bài toán điền biểu mẫu. Khác với rule-based, học máy cho phép hệ thống học từ dữ liệu đầu vào thực tế của người dùng để đưa ra dự đoán phù hợp cho các trường thông tin. Một số sản phẩm tiêu biểu sử dụng cách tiếp cận này là Google Smart Autofill và Facebook Autofill SDK – những hệ thống khai thác dữ liệu duyệt web hoặc tài khoản cá nhân để tự động hoàn thành các biểu mẫu đăng ký, đăng nhập hoặc khảo sát. Các thuật toán được sử dụng trong giai đoạn này bao gồm KNN (K-Nearest Neighbors), Naive Bayes, Decision Tree và các mạng nơ-ron đơn giản. Một ví dụ điển hình là CloudScan (Palm et al., 2020) – một hệ thống trích xuất hóa đơn sử dụng mạng hồi tiếp (RNN) để học cách gán thông tin từ các tài liệu đầu vào.

Cùng với sự phát triển của web ngữ nghĩa và các công nghệ ngữ nghĩa (semantic web), một nhánh nghiên cứu khác tập trung vào khai thác ngữ cảnh và ngữ nghĩa để nâng cao độ chính xác của hệ thống điền biểu mẫu. Phương pháp này không chỉ đơn giản là nhận diện chuỗi ký tự giống nhau, mà còn hiểu được mối liên hệ giữa các nhãn trường và mục đích sử dụng. Ví dụ, hệ thống có thể hiểu rằng “tên đầy đủ”, “full name” hay “họ và tên” đều ám chỉ cùng một loại thông tin, và do đó có thể dùng dữ liệu từ lịch sử phù hợp để điền. Nghiên cứu của Microsoft (2015) cho thấy việc phân tích ngữ cảnh xung quanh nhãn trường giúp cải thiện hiệu suất nhận dạng và gợi ý điền mẫu. Các kỹ thuật RDF, OWL trong Semantic Web cũng được ứng dụng để giúp hệ thống hiểu sâu hơn về mục tiêu của biểu mẫu. Chen et al. (2021) phát triển FormNet, một mô hình sử dụng học tuần tự để hiểu cấu trúc biểu mẫu và tự động hóa trích xuất trường dữ liệu từ các tài liệu scan như hóa đơn, biểu mẫu khảo sát.

Từ năm 2020 trở lại đây, xu hướng nổi bật trong nghiên cứu là ứng dụng trí tuệ nhân tạo, đặc biệt là các mô hình ngôn ngữ lớn (LLMs) như BERT (Devlin et al., 2019) và GPT (Brown et al., 2020), vào hệ thống điền biểu mẫu tự động. Các mô hình này có khả năng hiểu ngữ nghĩa sâu sắc của văn bản, học hỏi từ lịch sử người dùng và đưa ra gợi ý điền biểu mẫu một cách chính xác, linh hoạt và có tính cá nhân hóa cao. Reimers và Gurevych (2019) phát triển mô hình Sentence-BERT, sử dụng mạng Siamese để so sánh ngữ nghĩa giữa các chuỗi văn bản, từ đó cho phép hệ thống xác định sự tương đồng giữa các nhãn biểu mẫu và thông tin người dùng lưu trữ. Nghiên cứu của Li et al. (2021) cho thấy khi mô hình học sâu được huấn luyện với lịch sử điền biểu mẫu thực tế của từng người dùng, hệ thống có thể tạo ra các đề xuất điền mẫu gần như chính xác tuyệt đối đối với các trường quen thuộc, đồng thời linh hoạt thích ứng với biểu mẫu mới.

Mặc dù đã có nhiều công cụ hỗ trợ điền biểu mẫu tự động, phần lớn các hệ thống hiện tại chỉ dừng lại ở việc điền thông tin cơ bản như họ tên, email, địa chỉ... dựa trên dữ liệu được lưu trữ sẵn trong trình duyệt hoặc ứng dụng. Chúng chưa đủ thông minh để hiểu nội dung cụ thể của biểu mẫu hoặc đưa ra gợi ý phù hợp với từng người dùng trong những hoàn cảnh khác nhau.

Đặc biệt, chưa có hệ thống nào thực sự tập trung vào việc điền biểu mẫu từ tài liệu Word (định dạng .docx) được người dùng tải lên từ web, sau đó tự động phân tích nội dung, hiểu được ngữ cảnh sử dụng, và đưa ra gợi ý điền thông tin cá nhân hóa dựa trên lịch sử điền trước đó của người dùng. Cuối cùng, hệ thống còn phải cho phép tải về biểu mẫu đã được điền hoàn chỉnh – điều này gần như không có trong các giải pháp phổ biến hiện nay.

Bên cạnh đó, các công cụ hiện tại hầu hết không có khả năng học hỏi liên tục từ dữ liệu sử dụng thực tế – tức là chúng không thể tự cải thiện theo thời gian để đưa ra các gợi ý ngày càng chính xác hơn. Ngoài ra, việc tích hợp với môi trường giáo dục (như hệ thống quản lý học tập – LMS hoặc hệ thống thông tin sinh viên – SIS) thường bị bỏ qua.

Hệ thống trong nghiên cứu này ra đời nhằm giải quyết chính những điểm còn thiếu đó. Nó cho phép người dùng tải biểu mẫu Word lên một giao diện web, sau đó sử dụng các kỹ thuật trí tuệ nhân tạo (AI), xử lý ngôn ngữ tự nhiên (NLP) và mô hình GPT để phân tích văn bản, hiểu được mục đích biểu mẫu, và gợi ý tự động thông tin cần điền dựa vào lịch sử của từng người dùng. Người dùng có thể chỉnh sửa nếu cần, và sau đó tải về phiên bản hoàn chỉnh của biểu mẫu – tất cả được thực hiện nhanh chóng, chính xác, và dễ sử dụng, kể cả với người không có kiến thức kỹ thuật.

|  |  |  |
| --- | --- | --- |
| Phương pháp | Ưu điểm | Nhược điểm |
| Rule-based | Đơn giản, dễ triển khai | Khó mở rộng, không linh hoạt với biểu mẫu mới |
| Machine Learning | Học từ dữ liệu lịch sử, có khả năng thích nghi | Cần lượng dữ liệu huấn luyện lớn, độ phức tạp triển khai cao |
| Deep Learning | Hiệu quả cao với dữ liệu lớn, xử lý tốt biểu mẫu phi cấu trúc | Tốn tài nguyên tính toán, cần xử lý nhiều về hiệu năng |
| AI/LLM | Gợi ý thông minh, hiểu ngữ cảnh sâu | Chi phí cao, vấn đề bảo mật và đạo đức dữ liệu |

