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

Nguyễn Tuấn Đạt, Khoa Học Máy Tính, ntdat1610c@gmail.com

# Tóm tắt

*Bài báo này tập trung nghiên cứu và phát triển một hệ thống điền biểu mẫu tự động dựa trên lịch sử người dùng, ứng dụng trong lĩnh vực giáo dục nhằm giải quyết vấn đề lặp lại thao tác điền thông tin thủ công và tránh sai sót trong các biểu mẫu hành chính như đơn xin nghỉ học, đơn đăng ký học phần, hồ sơ học sinh – sinh viên, phiếu khảo sát, v.v. Đối tượng nghiên cứu là các cá nhân trong môi trường giáo dục, bao gồm học sinh, sinh viên, giảng viên và cán bộ quản lý. Phạm vi nghiên cứu được giới hạn trong việc xử lý các biểu mẫu định dạng DOCX, với khả năng mở rộng sang các định dạng khác trong tương lai.*

*Hệ thống đề xuất sử dụng mô hình học máy kết hợp TF-IDF, cosine similarity, fuzzy matching với mô hình ngữ nghĩa hiện đại (Sentence-BERT) để tăng độ chính xác khi so khớp tên trường trong các biểu mẫu và tích hợp mô hình ngôn ngữ của OpenAI để gợi ý và tự động hoàn thiện nội dung phù hợp với ngữ cảnh. Phương pháp nghiên cứu bao gồm: thu thập dữ liệu biểu mẫu thực tế, tiền xử lý và chuẩn hóa thông tin, xây dựng thuật toán so khớp các trường dữ liệu tương tự, và huấn luyện hệ thống đề xuất nội dung nâng cao dựa trên lịch sử cá nhân.*

*Phương pháp nghiên cứu bao gồm khảo sát thực tế tại các cơ sở giáo dục, mô hình hóa quá trình trích xuất và học dữ liệu từ lịch sử người dùng, cùng thử nghiệm khả năng đề xuất nội dung nâng cao bằng mô hình ngôn ngữ. Kết quả cho thấy hệ thống rút ngắn đáng kể thời gian thao tác (trung bình giảm 70–85%), đồng thời nâng cao chất lượng và độ chính xác của nội dung được điền. Người dùng cũng đánh giá cao tính thân thiện và khả năng tự học thích nghi của hệ thống.*

*Mặc dù hệ thống còn đang trong giai đoạn hoàn thiện, bài báo này đóng góp góc nhìn thiết thực vào tiềm năng của AI trong tự động hóa công việc giáo dục. Các hướng nghiên cứu tương lai bao gồm việc mở rộng hệ thống sang các định dạng biểu mẫu khác, tích hợp API của các hệ thống giáo dục hiện có (như LMS/SIS), và phát triển cơ chế học liên tục để nâng cao khả năng gợi ý theo thời gian. Qua đó, nghiên cứu này kỳ vọng sẽ thúc đẩy một cách tiếp cận thông minh và hiệu quả hơn trong quản lý giáo dục bằng công nghệ AI..*

**Keywords:** artificial intelligence, AI tools, education, educational technology, human-machine.

# Giới thiệu về hệ thống

Trong bối cảnh chuyển đổi số đang diễn ra mạnh mẽ trong nhiều lĩnh vực, đặc biệt là giáo dục, nhu cầu tối ưu hóa các quy trình hành chính đã trở nên cấp thiết nhằm nâng cao hiệu suất làm việc và giảm thiểu thao tác thủ công. Một trong những thách thức thường gặp trong môi trường giáo dục là việc điền các biểu mẫu hành chính như đơn xin nghỉ học, đơn đăng ký học phần, phiếu khảo sát, hồ sơ sinh viên, v.v. Những biểu mẫu này thường có cấu trúc lặp lại, yêu cầu nhập các thông tin quen thuộc nhưng vẫn phải thực hiện thủ công, gây tốn thời gian và dễ xảy ra sai sót.

Để giải quyết vấn đề này, nghiên cứu đề xuất và phát triển Hệ thống điền biểu mẫu tự động ứng dụng AI dựa trên lịch sử người dùng (AI-Powered Automatic Form-Filling System Based on User History). Đây là một hệ thống thông minh hỗ trợ tự động hóa quá trình điền biểu mẫu điện tử (định dạng DOCX) bằng cách sử dụng các công nghệ trí tuệ nhân tạo, học máy và xử lý ngôn ngữ tự nhiên.

Hệ thống ứng dụng các kỹ thuật như TF-IDF, độ đo Cosine, Word2Vec, Sentence-BERT để xác định mức độ tương đồng giữa các trường thông tin trên biểu mẫu và dữ liệu lịch sử của người dùng. Ngoài ra, với sự hỗ trợ từ các mô hình ngôn ngữ lớn (LLMs) như GPT, hệ thống có thể hiểu được ngữ cảnh trong các trường thông tin mở và đưa ra nội dung gợi ý phù hợp. Đặc biệt, hệ thống có khả năng học liên tục từ hành vi và tương tác của người dùng, giúp tăng mức độ cá nhân hóa và độ chính xác theo thời gian.

Hướng phát triển hiện tại tập trung vào việc xử lý các biểu mẫu ở định dạng DOCX, tuy nhiên hệ thống có khả năng mở rộng sang các định dạng khác như PDF, HTML hoặc tích hợp với các hệ thống quản lý học tập (LMS) và hệ thống quản lý thông tin sinh viên (SIS) trong tương lai.

