

NGHIÊN CỨU KỸ THUẬT WATERMARKING VÀ ỨNG DỤNG TRONG VĂN BẢN TIẾNG VIỆT

A CASE STUDY ON USING WATERMARKING TECHNIQUE AND ITS APPLICATION TO VIETNAMESE DOCUMENTS

Tác giả: Nguyễn Thị Hồng Thúy, Nguyễn Thị Hoa, [Võ Trung Hùng*](#)

Tóm tắt bằng tiếng Việt:

Trong bài báo này, chúng tôi trình bày kết quả nghiên cứu về kỹ thuật watermarking và cách áp dụng vào tài liệu tiếng Việt. Nghiên cứu tập trung phát triển kỹ thuật dịch chuyển kí tự để nhúng thông tin vào văn bản tiếng Việt. Dựa trên đặc điểm của chữ viết tiếng Việt, chúng tôi đề xuất một phương pháp để tăng số lượng thông tin được nhúng vào tài liệu. Khối lượng thông tin nhúng vào tài liệu tiếng Việt tăng lên nhiều so với kỹ thuật dịch chuyển kí tự chỉ dựa vào chữ viết La-tinh. Việc dấu thông tin vào tài liệu được chia thành 2 giai đoạn chính là nhúng và trích xuất. Phương pháp chúng tôi sử dụng là kỹ thuật di chuyển lên hoặc xuống các dấu của tiếng Việt với khoảng cách rất nhỏ. Chúng tôi đã phát triển và thử nghiệm thành công phần mềm nhúng và trích xuất thông tin trên các loại tài liệu tiếng Việt với nhiều định dạng khác nhau.

Từ khóa: văn bản tiếng Việt; xử lý ngôn ngữ tự nhiên; bảo mật; mã hóa; watermarking

Tóm tắt bằng tiếng Anh:

In this paper, we present the results of a research on watermarking technique and its application to the Vietnamese documents. This research focuses on a character displacement technique to embed information in the Vietnamese documents. Based on the Vietnamese characteristics, we propose a method to increase the amount of information to be embedded in documents. The amount of information embedded in the Vietnamese documents increase more than the character displacement technique based on Latin letters. The insertion of information in a document is divided into two main phases: embedding and extraction. Our method is based on moving up or down the signs of Vietnamese with very small distances. We have developed a tool to embed and extract information on a Vietnamese document, with many different formats.

Key words: Vietnamese document; NLP; protection; coding; watermarking