***Bảng 1.*** *So sánh các phương pháp tiếp cận*

Bên cạnh đó, các xu hướng phát triển trong tương lai của hệ thống điền biểu mẫu tự động bao gồm: tăng cường cá nhân hóa dựa trên hành vi người dùng theo thời gian thực; tích hợp học liên tục (continual learning) để hệ thống thích nghi với biểu mẫu mới mà không cần huấn luyện lại toàn bộ; hỗ trợ đa định dạng biểu mẫu (DOCX, PDF, HTML,...); và nâng cao khả năng tương thích đa nền tảng (web, mobile, desktop). Đặc biệt, vấn đề bảo mật và quyền riêng tư sẽ ngày càng được quan tâm, đòi hỏi các hệ thống điền biểu mẫu phải áp dụng các kỹ thuật mã hóa, xác thực người dùng và lưu trữ dữ liệu an toàn.

Tổng kết lại, quá trình phát triển của các hệ thống điền biểu mẫu tự động trong lịch sử cho thấy sự chuyển dịch rõ rệt từ những phương pháp đơn giản như điền thủ công theo quy tắc sang việc ứng dụng học máy và xử lý ngôn ngữ tự nhiên để tự động hóa và cá nhân hóa nội dung. Tuy nhiên, nhiều hệ thống trước đây vẫn còn hạn chế ở khả năng thích ứng ngữ cảnh, độ chính xác khi gợi ý, và chưa tận dụng tốt dữ liệu lịch sử người dùng. Hệ thống mà nghiên cứu này đề xuất đã kế thừa những ưu điểm từ các phương pháp truyền thống và hiện đại, đồng thời cải tiến bằng cách kết hợp các mô hình ngôn ngữ mạnh mẽ (như GPT), thuật toán so khớp ngữ nghĩa (TF-IDF, Cosine Similarity, Word2Vec), và khả năng học từ dữ liệu lịch sử người dùng để đưa ra các gợi ý điền biểu mẫu chính xác, phù hợp ngữ cảnh và cá nhân hóa. Chính nhờ những cải tiến đó, hệ thống có thể rút ngắn đáng kể thời gian xử lý, giảm sai sót, và nâng cao hiệu quả sử dụng trong thực tiễn – đặc biệt trong lĩnh vực giáo dục, nơi lượng biểu mẫu cần điền mỗi ngày là rất lớn và yêu cầu tính chính xác cao. Đây là bước tiến quan trọng so với các nghiên cứu trước, góp phần thúc đẩy ứng dụng trí tuệ nhân tạo vào công tác hành chính hiện đại.

# Các mã nguồn và sản phẩm thương mại

Trong lĩnh vực tự động hóa biểu mẫu, nhiều giải pháp đã được phát triển dưới dạng mã nguồn mở, mã nguồn đóng và các sản phẩm thương mại. Mỗi giải pháp mang những đặc điểm riêng về tính năng, khả năng mở rộng, mức độ tùy biến, bảo mật và đối tượng người dùng hướng đến. Việc phân tích và so sánh các hệ thống hiện có giúp làm rõ lợi thế cạnh tranh của hệ thống được đề xuất trong nghiên cứu này, đồng thời cung cấp cơ sở định hướng phát triển hiệu quả hơn trong tương lai.

Các giải pháp mã nguồn mở mang lại sự linh hoạt và khả năng tùy chỉnh cao với chi phí triển khai thấp hoặc miễn phí. AutoHotkey là một công cụ kịch bản phổ biến cho hệ điều hành Windows, cho phép người dùng tự động hóa các tác vụ như nhập liệu hoặc thao tác chuột, phù hợp với những người có kỹ năng kỹ thuật mong muốn kiểm soát sâu quy trình điền biểu mẫu. Trong khi đó, Selenium kết hợp với Python là một bộ công cụ mạnh mẽ cho việc tự động hóa các biểu mẫu web phức tạp, hỗ trợ tích hợp với trí tuệ nhân tạo (AI) để tăng tính thông minh trong điền biểu mẫu. Tuy nhiên, công cụ này đòi hỏi người dùng phải có kiến thức lập trình nhất định và thời gian thiết lập không nhỏ. Một giải pháp khác là FillForm – một tiện ích mở rộng trình duyệt đơn giản, giúp người dùng lưu và tự động điền lại thông tin trên các biểu mẫu lặp lại. Tuy vậy, FillForm lại không có khả năng học từ dữ liệu lịch sử hay xử lý các biểu mẫu có cấu trúc phức tạp.

Bên cạnh đó, các giải pháp mã nguồn đóng lại thường được tích hợp trực tiếp trong các hệ điều hành hoặc nền tảng phổ biến, mang lại sự tiện lợi cho người dùng cuối. Autofill Forms, như các tính năng điền tự động trong trình duyệt Chrome hoặc Safari, cho phép người dùng nhập nhanh các thông tin cá nhân cơ bản. Tuy nhiên, các tính năng này thiếu khả năng tùy chỉnh nâng cao và không hiểu được ngữ cảnh sử dụng biểu mẫu. Power Automate của Microsoft là một nền tảng mạnh mẽ cho việc tự động hóa quy trình nghiệp vụ, hỗ trợ nhiều hệ thống và tích hợp sâu vào hệ sinh thái doanh nghiệp. Mặc dù vậy, Power Automate có thể gây khó khăn cho người dùng không chuyên và đi kèm chi phí sử dụng đáng kể. Trong hệ sinh thái Apple, AutoFill giúp người dùng iOS và macOS nhanh chóng điền thông tin với mức độ bảo mật cao. Dẫu vậy, công cụ này không cung cấp khả năng tùy biến cho các loại biểu mẫu đặc thù hoặc khả năng học hỏi từ lịch sử sử dụng.