## Ứng dụng của hệ thống trong giáo dục

Trong bối cảnh chuyển đổi số diễn ra mạnh mẽ ở mọi lĩnh vực, giáo dục đang đứng trước yêu cầu cấp thiết phải thay đổi cả về phương pháp giảng dạy, quản lý lẫn cách tiếp cận người học. Trong đó, hoạt động quản trị hành chính trong các cơ sở giáo dục vốn vẫn phụ thuộc nhiều vào thao tác thủ công đang trở thành điểm nghẽn cản trở hiệu quả vận hành và trải nghiệm người dùng. Các biểu mẫu hành chính với cấu trúc lặp lại, khối lượng lớn, và yêu cầu độ chính xác cao là một ví dụ điển hình cho những quy trình có thể và nên được tự động hóa bằng công nghệ.

Việc ứng dụng trí tuệ nhân tạo (AI), đặc biệt là các mô hình ngôn ngữ tiên tiến như GPT, vào lĩnh vực điền biểu mẫu là một hướng tiếp cận mới, vừa phù hợp với nhu cầu thực tiễn, vừa có giá trị khoa học trong việc kết hợp giữa xử lý ngôn ngữ tự nhiên (NLP), học máy (machine learning), và tối ưu trải nghiệm người dùng. Hệ thống ứng dụng đồng thời các phương pháp như TF-IDF, Cosine Similarity, Word2Vec và Sentence-BERT nhằm nâng cao khả năng nhận diện và so sánh tên trường thông tin trong biểu mẫu, tăng tính linh hoạt và độ chính xác khi xử lý các biểu mẫu khác nhau. Tích hợp khả năng hiểu ngữ cảnh từ mô hình ngôn ngữ lớn GPT được sử dụng để hỗ trợ việc tạo nội dung cho các trường thông tin mở, giúp biểu mẫu không chỉ được điền chính xác theo dữ liệu cũ mà còn phù hợp với ngữ cảnh hiện tại của người dùng. Khả năng học hỏi và cá nhân hóa theo thời gian. Thay vì chỉ đưa ra gợi ý cố định, hệ thống có khả năng học từ những lựa chọn trước đó của người dùng để dần cải thiện độ phù hợp của nội dung đề xuất, hướng đến một hệ thống ngày càng thông minh và hiệu quả hơn.

Các hoạt động hành chính đóng vai trò quan trọng trong việc đảm bảo vận hành hiệu quả các quy trình học tập, giảng dạy và quản lý. Tuy nhiên, phần lớn các biểu mẫu liên quan như đơn xin nghỉ học, đơn yêu cầu xác nhận, biểu mẫu khảo sát, hồ sơ sinh viên,... đều yêu cầu người dùng phải tự điền tay theo từng trường thông tin cụ thể. Điều này không chỉ mất thời gian mà còn dễ dẫn đến các sai sót không đáng có.

Đối tượng nghiên cứu tập trung vào các cá nhân hoạt động trong môi trường giáo dục – từ học sinh, sinh viên đến giảng viên và cán bộ quản lý. Nghiên cứu hiện tại giới hạn trong việc xử lý biểu mẫu DOCX nhưng có định hướng mở rộng sang các định dạng phổ biến khác như PDF, HTML hoặc tích hợp trực tiếp vào các hệ thống LMS/SIS trong tương lai.

Với mục tiêu hướng đến một nền quản lý giáo dục thông minh, cá nhân hóa và hiệu quả hơn, nghiên cứu này không chỉ góp phần tiết kiệm thời gian và giảm tải khối lượng công việc hành chính, mà còn mở ra một hướng đi mới cho việc ứng dụng các mô hình ngôn ngữ lớn (Large Language Models – LLMs) như GPT vào các tình huống thực tiễn trong giáo dục, nghiên cứu này không chỉ góp phần giải quyết bài toán tiết kiệm thời gian và giảm tải công việc hành chính, mà còn mở ra một hướng đi tiềm năng cho các ứng dụng AI thực tiễn trong giáo dục. Tóm lại, nghiên cứu này góp phần tạo ra một công cụ hỗ trợ quản lý giáo dục thông minh, giúp giảm thiểu gánh nặng hành chính, nâng cao trải nghiệm người dùng và thúc đẩy xu hướng ứng dụng AI trong các quy trình hành chính vốn đang bị bỏ ngỏ trong ngành giáo dục.

# nghiên cứu

Hệ thống điền biểu mẫu tự động dựa trên lịch sử người dùng là giải pháp ứng dụng trí tuệ nhân tạo (AI) và xử lý ngôn ngữ tự nhiên (NLP) nhằm tự động hóa quá trình nhập liệu vào các biểu mẫu (đặc biệt là file .docx). Hệ thống này phân tích lịch sử điền biểu mẫu của người dùng, nhận diện các trường dữ liệu tương tự, từ đó tự động điền hoặc gợi ý thông tin phù hợp, giúp tiết kiệm thời gian và giảm sai sót.

## Các công nghệ nền tảng

2.1.1. Xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP)

NLP đóng vai trò trung tâm trong việc giúp hệ thống hiểu và phân tích nội dung biểu mẫu cũng như thông tin được người dùng nhập trước đó. Các kỹ thuật nổi bật bao gồm:

Tách từ và chuẩn hóa dữ liệu: Chuyển đổi văn bản tự do thành dạng dữ liệu chuẩn hóa để xử lý dễ dàng hơn.

