

# Một số hướng đề tài cho sinh viên đại học và học viên cao học

Năm học 2011-2012

Lê Thanh Hương – Viện Công nghệ Thông tin và Truyền thông – ĐHBKHN

Email: [huonglt@soict.hut.edu.vn](mailto:huonglt@soict.hut.edu.vn)

1. **Trích rút thông tin:** thu thập và phân tích các tài nguyên trên Internet, đưa vào một CSDL, cho phép người dùng có thể dễ dàng khai thác thông tin.

Các vấn đề nghiên cứu:

- Nghiên cứu phương pháp trích rút thực thể từ văn bản tiếng Việt
- Nghiên cứu phương pháp trích rút quan hệ giữa các thực thể từ văn bản tiếng Việt
- Nghiên cứu phương pháp xác định đồng tham chiếu thực thể từ các văn bản tiếng Việt
- Xây dựng ứng dụng dựa trên mạng ngữ nghĩa cho phép khai thác/truy vấn thông tin từ các nguồn dữ liệu trích rút

## 2. Tóm tắt văn bản

- a. Tóm tắt văn bản hướng truy vấn: Khi máy tìm kiếm trả về cho người dùng một danh sách các tài liệu liên quan, họ phải duyệt qua nội dung của toàn bộ các văn bản để tìm được thông tin mong muốn. Kỹ thuật tóm tắt hướng truy vấn cố gắng giảm bớt gánh nặng của người sử dụng bằng cách đưa ra nội dung cô đọng hơn liên quan trực tiếp đến câu truy vấn.
- b. Tóm tắt theo trích xuất: trích rút các câu quan trọng từ văn bản đầu vào và đưa vào văn bản tóm tắt.

Vấn đề chính: Nghiên cứu phương pháp chọn các câu quan trọng nhất để đưa vào văn bản tóm tắt

- c. Tóm tắt theo tóm lược: trên cơ sở các câu quan trọng từ văn bản đầu vào, tiến hành rút gọn câu và sinh văn bản tóm tắt dựa trên các kết quả rút gọn đó.

Vấn đề chính:

- Nghiên cứu phương pháp rút gọn câu nhằm loại bỏ các thông tin không quan trọng.
- Nghiên cứu phương pháp sinh văn bản tóm tắt

## 3. Hệ gợi ý.

Xây dựng hệ gợi ý trong một miền ứng dụng nhất định (như hỗ trợ tìm kiếm địa điểm du lịch, tìm kiếm tài liệu trên mạng, tự động cung cấp các quảng cáo phù hợp với người dùng, ...). Trên cơ sở truy vấn của người sử dụng và các thông tin liên quan (như user's profile, các tài liệu người dùng đã đọc, ...), hệ thống sẽ đưa ra các gợi ý phù hợp.