Các sản phẩm thương mại lại chú trọng vào việc cung cấp các chức năng chuyên biệt với hiệu suất cao, hướng đến các tình huống sử dụng cụ thể. RoboForm là một trình quản lý mật khẩu kết hợp chức năng điền biểu mẫu phức tạp, đảm bảo tính bảo mật, phù hợp với người dùng hoặc tổ chức có yêu cầu cao về bảo vệ thông tin cá nhân. Tuy nhiên, công cụ này không sử dụng AI và không học từ dữ liệu lịch sử của người dùng. TextExpander là một giải pháp giúp mở rộng đoạn văn bản, tiết kiệm thời gian nhập liệu trong các tác vụ lặp đi lặp lại, chẳng hạn như trả lời email chăm sóc khách hàng, nhưng thiếu khả năng xử lý ngữ nghĩa hoặc tự động học nội dung. Fillout là một nền tảng thiết kế biểu mẫu và khảo sát động với giao diện thân thiện, hỗ trợ các logic điều kiện và tích hợp API mạnh mẽ, thích hợp cho doanh nghiệp hoặc tổ chức cần thu thập dữ liệu, nhưng vẫn thiếu các tính năng gợi ý dựa trên lịch sử điền biểu mẫu hoặc khả năng cá nhân hóa nội dung.

Tóm lại, các giải pháp hiện tại dù có nhiều ưu điểm nhưng vẫn còn những hạn chế trong việc hiểu ngữ cảnh và đề xuất nội dung điền biểu mẫu mang tính cá nhân hóa cao. Hệ thống được đề xuất trong nghiên cứu này khắc phục những điểm yếu đó bằng cách kết hợp các kỹ thuật tiên tiến trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP), vector hóa văn bản như TF-IDF, Word2Vec, đo độ tương đồng ngữ nghĩa bằng Cosine Similarity và ứng dụng mô hình ngôn ngữ hiện đại như GPT. Nhờ vào các công nghệ này, hệ thống không chỉ hiểu được nội dung và mục đích của biểu mẫu mà còn có thể tự động phân tích ngữ cảnh và gợi ý thông tin phù hợp dựa trên lịch sử sử dụng của người dùng. Giao diện web đơn giản cho phép người dùng dễ dàng tải lên các biểu mẫu định dạng Word (.docx) mà không cần cài đặt phần mềm hay có kiến thức kỹ thuật chuyên sâu. Hệ thống này đặc biệt phù hợp trong môi trường giáo dục, nơi mà cả người dạy và người học thường xuyên phải xử lý nhiều loại biểu mẫu khác nhau. Bằng cách rút ngắn thời gian điền biểu mẫu, tăng độ chính xác và cải thiện tính cá nhân hóa, hệ thống mang lại hiệu quả và giá trị thực tiễn cao hơn so với các giải pháp hiện tại, góp phần nâng cao hiệu quả quản lý và hành chính trong giáo dục.

Bảng dưới đây tóm tắt những điểm mạnh và hạn chế chính của các giải pháp hiện có, cùng với nhóm người dùng mà mỗi giải pháp phù hợp nhất:

|  |  |  |  |
| --- | --- | --- | --- |
| Công cụ | Ưu điểm chính | Hạn chế chính | Phù hợp với ai |
| AutoHotkey | Linh hoạt, nhẹ, miễn phí | Không hỗ trợ AI hay NLP | Người dùng kỹ thuật |
| Selenium + Python | Mạnh mẽ, tích hợp AI dễ | Cần lập trình, thiết lập phức tạp | Lập trình viên, kỹ sư |
| Chrome/Apple Autofill | Dễ dùng, tích hợp sẵn | Không hiểu ngữ cảnh, không tùy biến sâu | Người dùng phổ thông |
| Power Automate | Tự động hóa quy trình phức tạp | Khó dùng với người không chuyên, tốn phí | Doanh nghiệp, tổ chức |
| RoboForm | Bảo mật cao, xử lý biểu mẫu phức tạp | Không học từ lịch sử người dùng, không AI | Nhân sự bảo mật cao |
| Fillout | Giao diện đẹp, hỗ trợ khảo sát động | Không gợi ý từ lịch sử, thiếu tính cá nhân hóa | Tổ chức khảo sát |

***Bảng 2:*** *So sánh các giải pháp*

Hệ thống được đề xuất trong nghiên cứu này mang lại nhiều lợi ích nổi bật, đặc biệt phù hợp cho người dùng trong lĩnh vực giáo dục. Trước hết, hệ thống cho phép người dùng dễ dàng tải lên các biểu mẫu ở định dạng Word (.docx) thông qua một giao diện web đơn giản, không cần cài đặt phần mềm hay có kiến thức kỹ thuật chuyên sâu. Sau đó, hệ thống sử dụng các công nghệ tiên tiến như GPT và xử lý ngôn ngữ tự nhiên để phân tích nội dung biểu mẫu, hiểu mục đích và ngữ cảnh sử dụng, từ đó đưa ra các gợi ý điền thông tin chính xác và mang tính cá nhân hóa dựa trên dữ liệu lịch sử của người dùng. Phương pháp tự động hóa thông minh này không chỉ giúp tiết kiệm thời gian, nâng cao độ chính xác mà còn cải thiện trải nghiệm người dùng nhờ các gợi ý theo ngữ cảnh.

Với khả năng tích hợp liền mạch giữa lịch sử người dùng, khả năng hiểu ngôn ngữ tự nhiên và trí tuệ nhân tạo hiện đại, hệ thống được đề xuất này tạo nên sự khác biệt rõ rệt so với các giải pháp hiện có. Đây là một hướng tiếp cận thiết thực và hiệu quả trong việc ứng dụng trí tuệ nhân tạo vào công tác hành chính giáo dục, góp phần đơn giản hóa quá trình xử lý biểu mẫu cho cả người học lẫn giảng viên, từ đó nâng cao hiệu quả quản lý giáo dục số.

# Ứng dụng công nghệ vào giáo dục

Ứng dụng hệ thống điền biểu mẫu tự động vào giáo dục là việc tích hợp công nghệ trí tuệ nhân tạo (AI) và xử lý ngôn ngữ tự nhiên (NLP) vào các quy trình điền biểu mẫu trong môi trường giáo dục nhằm tự động hóa, tối ưu hóa và đơn giản hóa công việc nhập liệu hành chính. Mục tiêu chính của việc ứng dụng này là: giảm thời gian và công sức nhập liệu thủ công cho sinh viên, giảng viên và cán bộ quản lý, tăng độ chính xác và nhất quán trong điền thông tin, gợi ý thông tin phù hợp dựa trên lịch sử cá nhân và ngữ cảnh, nâng cao hiệu quả quản lý hành chính giáo dục và góp phần vào chuyển đổi số.