Nhận diện thực thể có tên (Named Entity Recognition - NER): Xác định các thành phần như tên người, địa chỉ, ngày tháng... trong nội dung biểu mẫu.

Phân loại và gán nhãn trường dữ liệu (Field Classification & Labeling): Xác định vai trò của từng trường biểu mẫu (ví dụ: "Họ và tên", "Địa chỉ email").

So khớp ngữ nghĩa (Semantic Matching): Giúp nhận ra sự tương đồng về ý nghĩa giữa các biểu mẫu khác nhau dù sử dụng từ ngữ khác nhau.

2.1.2. So khớp trường dữ liệu (Field Matching Algorithms)

Để tự động hóa việc điền biểu mẫu, hệ thống cần xác định chính xác các trường dữ liệu tương ứng giữa các biểu mẫu. Việc này đòi hỏi:

So khớp dựa trên từ khóa (Keyword Matching): So sánh trực tiếp tên các trường.

So khớp dựa trên ngữ nghĩa (Semantic Similarity): Áp dụng các mô hình vector hóa văn bản như Word2Vec, BERT để đánh giá độ tương đồng về ngữ nghĩa.

So khớp dựa trên lịch sử (Historical Matching): So sánh với các biểu mẫu đã điền trước đó để xác định cách người dùng thường điền thông tin.

2.1.3. Tích hợp Trí tuệ nhân tạo (AI) và các mô hình ngôn ngữ lớn (LLM)

Các mô hình như GPT của OpenAI, BERT của Google, hoặc Claude, LLaMA đã mở ra khả năng sinh văn bản, gợi ý dữ liệu và hiểu ngữ cảnh vượt trội. Hệ thống điền biểu mẫu thông minh có thể tận dụng:

Sinh dữ liệu tự động dựa trên ngữ cảnh biểu mẫu.

Tự học và điều chỉnh gợi ý dựa trên hành vi người dùng.

Gợi ý thông minh các trường chưa có trong lịch sử nhưng có thể suy luận từ ngữ cảnh.

Việc tích hợp AI giúp hệ thống đạt đến mức độ tự động hóa cao, không chỉ giúp điền biểu mẫu chính xác mà còn linh hoạt với biểu mẫu mới.

## Lịch sử nghiên cứu và các công trình liên quan

2.2.1. Các nghiên cứu về tự động điền biểu mẫu

Từ đầu những năm 2000, nhiều công trình nghiên cứu đã xuất hiện xoay quanh việc xây dựng hệ thống điền biểu mẫu tự động, ban đầu dựa vào rule-based system, sau đó phát triển sang machine learning (ML) và deep learning (DL):

Liu et al. (2004) nghiên cứu cách hệ thống nhận diện các trường dữ liệu từ giao diện HTML, từ đó tự động điền dựa trên bộ từ điển được xây dựng trước.

Kumar et al. (2012) triển khai mô hình sử dụng Naïve Bayes để dự đoán giá trị của các trường trong biểu mẫu.

Chen et al. (2018) áp dụng deep learning để học hành vi điền biểu mẫu của người dùng từ dữ liệu lịch sử và đạt được độ chính xác cao hơn các phương pháp truyền thống.

2.2.2. Ứng dụng NLP và AI trong điền biểu mẫu

Các công trình gần đây đi theo hướng kết hợp NLP và AI để xử lý các biểu mẫu có cấu trúc phức tạp như .pdf, .docx, và thậm chí là hình ảnh scan:

Google Smart Fill (2020) là ví dụ tiêu biểu khi sử dụng NLP và AI để tự động điền thông tin dựa trên nội dung bảng tính.

Các mô hình ngôn ngữ lớn (LLM) như GPT-3 được chứng minh có khả năng phân tích cấu trúc văn bản phi cấu trúc và đưa ra gợi ý điền dữ liệu sát với nhu cầu thực tế người dùng.

2.2.3. So sánh các phương pháp tiếp cận

|  |  |  |
| --- | --- | --- |
| 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 |

## 2.4. Xu hướng phát triển trong tương lai

Dựa trên những nghiên cứu hiện tại và tiềm năng của công nghệ AI, xu hướng trong tương lai của hệ thống điền biểu mẫu tự động có thể bao gồm:

Cá nhân hóa gợi ý: Hệ thống học sâu từ hành vi người dùng và tự động điều chỉnh gợi ý phù hợp theo từng cá nhân.

Tích hợp mô hình AI thế hệ mới: GPT-4, LLaMA, Claude và các mô hình multimodal cho khả năng xử lý văn bản, hình ảnh, âm thanh cùng lúc.

Triển khai đa nền tảng: Hệ thống hỗ trợ đồng bộ trên web, mobile và desktop, tương thích với nhiều định dạng biểu mẫu (DOCX, PDF, HTML...).

Học liên tục (Continual Learning): Cho phép hệ thống cập nhật và thích nghi với biểu mẫu và hành vi mới mà không cần huấn luyện lại toàn bộ.

Đảm bảo an toàn và bảo mật dữ liệu: Ứng dụng các biện pháp mã hóa, lưu trữ phân tán và xác thực người dùng để đảm bảo quyền riêng tư.

# Phân tích, đánh giá và so sánh các hệ thống điền biểu mẫu tự động hiện có

## 3.1. Các giải pháp mã nguồn mở

3.1.1. Docassemble

Tính năng nổi bật: Nền tảng mã nguồn mở hỗ trợ tạo hệ thống hỏi đáp, điền biểu mẫu tự động dựa trên kịch bản; tích hợp tốt với NLP, có thể mở rộng bằng Python và YAML.

