

Ứng dụng học sâu và nhận diện ký tự quang học trong số hóa tài liệu báo cáo tài chính

Nguyễn Quang Học • Đặng Thiên Vũ • Trần Thị Minh Hiền • Lê Hoàng Sử^(*)

Ngày nhận bài: 01/4/2024 | Biên tập xong: 02/8/2024 | Duyệt đăng: 10/8/2024

TÓM TẮT: Trong bối cảnh của cuộc cách mạng số hóa toàn cầu, việc sử dụng phương pháp nhập liệu thủ công để số hóa bảng biểu trong báo cáo tài chính (BCTC) đã trở nên lỗi thời và không đáp ứng được nhu cầu về thời gian và chi phí trong thời đại hiện nay. Để khắc phục hạn chế này, nghiên cứu tập trung đề xuất một phương pháp tự động nhận diện bảng biểu trong các BCTC từ hình ảnh thông qua công cụ PaddleOCR. Nghiên cứu tận dụng những mô hình học sâu và công nghệ nhận diện ký tự quang học (OCR) của công cụ mã nguồn mở này để tiến hành các công đoạn phát hiện bảng, phát hiện và nhận diện văn bản, dự đoán cấu trúc bảng và tọa độ ô và cuối cùng là tái tạo lại bảng tương đồng dưới dạng file Excel, HTML. Nghiên cứu thực nghiệm và so sánh với các bảng thực tế, đạt được độ chính xác trung bình trong việc tái tạo cấu trúc bảng và nhận diện các cột nội dung quan trọng là 95% đối với dạng bảng có các đường viền đầy đủ và 83% đối với dạng bảng ít viền. Kết quả khả quan này khẳng định tính ứng dụng của công cụ cho giải pháp số hóa tài liệu chứa bảng biểu, giảm bớt thời lượng xử lý các công việc nhập liệu.

TỪ KHÓA: Trí tuệ nhân tạo, học sâu, nhận diện ký tự quang học, báo cáo tài chính, số hóa tài liệu.

Mã phân loại JEL: C45, G39, O32.

DOI: <https://doi.org/10.63065/ajeb.vn.2024.221.101386>.

1. Giới thiệu

Trong bối cảnh số hóa toàn cầu hiện nay, các ngành công nghiệp truyền thống đang dần phải bắt kịp xu hướng và sự phát triển của thời đại. Ngay tại Việt Nam, làn sóng chuyển đổi số cũng đang lan rộng tới tất cả các cấp quản lý, từ Chính phủ đến các doanh nghiệp và địa phương. Cụ thể, năm 2020, Chính phủ đã chính thức phê duyệt “Chương trình

Chuyển đổi số quốc gia đến năm 2025, với tầm nhìn đến năm 2030”, tạo cơ sở cho các địa phương và doanh nghiệp tham gia vào

^(*) Lê Hoàng Sử - Trường Đại học Kinh tế - Luật, Đại học Quốc gia TP.HCM, Số 669 Quốc lộ 1, Khu phố 6, Phường Linh Xuân, TP. Thủ Đức, TP. Hồ Chí Minh;
Email: sulh@uel.edu.vn.