Hệ thống này có thể tự động thu thập, phân tích dữ liệu cá nhân từ các lần điền trước hoặc từ hồ sơ học sinh – sinh viên, sau đó tự động điền các biểu mẫu như: đơn xin nghỉ học, đơn xin hỗ trợ học phí, phiếu khảo sát, hồ sơ đăng ký môn học, v.v., nhu cầu tối ưu hóa các hoạt động hành chính là một trong những yếu tố then chốt để nâng cao chất lượng quản lý và phục vụ người học. Việc xử lý các biểu mẫu hành chính như đơn xin nghỉ học, đăng ký học phần, phiếu khảo sát, hồ sơ sinh viên... vẫn còn được thực hiện thủ công tại nhiều cơ sở giáo dục, gây ra áp lực lớn cho cả người học lẫn cán bộ quản lý. Chính vì vậy, việc ứng dụng một hệ thống điền biểu mẫu tự động thông minh, có khả năng học hỏi và thích nghi theo lịch sử sử dụng cá nhân, mang đến một hướng đi hiệu quả nhằm giảm thiểu gánh nặng hành chính và nâng cao trải nghiệm người dùng trong môi trường giáo dục.

Hệ thống điền biểu mẫu tự động sử dụng trí tuệ nhân tạo kết hợp với các kỹ thuật xử lý ngôn ngữ tự nhiên có thể tích hợp trực tiếp vào các nền tảng quản lý giáo dục hiện hành như hệ thống quản lý học tập (LMS) hoặc hệ thống thông tin sinh viên (SIS). Khi áp dụng vào giáo dục, hệ thống này không chỉ hỗ trợ sinh viên tự động hoàn thiện các biểu mẫu quen thuộc một cách nhanh chóng, chính xác mà còn giúp giảng viên, cán bộ quản lý xử lý khối lượng lớn dữ liệu hành chính trong thời gian ngắn. Đặc biệt, bằng việc ghi nhớ và phân tích lịch sử tương tác của từng cá nhân, hệ thống có khả năng đưa ra các gợi ý điền thông tin phù hợp với ngữ cảnh, từ đó giảm thiểu tối đa sai sót và sự lặp lại không cần thiết trong quá trình nhập liệu.

Khả năng học hỏi và cập nhật liên tục từ dữ liệu đầu vào cũng cho phép hệ thống thích ứng tốt với sự thay đổi của các biểu mẫu theo thời gian hoặc yêu cầu đặc thù của từng cơ sở giáo dục. Chẳng hạn, nếu một trường đại học thường xuyên cập nhật các biểu mẫu mới phục vụ cho hoạt động ngoại khóa hoặc chương trình học bổng, hệ thống có thể nhanh chóng thích nghi và hỗ trợ người dùng điền các biểu mẫu đó mà không cần lập trình lại. Tính năng này mở ra tiềm năng ứng dụng rộng lớn không chỉ ở cấp độ trường học mà còn ở cấp hệ thống giáo dục quốc gia, nơi việc đồng bộ và xử lý biểu mẫu của hàng triệu học sinh – sinh viên là một bài toán lớn.

Ngoài ra, hệ thống còn góp phần nâng cao tính minh bạch và chuẩn hóa trong hoạt động quản lý giáo dục. Các biểu mẫu được hoàn thiện một cách nhất quán và đúng định dạng, giúp công tác lưu trữ, tìm kiếm và phân tích dữ liệu trở nên dễ dàng hơn. Trong dài hạn, việc ứng dụng hệ thống này còn có thể kết hợp với các công cụ phân tích dữ liệu để đưa ra những nhận định hỗ trợ ra quyết định trong công tác điều hành giáo dục, chẳng hạn như dự đoán nhu cầu môn học, phát hiện bất thường trong hồ sơ sinh viên hoặc đánh giá mức độ hài lòng của người học thông qua phân tích phản hồi khảo sát.

Tóm lại, việc tích hợp hệ thống điền biểu mẫu tự động vào môi trường giáo dục không chỉ giải quyết các vấn đề trước mắt về hiệu suất hành chính mà còn tạo nền tảng vững chắc để hiện đại hóa công tác quản lý, thúc đẩy sự phát triển bền vững của ngành giáo dục trong kỷ nguyên số. Đây là bước đi cần thiết nhằm đưa trí tuệ nhân tạo vào phục vụ thực tiễn một cách thiết thực, hiệu quả và có định hướng lâu dài.

# Giải thích cơ sở lý thuyết của hệ thống điền biểu mẫu tự động

Hệ thống điền biểu mẫu tự động trong nghiên cứu này được xây dựng dựa trên hai nền tảng chính: kỹ thuật so khớp tên trường (field matching) và trí tuệ nhân tạo (AI). Mục tiêu của hệ thống là giúp người dùng điền các biểu mẫu một cách nhanh chóng và chính xác bằng cách hiểu nội dung biểu mẫu và gợi ý thông tin phù hợp dựa trên lịch sử sử dụng của từng người dùng.

Trước tiên, khi người dùng tải lên một biểu mẫu Word (.docx), hệ thống sẽ quét và xác định những chỗ cần điền thông tin – thường được đánh dấu bằng các ký hiệu đặc biệt như [\_123\_]. Sau đó, hệ thống sẽ tìm xung quanh các ký hiệu đó để xác định tên của từng trường dữ liệu (ví dụ như “Họ tên”, “Ngành học”, “Lớp”,...). Để làm điều này, hệ thống sử dụng các kỹ thuật phân tích văn bản cơ bản, kết hợp với cách tổ chức bố cục trong biểu mẫu như văn bản thông thường hoặc bảng.

Sau khi biết các trường cần điền, hệ thống sử dụng nhiều phương pháp để tìm ra nội dung phù hợp. Các thuật toán được áp dụng bao gồm những cách đo độ giống nhau giữa các từ và câu, chẳng hạn như: Levenshtein Distance là đo số lần cần thay đổi để biến một từ thành từ khác, Jaccard và Cosine Similarity sẽ so sánh mức độ giống nhau giữa các cụm từ, TF-IDF xác định từ nào quan trọng trong câu, Sentence-BERT giúp hệ thống “hiểu ý” của câu thay vì chỉ so sánh từng từ riêng lẻ.