Ưu điểm: Miễn phí, cộng đồng phát triển năng động, tài liệu hướng dẫn đầy đủ, khả năng tích hợp với nhiều công cụ như PostgreSQL, Redis, API bên ngoài.

Nhược điểm: Yêu cầu kiến thức kỹ thuật cao để cài đặt và tùy biến; giao diện chưa thân thiện với người dùng không chuyên.

Ứng dụng thực tiễn: Thường được sử dụng trong lĩnh vực pháp lý, khảo sát và hành chính công.

3.1.2. OpenFaaS Form Auto-Filler

Tính năng nổi bật: Ứng dụng kiến trúc serverless để tự động hóa quá trình điền biểu mẫu, hoạt động nhẹ, dễ tích hợp vào quy trình CI/CD.

Ưu điểm: Khả năng mở rộng cao, tối ưu tài nguyên hệ thống, thích hợp triển khai đám mây hoặc container.

Nhược điểm: Tài liệu hỗ trợ hạn chế, nhiều tính năng nâng cao cần tự phát triển hoặc tích hợp thủ công.

Mức độ phù hợp: Phù hợp với các nhóm kỹ thuật có nhu cầu tích hợp hệ thống tùy biến cao.

3.1.3. Các thư viện Python (python-docx, pdfplumber, PyMuPDF…)

Tính năng nổi bật: Cung cấp các hàm API hỗ trợ đọc, chỉnh sửa, trích xuất và chèn dữ liệu vào file .docx và .pdf.

Ưu điểm: Linh hoạt, dễ sử dụng, có thể xây dựng hệ thống tùy chỉnh từ đầu.

Nhược điểm: Không có giao diện đồ họa sẵn, yêu cầu xây dựng toàn bộ logic nghiệp vụ và mô hình AI phụ trợ.

Tính mở rộng: Rất cao khi kết hợp cùng thư viện NLP và mô hình học máy khác.

## 3.3. Các giải pháp mã nguồn đóng và sản phẩm thương mại

3.3.1. Adobe Acrobat Pro DC

Tính năng: Tự động nhận diện trường biểu mẫu, hỗ trợ điền dữ liệu, chữ ký điện tử, xác thực định dạng dữ liệu.

Ưu điểm: Đầy đủ tính năng, hỗ trợ tốt nghiệp vụ hành chính và doanh nghiệp.

Nhược điểm: Chi phí sử dụng cao, khó tùy biến chức năng, phụ thuộc vào hệ sinh thái Adobe.

Độ tin cậy: Cao, bảo mật tốt, phù hợp với doanh nghiệp lớn.

3.3.2. Kofax Power PDF

Tính năng: Hỗ trợ AI nhận diện trường dữ liệu, chuyển đổi định dạng linh hoạt (PDF sang Word, Excel...), tự động hóa điền biểu mẫu.

Ưu điểm: Giao diện dễ sử dụng, hỗ trợ nhiều chức năng bảo mật như mã hóa, phân quyền truy cập.

Nhược điểm: Chủ yếu tập trung vào PDF, thiếu khả năng tích hợp ngữ cảnh lịch sử người dùng.

Ứng dụng: Thường được dùng trong ngành tài chính, kế toán, hành chính văn phòng.

3.3.3. Formstack Documents

Tính năng: Tự động tạo và điền biểu mẫu dựa trên dữ liệu từ các nền tảng như Google Sheets, Salesforce, Zapier.

Ưu điểm: Giao diện kéo-thả, dễ sử dụng, nhiều tính năng nâng cao như logic điều kiện, gợi ý nội dung.

Nhược điểm: Chi phí cao, phụ thuộc vào kết nối internet và đám mây.

Tính mở rộng: Cao, nhưng phụ thuộc vào gói dịch vụ đăng ký.

3.3.4. Foxit PDF Editor

Tính năng: Soạn thảo, chỉnh sửa biểu mẫu PDF, tạo trường điền thông minh, hỗ trợ nhiều định dạng ngôn ngữ.

Ưu điểm: Nhẹ, nhanh, chi phí thấp hơn so với Adobe, tương thích nhiều hệ điều hành.

Nhược điểm: Tính năng AI hạn chế, không hỗ trợ tự động gợi ý từ lịch sử điền biểu mẫu.

Khả năng áp dụng: Thích hợp cho tổ chức vừa và nhỏ.

## 3.4. So sánh các giải pháp

|  |  |  |
| --- | --- | --- |
| Tiêu chí | Mã nguồn mở (Docassemble, OpenFaaS, Thư viện Python) | Mã nguồn đóng / Thương mại (Adobe, Kofax, Formstack, Foxit) |
| Tính năng chính | Linh hoạt, có thể lập trình mở rộng | Đầy đủ, sẵn dùng, tối ưu cho người dùng cuối |
| Chi phí | Miễn phí, tiết kiệm ngân sách | Trả phí theo tháng/năm hoặc bản quyền vĩnh viễn |
| Mức độ tùy biến | Cao – tự phát triển theo nhu cầu | Thấp – giới hạn trong chức năng nhà cung cấp |
| Hỗ trợ kỹ thuật | Cộng đồng và tài liệu | Chuyên nghiệp, hỗ trợ trực tiếp qua email/chat |
| Bảo mật | Do người triển khai kiểm soát | Tích hợp các chuẩn bảo mật doanh nghiệp |
| Yêu cầu kỹ thuật | Cao – cần lập trình và triển khai | Thấp – người dùng phổ thông có thể sử dụng |
| Đối tượng sử dụng | Nhóm nghiên cứu, lập trình viên, tổ chức nhỏ | Doanh nghiệp lớn, tổ chức cần giải pháp ổn định và nhanh chóng |

