

NGHIÊN CỨU VÀ XÂY DỰNG MÔI TRƯỜNG QUẢN LÝ, TRUY CẬP
TÀI NGUYÊN PHỤC VỤ XỬ LÝ TIẾNG VIỆT VÀ TIẾNG DÂN TỘC THIỂU SỐ

*RESEARCHING AND BUILDING AN ENVIRONMENT FOR ACCESSING AND MANAGING
RESOURCES FOR VIETNAMESE AND ETHNIC MINORITY LANGUAGE PROCESSING*

Tác giả: *Huỳnh Công Pháp, Văn Đỗ Cẩm Vân*

*Trường Cao đẳng Công nghệ Thông tin, Đại học Đà Nẵng;
hcpchap@gmail.com; van156dvn@gmail.com*

Tóm tắt:

Xử lý tiếng Việt (TV) và tiếng dân tộc thiểu số (TDTTS) là một lĩnh vực nghiên cứu rất được quan tâm và cấp bách đối với Việt Nam. Vấn đề đặt ra là làm thế nào để tạo điều kiện thuận lợi cho cộng đồng nghiên cứu về lĩnh vực này nhằm nâng cao và phát huy hơn nữa hiệu quả quá trình nghiên cứu. Bởi lẽ, các nguồn tài nguyên phục vụ xử lý TV và TDTTS được lưu trữ phân tán, nên việc khai thác các nguồn tài nguyên này gặp nhiều khó khăn. Do đó, ý tưởng và giải pháp của bài báo trước hết là tập hợp và xây dựng hệ thống truy cập tập trung các tài nguyên phục vụ xử lý TV và TDTTS. Sau đó, hợp nhất chúng, tạo nên nguồn dữ liệu lớn hơn. Cuối cùng, xây dựng một hệ thống hoạt động theo kiến trúc hướng dịch vụ (SOA) sao cho việc truy xuất khai thác các tài nguyên và công cụ xử lý TV và TDTTS dễ dàng và hiệu quả hơn.

Từ khóa: Xử lý tiếng Việt; Xử lý tiếng dân tộc thiểu số; Tập hợp; Môi trường quản lý; Hợp nhất; Khai thác.

Abstract:

Nowadays, processing Vietnamese and ethnic minority languages is an urgent and interesting issue for many organizations and individuals. However, the resources which are available to the research of processing Vietnamese language as well as some languages of minorities were stored in a discrete way. This is the reason why the exploitation and the inheritance of these resources meet a lot of difficulties. Therefore, the paper proposes some measures to improve the situation. Firstly, collecting and constructing a centralized system of scientific documents on this field. Secondly, merging them in order to create the bigger resources. Finally, constructing the service – oriented architecture system which allows users to access and exploit the tools of processing Vietnamese and minority languages effectively.

Key words: Process Vietnamese language; Process minority languages; Collect; Management environment; Merge; Exploit.