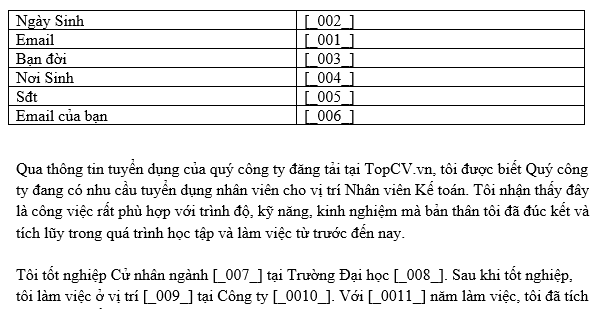
Tuy nhiên, hệ thống không chỉ dựa vào việc so sánh câu chữ. Nó còn sử dụng trí tuệ nhân tạo GPT-4 của OpenAI, một mô hình ngôn ngữ rất mạnh, có khả năng hiểu được mục đích của biểu mẫu và đề xuất nội dung phù hợp. Khi cần điền thông tin, hệ thống sẽ “hỏi” mô hình AI này bằng cách gửi nội dung như: “Biểu mẫu này là đơn xin nghỉ học, trường cần điền là ‘Lý do nghỉ’. Người dùng trước đây từng điền các lý do như ‘bị ốm’, ‘việc gia đình’...” Từ đó, AI sẽ đưa ra các gợi ý như “bị ốm” hoặc “có việc cá nhân”.

Hệ thống cũng có khả năng học hỏi từ lịch sử người dùng, nghĩa là những gì người dùng từng điền trước đây sẽ được ghi nhớ để cải thiện các gợi ý trong tương lai. Đồng thời, để tiết kiệm thời gian xử lý, hệ thống áp dụng cơ chế ghi nhớ (caching) – những kết quả gợi ý đã từng được sinh ra sẽ được lưu lại và sử dụng lại nếu gặp biểu mẫu tương tự, giúp phản hồi nhanh hơn.

Nhờ kết hợp giữa các kỹ thuật truyền thống và trí tuệ nhân tạo hiện đại, hệ thống có thể điền biểu mẫu một cách thông minh, hiểu đúng ngữ cảnh, và cá nhân hóa nội dung cho từng người dùng, ngay cả khi họ không có kiến thức chuyên sâu về công nghệ.

Để giúp hệ thống tự động điền biểu mẫu, chúng tôi xây dựng một hệ thống để tự động lấy ra các thông tin cần điền (gọi là "trường dữ liệu") từ file Word. Hệ thống này dựa vào các ký hiệu đặc biệt như [\*123\*] và các chữ xung quanh để xác định thông tin cần lấy. Hệ thống sử dụng các kỹ thuật xử lý văn bản để trích xuất tên của trường dữ liệu. Hệ thống này hoạt động theo hai bước chính: 1. Lấy thông tin từ văn bản thường: Hệ thống dùng một công cụ đặc biệt gọi là "biểu thức chính quy" để tìm các mã trường có dạng [\*123\*] trong văn bản. Khi tìm thấy mã trường, hệ thống sẽ dò ngược lại để tìm tên của trường dữ liệu theo các bước sau: bỏ qua các khoảng trắng, kiểm tra các từ chỉ ngày, tháng, năm, tìm chữ cái viết hoa đầu tiên và lấy toàn bộ cụm từ từ đó đến mã trường. Hệ thống cũng có một danh sách các từ đặc biệt để xử lý riêng, giúp xác định chính xác các trường ngày tháng năm. 2. Phân tích cấu trúc bảng: Hệ thống xem xét tất cả các bảng trong file Word và phân tích từng hàng. Hệ thống cho rằng bảng được chia làm hai cột, trong đó cột đầu tiên chứa tên của trường dữ liệu và cột thứ hai chứa mã trường có dạng [\*123\*]. Tương tự như khi xử lý văn bản thường, hệ thống dùng "biểu thức chính quy" để tìm mã trường

Bảng sau trình bày định dạng phù hợp trong tài liệu Word nhằm trích xuất tên trường hiển thị trên Web, bao gồm cả bảng biểu và đoạn văn:



***Hình 3****: Định dạng trong word*

Hệ thống kết hợp các trường được trích xuất từ văn bản thường và từ bảng, đồng thời loại bỏ các trường trùng lặp dựa trên mã trường:

Tập hợp (Set): Hệ thống sử dụng cấu trúc dữ liệu tập hợp (set) để theo dõi các mã trường đã xử lý, giúp loại bỏ trùng lặp hiệu quả với độ phức tạp O(1).

Sắp xếp theo vị trí: Hệ thống sắp xếp các trường theo vị trí từ trên xuống dưới trong tài liệu, giúp duy trì thứ tự logic của các trường.

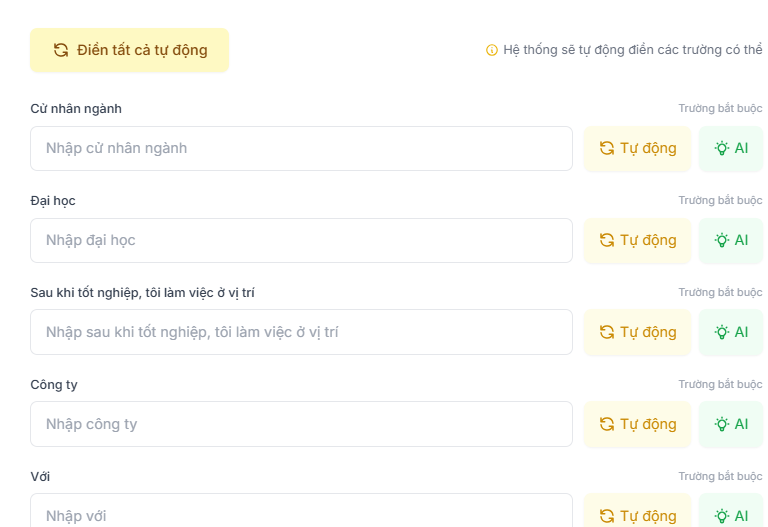
Ưu tiên trường xuất hiện đầu tiên: Khi có nhiều trường có cùng mã trường, hệ thống ưu tiên trường xuất hiện đầu tiên trong tài liệu.

Ứng dụng trong hệ thống gợi ý giá trị

Sau khi trích xuất được các tên trường từ tài liệu Word, hệ thống sẽ sử dụng các phương pháp so khớp tên trường và gợi ý giá trị đã được mô tả trong các phần trước để đề xuất giá trị phù hợp cho từng trường.