# Ứng dụng hệ thống điền biểu mẫu tự động vào giáo dục

## 4.2. Ứng dụng trong giảng dạy

4.2.1. Hỗ trợ giáo viên

Hệ thống điền biểu mẫu tự động mang đến nhiều tiện ích cho giáo viên ở các cấp học:

Tạo và quản lý biểu mẫu kiểm tra, khảo sát: Thay vì phải chuẩn bị thủ công, giáo viên có thể dễ dàng tạo biểu mẫu kiểm tra trắc nghiệm, phiếu khảo sát học sinh hay bảng đánh giá tiết học chỉ bằng vài thao tác.

Tự động hóa chấm điểm: Thông qua tích hợp với các hệ thống LMS (Learning Management System) hoặc công cụ chấm điểm tự động, giáo viên có thể tiết kiệm thời gian, đồng thời đảm bảo tính khách quan trong chấm điểm.

Quản lý hồ sơ học sinh: Việc nhập liệu thông tin học sinh như lý lịch, điểm số, nhận xét quá trình học tập... được đơn giản hóa, giảm thiểu sai sót nhờ chức năng tự động điền và đồng bộ dữ liệu.

4.2.2. Hỗ trợ học sinh/sinh viên

Từ góc độ người học, hệ thống cũng đem lại những lợi ích thiết thực:

Đăng ký môn học, lớp học: Sinh viên chỉ cần điền một biểu mẫu duy nhất, hệ thống sẽ tự động kiểm tra điều kiện tiên quyết và sắp xếp lớp phù hợp.

Nộp bài tập, hồ sơ: Các biểu mẫu giúp quản lý tiến độ nộp bài, thông báo nhắc nhở hạn nộp và hỗ trợ phản hồi tự động từ giáo viên.

Khảo sát ý kiến: Sinh viên dễ dàng tham gia khảo sát đánh giá giảng viên, chương trình học, từ đó đóng góp ý kiến giúp nhà trường cải thiện chất lượng đào tạo.

Ví dụ minh họa:

Một trường đại học áp dụng hệ thống khảo sát đầu khóa, sinh viên chỉ cần mất vài phút để hoàn tất. Dữ liệu được phân tích và phản hồi đến giảng viên để điều chỉnh giáo trình kịp thời.

Giáo viên trung học sử dụng biểu mẫu tự động cho điểm danh và nhận xét hàng tuần, giúp phụ huynh theo dõi sát sao quá trình học tập của con em.

## 4.3. Ứng dụng trong quản lý giáo dục

Hệ thống cũng đóng vai trò quan trọng trong việc nâng cao hiệu quả quản lý hành chính tại các cơ sở giáo dục:

Quản lý hồ sơ cán bộ, giáo viên: Tự động điền và cập nhật các thông tin cá nhân, quá trình công tác, kết quả đánh giá năng lực... giúp tiết kiệm thời gian cho phòng nhân sự.

Báo cáo, thống kê: Với dữ liệu được thu thập và tổ chức tự động, hệ thống có thể xuất báo cáo định kỳ về chất lượng giảng dạy, tỷ lệ học sinh lên lớp, kết quả học tập... từ đó hỗ trợ nhà trường đưa ra quyết định nhanh chóng, chính xác.

Quản lý tuyển sinh: Ứng viên chỉ cần điền biểu mẫu đăng ký một lần duy nhất. Hệ thống sẽ tự động xét tuyển dựa trên tiêu chí đầu vào và gửi thông báo kết quả.

Ví dụ minh họa:

Phòng đào tạo sử dụng hệ thống để tổng hợp dữ liệu đăng ký học phần và lập kế hoạch giảng dạy, phân lớp chỉ trong vài phút.

Ban giám hiệu nhận được báo cáo hàng tuần qua email về tỷ lệ chuyên cần, thống kê kết quả học tập theo lớp/khoa.

## 4.4. Ứng dụng trong hướng nghiệp giáo dục

Trong thời đại số, công tác hướng nghiệp không thể thiếu các công cụ tự động giúp cá nhân hóa trải nghiệm cho học sinh:

Khảo sát năng lực, sở thích: Học sinh tham gia khảo sát qua biểu mẫu điện tử, hệ thống phân tích và đưa ra các gợi ý ngành nghề dựa trên sở thích, năng lực, điểm mạnh cá nhân.

Tư vấn hướng nghiệp: Kết hợp với chatbot hoặc trợ lý ảo, hệ thống có thể phản hồi các câu hỏi của học sinh về ngành học, yêu cầu đầu vào, cơ hội nghề nghiệp một cách nhanh chóng và chính xác.

Quản lý hoạt động ngoại khóa: Học sinh đăng ký, theo dõi tiến độ, và đánh giá hoạt động qua hệ thống biểu mẫu tích hợp, từ đó giúp nhà trường quản lý hiệu quả và đưa ra chương trình hướng nghiệp phù hợp hơn.

Ví dụ minh họa:

Trung tâm hướng nghiệp tổ chức khảo sát quy mô lớn cho học sinh cuối cấp, sau đó gửi bản phân tích chi tiết về nghề nghiệp phù hợp cho từng em qua email.

