

MỘT MÔ HÌNH SEARCH ENGINE CHO TIẾNG VIỆT

Hoàng Kiếm, Huỳnh Thụy Bảo Trân

Khoa CNTT, Trường Đại học Khoa học Tự Nhiên - ĐHQG tp.HCM

Tóm tắt:

Báo cáo trình bày một số kết quả tìm hiểu về nguyên lý vận hành của hệ thống search engine. Trước hết, chúng tôi phân tích đặc điểm loại hình ngôn ngữ Việt và so sánh những thuận lợi, cũng như khó khăn trong việc tìm kiếm dựa trên ngôn ngữ Ấn - Âu và tiếng Việt. Sau đó, dựa trên những phân tích trên, chúng tôi đề xuất một mô hình search engine có hỗ trợ tìm kiếm thông tin tiếng Việt, dựa trên việc xây dựng vector chỉ mục nghịch đảo của tài liệu.

A PARTICULAR SEARCH ENGINE MODEL FOR VIETNAMESE LANGUAGE

Hoang Kiem, Huynh Thuy Bao Tran

Department of Information Technology, University of Natural Sciences -
VNU.HCM

Abstract:

This article represents some results of our work in understanding the operational principles of a search engine system. We begin with the analysis of Vietnamese characteristics and comparing the difficult issues in searching by Indian-European and Vietnamese language. After that, based on these considerations, we propose a new model for Vietnamese-supported search engine using the concept of inverted index vectors be obtained from documents.