Việc trích xuất chính xác tên trường là bước đầu tiên và quan trọng trong quy trình gợi ý giá trị tự động, vì nó ảnh hưởng trực tiếp đến độ chính xác của các gợi ý được đưa ra.

Sau khi áp dụng các kỹ thuật trích xuất, hình ảnh dưới đây minh họa kết quả trích xuất tên trường được hiển thị trên giao diện Web:



***Hình 4****: Upload tài liệu docx lên Web*

Phương pháp này được triển khai trong lớp EnhancedFieldMatcher và sử dụng nhiều thuật toán so khớp văn bản khác nhau để xác định độ tương đồng giữa các trường. Các thuật toán so khớp văn bản Levenshtein Distance (Edit Distance): Đây là thuật toán đo lường khoảng cách giữa hai chuỗi bằng cách đếm số thao tác tối thiểu (chèn, xóa, thay thế) cần thiết để biến đổi một chuỗi thành chuỗi khác. Công thức toán học: Gọi a và b là hai chuỗi cần so sánh, độ dài lần lượt là |a| và |b|. Khoảng cách Levenshtein lev(a,b) được tính như sau:

Levenshtein Distance đo số lần cần thay đổi để biến một từ thành từ khác. Gọi a và b là hai chuỗi cần so sánh, độ dài lần lượt là ∣a∣ và ∣b∣. Khoảng cách Levenshtein lev(a,b) được tính như sau:

lev(a,b) =

max(|a|, |b|) nếu min(|a|, |b|) = 0

min(

lev(a[1:], b) + 1, // xóa

lev(a, b[1:]) + 1, // chèn

lev(a[1:], b[1:]) + (a[0] != b[0]) // thay thế

)

nếu min(|a|, |b|) > 0​

Trong hệ thống, nhiều thuật toán đo độ tương đồng văn bản được kết hợp để nâng cao độ chính xác trong việc so khớp nội dung. Cụ thể, thuật toán difflib.SequenceMatcher của Python được sử dụng để so sánh độ tương đồng giữa hai chuỗi ký tự dựa trên các đoạn con trùng khớp. Jaccard Similarity đo lường độ tương đồng giữa hai tập hợp bằng cách tính tỷ lệ giữa kích thước giao và hợp của hai tập. Trong ngữ cảnh so khớp văn bản, hai chuỗi được chuyển thành tập hợp các từ, sau đó áp dụng công thức:

J(A, B) = |A ∩ B| / |A ∪ B|

Trong đó A và B là tập hợp các từ trong hai văn bản, |A ∩ B| là số từ xuất hiện ở cả hai văn bản, và |A ∪ B| là tổng số từ không trùng lặp trong cả hai văn bản. Cosine Similarity đo lường góc giữa hai vector trong không gian nhiều chiều, thể hiện mức độ tương đồng giữa hai văn bản được biểu diễn thành vector tần suất từ, với công thức:

cos(θ) = (A ⋅ B) / (||A|| × ||B||)

Trong đó A ⋅ B là tích vô hướng của hai vector, ||A|| và ||B|| là độ dài Euclidean. Hệ thống sử dụng hàm cosine\_similarity từ thư viện sklearn.metrics.pairwise để tính toán chỉ số này. TF-IDF (Term Frequency-Inverse Document Frequency) là một chỉ số đánh giá tầm quan trọng của một từ trong một văn bản, so với toàn bộ tập văn bản, được tính bằng công thức:

tf-idf(t, d, D) = tf(t, d) × idf(t, D)

idf(t, D) = log(N / df(t))

Với tf(t, d) là tần suất xuất hiện của từ t trong văn bản d, N là tổng số văn bản và df(t) là số văn bản chứa từ t. Vector TF-IDF được tính bằng thư viện sklearn.feature\_extraction.text.TfidfVectorizer. Sentence Transformers sử dụng mô hình học sâu để biểu diễn các câu dưới dạng vector ngữ nghĩa. Hệ thống sử dụng mô hình all-MiniLM-L6-v2 từ thư viện sentence\_transformers để mã hóa văn bản và tính độ tương đồng giữa hai câu bằng công thức:

similarity = cos(encode(text\_1), encode(text\_2))

Trong đó encode() là hàm chuyển văn bản thành vector embedding. Ngoài ra, hệ thống sử dụng phương pháp kết hợp (ensemble) nhiều thuật toán để tăng độ chính xác của việc so khớp nội dung. Các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) được áp dụng nhằm nâng cao chất lượng dữ liệu đầu vào, bao gồm chuẩn hóa mã Unicode về dạng chuẩn NFC, loại bỏ dấu câu và các ký tự đặc biệt nhưng vẫn giữ nguyên dấu tiếng Việt, cũng như chuẩn hóa từ đồng nghĩa bằng cách ánh xạ các cách viết khác nhau của cùng một khái niệm về một dạng thống nhất thông qua từ điển đồng nghĩa.

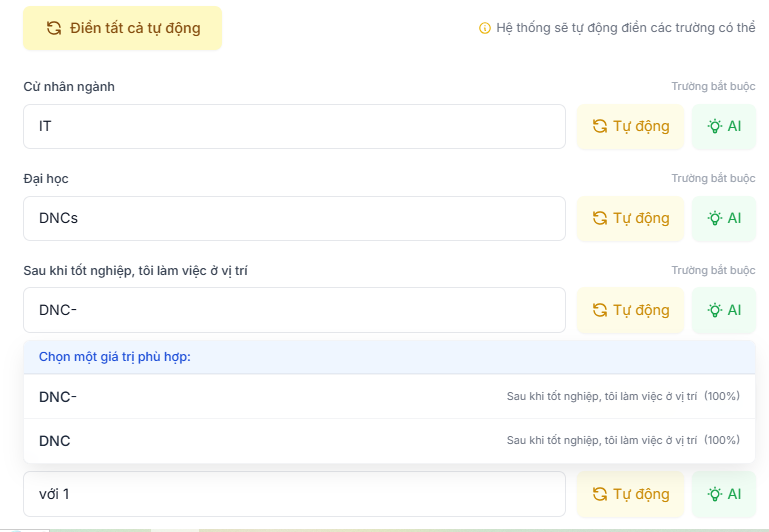
Bên cạnh đó, Hệ thống này sử dụng trí tuệ nhân tạo GPT-4 của OpenAI, được triển khai trong lớp AIFieldMatcher, để tự động gợi ý nội dung phù hợp khi người dùng điền vào các biểu mẫu. Khi cần điền một trường cụ thể (ví dụ: “Ngành học”), hệ thống sẽ gửi đến GPT-4 ba thông tin chính gồm tên trường cần điền, danh sách các giá trị mà người dùng đã từng nhập trước đó (chẳng hạn như “Khoa học máy tính” hoặc “Công nghệ thông tin”), và ngữ cảnh sử dụng của biểu mẫu (ví dụ: giáo dục, nghỉ học, khảo sát, v.v.). Dựa trên các dữ liệu này, GPT-4 sẽ hiểu được ý định và hoàn cảnh sử dụng biểu mẫu, từ đó tạo ra các gợi ý thông minh và phù hợp. Ví dụ, nếu người dùng từng nhập “Khoa học máy tính” và “Công nghệ thông tin” cho trường “Ngành học”, trong một biểu mẫu thuộc lĩnh vực giáo dục, GPT-4 có thể đề xuất lại các ngành đó để người dùng chọn, đồng thời đưa ra một giá trị mặc định. Các gợi ý này được hiển thị trên giao diện để người dùng có thể chọn hoặc chỉnh sửa lại nếu cần. Nhờ vào cơ chế này, việc điền biểu mẫu trở nên nhanh chóng, thuận tiện và mang tính cá nhân hóa cao.