Học sinh đăng ký tham dự hội thảo nghề nghiệp thông qua biểu mẫu điện tử, hệ thống gửi mã QR xác nhận, lịch trình và thông báo tự động trước sự kiện.

## 4.5. Lợi ích khi áp dụng hệ thống điền biểu mẫu tự động trong giáo dục

Việc triển khai hệ thống này mang lại nhiều lợi ích vượt trội:

Tiết kiệm thời gian: Giảm bớt khối lượng công việc thủ công lặp đi lặp lại cho giáo viên, cán bộ quản lý.

Tăng độ chính xác: Hạn chế lỗi nhập liệu, sai sót khi tổng hợp thủ công.

Nâng cao trải nghiệm người học: Quy trình số hóa, đơn giản và minh bạch hơn.

Hỗ trợ ra quyết định: Cung cấp dữ liệu tổng hợp, phân tích trực quan giúp nhà trường đưa ra quyết sách chính xác hơn.

Cá nhân hóa tư vấn: Phân tích dữ liệu để đưa ra gợi ý hướng nghiệp, lộ trình học tập phù hợp với từng học sinh/sinh viên.

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

## 5.2. Xử lý ngôn ngữ tự nhiên (NLP)

5.2.1. Định nghĩa và vai trò

Xử lý ngôn ngữ tự nhiên là nhánh của trí tuệ nhân tạo tập trung vào việc cho phép máy tính hiểu, diễn giải và sinh ra ngôn ngữ tự nhiên của con người. Trong hệ thống điền biểu mẫu tự động, NLP hỗ trợ:

Phân tích cấu trúc ngôn ngữ học trong biểu mẫu (tiêu đề trường, nội dung gợi ý).

Trích xuất thực thể (Entity Recognition) như tên người, địa chỉ, mã số, ngày tháng,...

Tiền xử lý văn bản để loại bỏ nhiễu, chuẩn hóa ngữ liệu.

5.2.2. Các bước tiền xử lý văn bản cơ bản

Chuẩn hóa văn bản: chuyển văn bản về dạng chuẩn (chữ thường, bỏ dấu câu,...)

Tách từ (Tokenization): chia văn bản thành các đơn vị nhỏ như từ hoặc cụm từ.

Stemming/Lemmatization: đưa từ về dạng gốc để giảm độ biến thiên ngôn ngữ.

Loại bỏ stopword: như “là”, “và”, “của”, giúp tập trung vào từ khóa chính.

## 5.3. Thuật toán so khớp trường dữ liệu

5.3.1. Bài toán so khớp

Khi hệ thống nhận được một biểu mẫu cần điền, nó cần xác định trường dữ liệu phù hợp từ tập dữ liệu nguồn để tự động điền vào. Điều này yêu cầu một phương pháp đo lường mức độ tương đồng ngữ nghĩa giữa các đoạn văn bản – thường là giữa nhãn trường biểu mẫu và nhãn dữ liệu.

5.3.2. Các công thức đo độ tương đồng văn bản

a) Cosine Similarity

Dùng để đo góc giữa hai vector đặc trưng biểu diễn văn bản (sau khi chuyển sang không gian vector bằng kỹ thuật embedding hoặc bag-of-words):



Giá trị gần 1 cho thấy mức độ tương đồng cao, gần 0 là không tương đồng.

b) Jaccard Similarity

Phù hợp khi biểu diễn văn bản dưới dạng tập hợp từ:



Thích hợp cho các văn bản ngắn như tiêu đề hoặc nhãn biểu mẫu.

c) Levenshtein Distance (Khoảng cách chỉnh sửa)

Dùng để đo số bước thay đổi tối thiểu để chuyển văn bản A thành B. Rất hiệu quả khi có sai chính tả, lỗi đánh máy.

5.3.3. Ứng dụng trong so khớp dữ liệu biểu mẫu

So sánh nhãn trường biểu mẫu với nhãn hoặc giá trị dữ liệu nguồn.

Lựa chọn trường có độ tương đồng cao nhất để gán thông tin.

Cải tiến độ chính xác thông qua kết hợp nhiều công thức (ensemble similarity).

## 5.4. Mô hình AI và kỹ thuật embedding

5.4.1. Embedding: Biểu diễn ngữ nghĩa bằng vector

Embedding là kỹ thuật biến đổi từ/ngữ sang dạng vector có mang ý nghĩa ngữ cảnh. So với biểu diễn truyền thống như one-hot, embedding cung cấp sự hiểu biết sâu sắc hơn về ngữ nghĩa và mối quan hệ giữa các từ.

Các kỹ thuật embedding phổ biến:

Word2Vec: học biểu diễn từ qua ngữ cảnh cận kề (CBOW, Skip-gram).

GloVe: biểu diễn từ dựa trên thống kê đồng xuất hiện toàn cục.

FastText: xử lý được các từ hiếm nhờ học embedding theo n-gram ký tự.

Embedding được dùng để:

Biến nhãn trường và dữ liệu thành vector.

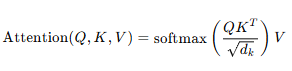
So sánh vector qua các hàm tương đồng như cosine.

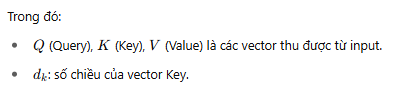
Cung cấp input cho các mô hình học sâu.