Hệ thống vận hành dựa trên mô hình ngôn ngữ lớn GPT-4 – một mô hình được huấn luyện từ lượng lớn văn bản và có khả năng hiểu ngữ cảnh rất tốt. Để định hướng cho GPT-4 tạo ra các gợi ý chính xác, hệ thống sử dụng kỹ thuật gọi là "Prompt Engineering", tức là thiết kế các câu lệnh đầu vào thật cụ thể và rõ ràng. Ví dụ, nếu trường cần gợi ý là “Ngành học”, hệ thống sẽ tạo prompt có nội dung như “Bạn là một hệ thống gợi ý biểu mẫu thông minh... Trường cần gợi ý: 'Ngành học'. Lịch sử người dùng từng nhập: Khoa học máy tính, Công nghệ thông tin. Biểu mẫu này thuộc ngữ cảnh: giáo dục...” và yêu cầu mô hình tạo ra các gợi ý phù hợp. Ngoài ra, để hiểu rõ hơn biểu mẫu mà người dùng đang sử dụng, hệ thống còn có thể phân tích toàn bộ nội dung biểu mẫu thông qua GPT-4 nhằm xác định mục đích sử dụng, đối tượng sử dụng, các trường thông tin chính và ngữ cảnh (như hành chính, pháp lý, giáo dục...). Tất cả thông tin phân tích được hệ thống tóm tắt lại để làm cơ sở cho việc sinh gợi ý.

Sau khi GPT-4 trả về kết quả, hệ thống sẽ dùng các biểu thức chính quy (regex) để trích xuất phần gợi ý và giá trị mặc định. Ví dụ, nếu mô hình trả về chuỗi như “Gợi ý: [Khoa học máy tính, Công nghệ thông tin], Mặc định: Công nghệ thông tin”, hệ thống sẽ tách riêng từng gợi ý và lưu lại giá trị mặc định để hiển thị cho người dùng. Với cách tiếp cận này, hệ thống không chỉ giúp rút ngắn thời gian nhập liệu mà còn mang đến trải nghiệm hiện đại, cá nhân hóa, và dễ sử dụng, kể cả với người không có chuyên môn kỹ thuật.

Hệ thống sử dụng cơ chế caching để tối ưu hiệu suất, "Caching" là kỹ thuật lưu tạm thời kết quả xử lý trước đó để tăng tốc độ phản hồi cho những lần sử dụng tiếp theo. Ví dụ, nếu người dùng vừa nhập một biểu mẫu có trường “Tên trường đại học”, hệ thống sẽ lưu lại gợi ý tương ứng. Khi gặp biểu mẫu tương tự, thay vì gọi lại AI hoặc tính toán lại từ đầu, hệ thống sẽ sử dụng kết quả đã lưu sẵn để trả về nhanh hơn.

Hệ thống kết hợp cả hai hướng – vừa so khớp tên trường, vừa dùng AI – để đưa ra gợi ý thông minh và cá nhân hóa. Các yếu tố ảnh hưởng đến kết quả gợi ý bao gồm độ tương đồng giữa các trường (sử dụng các thuật toán so khớp văn bản), tần suất sử dụng (các giá trị được sử dụng nhiều lần có trọng số cao hơn), thời gian sử dụng (các giá trị gần đây có trọng số cao hơn), ngữ cảnh biểu mẫu (được trích xuất từ nội dung biểu mẫu), và sở thích người dùng (dựa trên lịch sử sử dụng của từng người dùng).



***Hình 4****: chức năng điền tự động*

Nhờ kết hợp giữa các kỹ thuật truyền thống và trí tuệ nhân tạo hiện đại, hệ thống có thể điền biểu mẫu một cách thông minh, hiểu đúng ngữ cảnh, và cá nhân hóa nội dung cho từng người dùng, ngay cả khi họ không có kiến thức chuyên sâu về công nghệ. Hệ thống cũng xây dựng phương pháp trích xuất tự động các trường cần điền từ tài liệu Word dựa vào ký hiệu đặc trưng [123] và văn bản ngữ cảnh xung quanh. Hệ thống sử dụng các kỹ thuật xử lý văn bản để trích xuất tên trường từ tài liệu Word, bao gồm trích xuất trường từ văn bản thường (sử dụng biểu thức chính quy \d+ và thuật toán tìm ngược), phân tích cấu trúc bảng, kết hợp và loại bỏ trùng lặp (sử dụng tập hợp và sắp xếp theo vị trí), và ứng dụng trong hệ thống gợi ý giá trị. Việc trích xuất chính xác tên trường là bước đầu tiên và quan trọng trong quy trình gợi ý giá trị tự động, vì nó ảnh hưởng trực tiếp đến độ chính xác của các gợi ý được đưa ra.

# Kết luận

Nghiên cứu đã trình bày và hiện thực hóa một cách tiếp cận mới trong việc tự động hóa quy trình điền biểu mẫu hành chính thông qua hệ thống tích hợp trí tuệ nhân tạo, đặc biệt là các kỹ thuật xử lý ngôn ngữ tự nhiên và mô hình học sâu. Với mục tiêu nâng cao hiệu quả, độ chính xác và tính cá nhân hóa trong xử lý biểu mẫu, hệ thống đề xuất không chỉ giúp rút ngắn thời gian thao tác cho người dùng mà còn cải thiện đáng kể chất lượng dữ liệu được nhập vào. Thông qua quá trình triển khai thực nghiệm, hệ thống cho thấy khả năng học hỏi từ lịch sử sử dụng để đưa ra gợi ý điền nội dung thông minh và phù hợp theo ngữ cảnh, đồng thời xử lý tốt các biểu mẫu có cấu trúc đa dạng và độ phức tạp khác nhau. Việc áp dụng các phương pháp so khớp trường dữ liệu như Cosine Similarity, Jaccard Index, và các kỹ thuật embedding hiện đại như Word2Vec, BERT, hay Sentence-BERT đã góp phần nâng cao độ chính xác trong việc xác định tương quan ngữ nghĩa giữa các nhãn trường và dữ liệu nguồn. Bên cạnh đó, nghiên cứu cũng đã phân tích kỹ lưỡng các hệ thống và công cụ điền biểu mẫu hiện có trên thị trường, từ mã nguồn mở đến các sản phẩm thương mại, qua đó khẳng định vị trí và tiềm năng vượt trội của hệ thống được đề xuất trong việc cá nhân hóa và mở rộng ứng dụng thực tiễn. Để kiểm chứng hiệu quả thực tế của hệ thống, nhóm nghiên cứu đã tiến hành thử nghiệm trên một tập dữ liệu gồm 150 biểu mẫu hành chính thực tế được sử dụng tại một trường đại học tư thục. Kết quả cho thấy hệ thống đạt tỷ lệ chính xác trung bình khi gợi ý đúng thông tin là 91.2%, trong đó các trường phổ biến như “Họ tên”, “Ngành học”, “Lớp”, “Mã số sinh viên” có độ chính xác trên 95%. Thời gian trung bình để hoàn thành một biểu mẫu sau khi hệ thống gợi ý là 1 phút 47 giây, giảm gần 80% thời gian so với nhập tay truyền thống (trung bình 8 phút). Ngoài ra, khảo sát nhanh trên 35 người dùng thử cho thấy 94% đánh giá hệ thống dễ sử dụng, và 88% cho biết họ sẵn sàng sử dụng hệ thống nếu được triển khai thực tế trong trường học. Tuy nhiên, hệ thống vẫn tồn tại một số hạn chế, đặc biệt là về hiệu năng xử lý ở quy mô lớn, khả năng thích ứng với biểu mẫu hoàn toàn mới hoặc dữ liệu không chuẩn hóa. Trong tương lai, việc tích hợp học liên tục (continual learning), tối ưu hóa hiệu suất, tăng cường bảo mật dữ liệu và khả năng tương thích đa nền tảng sẽ là những hướng phát triển cần thiết nhằm hoàn thiện hệ thống một cách toàn diện. Với những kết quả đã đạt được, nghiên cứu này không chỉ cung cấp một giải pháp công nghệ thiết thực cho ngành giáo dục mà còn mở ra triển vọng ứng dụng rộng rãi trong các lĩnh vực hành chính công, tài chính, y tế và doanh nghiệp, góp phần vào tiến trình chuyển đổi số bền vững trong kỷ nguyên trí tuệ nhân tạo.

**TÀI LIỆU THAM KHẢO**

De Luca, A., et al. (2007). “The impact of autofill on the usability of mobile form filling.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. https://www.medien.informatik.uni-muenchen.de/pubdb/publications/pub/deluca2007pmc/deluca2007pmc.pdf.

Palm, R. B., et al. (2020). “CloudScan: A configuration-free invoice analysis system using RNNs.” IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(4), 1169–1185. https://doi.org/10.1109/TPAMI.2019.2944806.

Reimers, Nils, and Iryna Gurevych. “Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks.” \*Proceedings of EMNLP-IJCNLP 2019\*, 2019, pp. 3982–3992. [Online] Available: https://doi.org/10.18653/v1/D19-1410.

Chen, Xiang, et al. “FormNet: Structural Encoding via Sequential Pattern for Form Information Extraction.” \*Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics\*, 2021, pp. 2425–2435. [Online] Available: https://doi.org/10.18653/v1/2021.acl-long.188.

Devlin, J., et al. (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” NAACL-HLT 2019. https://doi.org/10.18653/v1/N19-1423.

Zawacki-Richter, Olaf, et al. “Systematic Review of Research on Artificial Intelligence Applications in Higher Education.” \*International Journal of Educational Technology in Higher Education\*, vol. 16, 2019, pp. 1–27. [Online] Available: https://doi.org/10.1186/s41239-019-0171-0.

Brown, T. B., et al. (2020). “Language Models Are Few-Shot Learners.” NeurIPS 2020, 33, 1877–1901. https://arxiv.org/abs/2005.14165.

Reimers, N., & Gurevych, I. (2019). “Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks.” EMNLP-IJCNLP 2019. <https://doi.org/10.18653/v1/D19-1410>.

**Thông tin về tác giả**

|  |  |
| --- | --- |
| 81f5efd849befce0a5af | Nguyễn Tuấn Đạt:  Sinh viên năm tư ngành Khoa học máy, Trường Đại học Nam Cần Thơ.  Thực tập tại công ty CÔNG TY TNHH MỘT THÀNH VIÊN CÔNG NGHỆ KỸ THUẬT TIÊN PHONG.  Lĩnh vực quan tâm: Khoa học máy tính, Trí tuệ nhân tạo.  Điện thoại:0353730323 |
| b772f6400620b37eea31 | Nguyễn Anh Duy:  Cử nhân ngành CNTT (Trường Đại học Nam Cần Thơ). Giám đốc dự án Adhightech Ltd., Vietnam.  Lĩnh vực quan tâm: máy học (ML), công nghệ AI.  Liên lạc: nguyenanhduy@adhigtechn.com |
| thaynam7 | ThS. Trần Thanh Nam:  Giảng viên khoa Công nghệ thông tin, Trường Đại Học Nam Cần Thơ  Điện thoại:0834734099 |
| 2 | ThS. Trần Văn Thiện:  Giảng viên khoa công nghệ thông tin, Trường Đại Học Nam Cần Thơ  Điện thoại:0902091071 |