5.4.2. Transformer và Attention

Transformer là kiến trúc mạng nơ-ron đột phá trong NLP, sử dụng cơ chế Attention để mô hình hóa quan hệ ngữ nghĩa giữa các từ ở khoảng cách xa.

Attention cơ bản:





Ứng dụng trong hệ thống:

Mô hình như BERT hoặc GPT có thể được tích hợp để hiểu ngữ cảnh sâu rộng, dự đoán nhãn trường hoặc sinh văn bản điền phù hợp.

Có thể huấn luyện mô hình fine-tuned để giải bài toán điền mẫu trong lĩnh vực cụ thể như giáo dục, hành chính, y tế,...

# 6. Kết luận

## 6.1 Tổng kết kết quả đạt được

Trong quá trình nghiên cứu và triển khai, hệ thống đã đạt được nhiều kết quả đáng ghi nhận, thể hiện ở các khía cạnh sau:

Tự động nhận diện và so khớp chính xác các trường dữ liệu giữa biểu mẫu và dữ liệu nguồn, nhờ vào sự kết hợp giữa kỹ thuật trích xuất từ khóa, phân tích ngữ nghĩa và đo lường độ tương đồng ngữ cảnh.

Áp dụng hiệu quả các công thức toán học như Cosine Similarity và Jaccard Index trong việc đánh giá mức độ tương đồng giữa các chuỗi văn bản, từ đó hỗ trợ quá trình ánh xạ trường dữ liệu một cách chính xác.

Tích hợp thành công các phương pháp embedding (TF-IDF, Word2Vec, BERT) và mô hình transformer, giúp hệ thống không chỉ hiểu được nội dung bề mặt mà còn nắm bắt được mối liên hệ ngữ nghĩa sâu xa giữa các cụm từ và ngữ cảnh sử dụng.

Xây dựng giao diện người dùng trực quan, thân thiện, dễ sử dụng với khả năng hỗ trợ nhập liệu thông minh, rút ngắn thời gian thao tác và giảm thiểu sai sót từ phía người dùng.

## 6.2 Giá trị thực tiễn và khả năng ứng dụng

Hệ thống mang lại nhiều giá trị thiết thực, có khả năng triển khai rộng rãi trong các hoạt động hành chính, giáo dục, y tế, tài chính, và doanh nghiệp, cụ thể như:

Tiết kiệm thời gian cho người dùng khi điền các biểu mẫu lặp đi lặp lại.

Giảm thiểu rủi ro sai sót do nhập liệu thủ công, nhất là trong các quy trình phức tạp hoặc mang tính pháp lý cao.

Tăng khả năng tái sử dụng và mở rộng, cho phép tích hợp dễ dàng với các phần mềm quản lý hiện có thông qua API.

Thúc đẩy chuyển đổi số, góp phần cải tiến quy trình quản lý thông tin trong tổ chức.

## 6.3 Hạn chế và hướng phát triển

Mặc dù đạt được nhiều kết quả khả quan, hệ thống vẫn còn một số hạn chế cần được cải thiện:

Phụ thuộc vào chất lượng và tính nhất quán của dữ liệu đầu vào, đặc biệt là khi dữ liệu chứa nhiều lỗi chính tả hoặc thiếu ngữ cảnh.

Khả năng xử lý các trường hợp ngoại lệ (ví dụ: trường thông tin không phổ biến, thông tin ẩn) vẫn còn hạn chế.

Hiệu năng xử lý có thể bị ảnh hưởng khi mở rộng quy mô dữ liệu lớn hoặc biểu mẫu có cấu trúc phức tạp.

Để khắc phục và nâng cao hiệu quả hệ thống trong tương lai, một số hướng nghiên cứu và phát triển tiềm năng bao gồm:

Tăng cường khả năng học hỏi và thích nghi theo hành vi người dùng, nhằm cá nhân hóa kết quả điền biểu mẫu theo từng đối tượng cụ thể.

Ứng dụng mô hình học sâu thế hệ mới (như GPT, T5, LLaMA) để nâng cao khả năng hiểu ngữ cảnh và sinh văn bản chính xác, tự nhiên hơn.

Tối ưu hóa hiệu suất hệ thống và đảm bảo an toàn dữ liệu, áp dụng các biện pháp bảo mật tiên tiến như mã hóa đầu cuối và xác thực nhiều lớp.

Mở rộng khả năng tích hợp với nhiều hệ thống nguồn khác nhau, chẳng hạn như cơ sở dữ liệu lớn, dịch vụ điện toán đám mây, hoặc các hệ thống quản lý quy trình nghiệp vụ (BPM).

## 6.4 Kết luận chung

Hệ thống điền biểu mẫu tự động dựa trên lịch sử người dùng không chỉ là một công cụ hỗ trợ hiệu quả cho công việc hàng ngày, mà còn là một bước tiến công nghệ quan trọng trong quá trình chuyển đổi số và tự động hóa thông minh. Việc ứng dụng trí tuệ nhân tạo vào lĩnh vực nhập liệu giúp nâng cao năng suất, độ chính xác và trải nghiệm người dùng. Với định hướng phát triển bền vững và không ngừng hoàn thiện, hệ thống hứa hẹn sẽ tiếp tục mở rộng và đóng góp tích cực vào nhiều lĩnh vực trong tương lai.

# References

1. Liu, C.-C., Liao, M.-G., Chang, C.-H., Lin, H.-M. (2022). An analysis of children’ interaction with an AI chatbot and its impact on their interest in reading. Computers & Education, 189, 104576, [https://doi.org/10.1016/j.compedu.2022.104576.](https://doi.org/10.1016/j.compedu.2022.104576)
2. Marques, C., Bachega, S.J., Tavares, D.M. (2019). Framework proposal for the environmental impact assessment of universities in the context of Green IT, Journal of Cleaner Production, 241, Article 118346, https://doi.org/10.1016/j.jclepro.2019. 118346.
3. McGrath, C., Cerratto Pargman, T., Juth, N., Palmgren, P.J. (2023). University teachers’ perceptions of responsibility and artificial intelligence in higher education – An experimental philosophical study. Computers and Education: Artificial Intelligence, Vol. 4, 100139, [https://doi.org/10.1016/j.caeai.2023.100139.](https://doi.org/10.1016/j.caeai.2023.100139)
4. Miller, J. (2020). Climate change solutions: The role of technology, House of Commons Library, https://commonslibrary.parliament.uk/climate-change-solutions-the-role-of- technology/.
5. Mizumoto, A., Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring, Research Methods in Applied Linguistics, Vol. 2, 100050, [https://doi.org/10.1016/j.rmal.2023.100050.](https://doi.org/10.1016/j.rmal.2023.100050)
6. Mollick, E., Mollick, L. (2023). Using AI to Implement Effective Teaching Strategies in Classrooms: Five Strategies, Including Prompts, Wharton School of the University of Pennsylvania & Wharton Interactive, March 16, [https://ssrn.com/abstract=4391243](https://ssrn.com/abstract%3D4391243) .
7. Moore, G.E. (1998). Cramming More Components Onto Integrated Circuits,

*Proceedings of the IEEE*, 86(1), 82-85, [https://doi.org/10.1109/JPROC.1998.658762.](https://doi.org/10.1109/JPROC.1998.658762)

1. Perkins, K.M., Munguia, N., Moure-Eraso, R., Delakowitz, B., Giannetti, B.F., Liu, G., Velazquez, L. (2018). International perspectives on the pedagogy of climate change. Journal of Cleaner Production, 200, 1043-1052, https://doi.org/10.1016/j.jclepro. 2018.07.296.
2. Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond, Journal of University Teaching & Learning Practice, 20(2), [https://doi.org/10.53761/1.20.02.07.](https://doi.org/10.53761/1.20.02.07)
3. Ray, P.P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, Internet of Things and Cyber- Physical Systems, Vol. 3, 121-154, [https://doi.org/10.1016/j.iotcps.2023.04.003.](https://doi.org/10.1016/j.iotcps.2023.04.003)
4. Runge, I., Lazarides, R., Rubach, C., Richter, D. (2023). Teacher-reported instructional quality in the context of technology-enhanced teaching: The role of teachers’ digital competence-related beliefs in empowering learners. Computers & Education, No. 198, 104761, https://doi.org/10.1016/j.compedu.2023.104761.
5. Schroder, A., Prockl, G., Constantiou, I. (2021). How Digital Platforms with a Social Purpose Trigger Change towards Sustainable Supply Chains, Proceedings of the 54th Hawaii International Conference on System sciences, https://doi.org/10.24251/ HICSS.2021.580.
6. Sipică, I.S., Toma, E. (2022). The main challenges in education in the context of the digital transformation of learning. Scientific Papers Series, Management, Economic Engineering in Agriculture and Rural Development, 22(3), https://managementjournal

.usamv.ro/pdf/vol.22\_3/Art73.pdf.

1. Sloan, K. (2020). E-Learning: How the coronavirus has changed education forever. Cybint, retrieved from: https:/[/www.cybintsolutions.com/e-learning-how-the-corona](http://www.cybintsolutions.com/e-learning-how-the-corona) virus-has-changed-education-forever/.
2. Tahiru, F. (2021). AI in Education: A Systematic Literature Review, Journal of Cases on Information Technology (JCIT), 23(1), 1-20, <http://doi.org/10.4018/JCIT.2021> 010101.
3. Tam, G., El-Azar, D. (2020). 3 ways the coronavirus pandemic could reshape education. World Economic Forum, retrieved from: https:/[/www.weforum.org/agenda/2020/03/3-](http://www.weforum.org/agenda/2020/03/3-) ways-coronavirus-is-reshaping-education-and-what-changes-might-be-here-tostay/.
4. UNESCO (2020). Global Education Coalition. COVID-19 Education Response. Retrieved from: https://en.unesco.org/covid19/educationresponse/globalcoalition.
5. Volchik, V., Oganesyan, A., Olejarz, T. (2018). Higher education as a factor of socio- economic performance and development, Journal of International Studies, 11(4), 326-340, [https://doi.org/10.14254/2071-8330.2018/11-4/23.](https://doi.org/10.14254/2071-8330.2018/11-4/23)
6. Wand, X., Li, L., Tan, S.C., Yang, L., Lei, J. (2023). Preparing for AI-enhanced education: Conceptualizing and empirically examining teachers’ AI readiness, Computers in Human Behavior, No. 146, 107796, https://doi.org/10.1016/j.chb.2023

.107798.

1. Werfhorst, H.G., Kessenich, E., Geven, S. (2022). The digital divide in online education: Inequality in digital readiness of students and schools, Computers and Education Open, Vol. 3, 100100. [https://doi.org/10.1016/j.caeo.2022.100100.](https://doi.org/10.1016/j.caeo.2022.100100)