

XÂY DỰNG MỘT SỐ CÔNG CỤ HỖ TRỢ TRA CỨU, TỔNG HỢP THÔNG TIN ĐA PHƯƠNG TIỆN TRONG CÁC THƯ VIỆN SỐ

Chủ nhiệm đề tài nghiên cứu: **ĐỖ PHÚC, MS.**, TT Phát triển CNTT, ĐHQG TP. HCM & **NGUYỄN MINH HIỆP, MS.**, Thư viện Cao học, ĐH Khoa học Tự Nhiên

● CƠ QUAN CHỦ TRÌ ĐỀ TÀI

TT Phát triển Công nghệ Thông tin, Đại học Quốc gia TP. HCM
34 Trương Định Quận 3, TP. HCM, ĐT: 8222678

● CƠ QUAN VÀ CÁN BỘ PHỐI HỢP CHÍNH

1. Cơ quan phối hợp chính
 - Trung tâm Phát triển Công nghệ Thông tin, ĐHQG-HCM
 - Khoa Công nghệ Thông tin, ĐH Khoa học Tự Nhiên, ĐHQG-HCM
 - [Thư Viện Cao học](#), ĐH Khoa học Tự Nhiên, ĐHQG-HCM
2. Cán bộ phối hợp chính

TT	Họ và tên	Học vị	Cơ quan công tác
1.	Nguyễn Minh Hiệp	Th. Sĩ	Thư viện Cao học, ĐH Khoa học tự nhiên
2.	Đỗ Phúc	Th. Sĩ	Trung tâm Phát triển CNTT, ĐHQG-HCM
3.	Võ Đức Khánh	Th. Sĩ	Khoa Công nghệ Thông tin, ĐH Khoa học tự nhiên
4.	Vũ Hải Quân	Th. Sĩ	Khoa Công nghệ Thông tin, ĐH Khoa học tự nhiên
5.	Phạm Khắc Hậu	CN	Trung tâm Phát triển CNTT, ĐHQG-HCM
6.	Vũ Thanh Hiền	Th. Sĩ	Trung tâm Phát triển CNTT, ĐHQG-HCM
7.	Nguyễn Hà Giang	CN	Trung tâm Phát triển CNTT, ĐHQG-HCM
8.	Mai Tiến Dũng	CN	Trung tâm Phát triển CNTT, ĐHQG-HCM
9.	Lê Ngọc Oánh	Th. Sĩ	Thư Viện ĐH Mở-Bán công TP.HCM
10.	Dương Thúy Hương	CN	Thư Viện Cao học, ĐH Khoa học tự nhiên

□ TÊN CƠ QUAN ỨNG DỤNG KẾT QUẢ NGHIÊN CỨU

- [Thư Viện Cao học](#), ĐH Khoa học Tự Nhiên, ĐHQG-HCM
- Thư Viện Khoa học tổng hợp TP. HCM
- Trung tâm Lưu trữ Quốc gia tại TP. HCM
- Trung tâm Thông tin Khoa học & Công nghệ TP. HCM

● MỤC TIÊU CỦA ĐỀ TÀI, LÝ DO THỰC HIỆN ĐỀ TÀI

Việc nghiên cứu ứng dụng Công nghệ Thông tin vào công tác thư viện đã và đang được quan tâm. Sự phát triển nhanh chóng của Công nghệ Thông tin và Internet vào những thập niên cuối của thế kỷ 20 đã tạo ra số lượng thông tin khổng lồ. Các thư viện số với các nguồn thông tin có phong phú, đa dạng (văn bản, hình ảnh, siêu văn bản, đa phương tiện ...) đã hình thành và lưu trữ một lượng thông tin khổng lồ, đa dạng, phức tạp nên nhu cầu xây dựng các công cụ hỗ trợ việc tìm, đọc và tổng hợp nhanh chóng các nguồn thông tin theo các yêu cầu của nhà nghiên cứu là một nhu cầu bức thiết. Hiện nay trên thế giới đã xuất hiện một số công cụ hỗ trợ

từng phần các công việc nêu trên như: Phần mềm ELIB (Electronic Library on Web) hỗ trợ chức năng tìm kiếm thông tin như: tìm lướt, tìm theo toàn văn, tìm với các toán tử logic. Phần mềm Text Analyst hỗ trợ chức năng đọc và tổng hợp văn bản. Một số phần mềm hỗ trợ phân tích và diễn dịch ảnh. Các phần mềm này đã đóng góp vào công việc tra cứu và tổng hợp thông tin.

Tại Việt nam, trong những năm qua ngành thư viện và văn thư lưu trữ đã có những bước phát triển đáng kể. Việc đưa vào sử dụng các công cụ tin học trong lưu trữ tìm kiếm thông tin dần dần đã được chú ý. Tuy vậy, các công trình nghiên cứu đầu tư vào lĩnh vực nâng cao hiệu quả tìm, đọc và tổng hợp thông tin đặc biệt là thông tin đa phương tiện vẫn chưa được quan tâm đúng mức. Đề tài được xây dựng nhằm mục tiêu triển khai ứng dụng ban đầu các công nghệ tiên tiến của Công nghệ Tri thức vào lĩnh vực thư viện, góp phần thúc đẩy việc nghiên cứu, ứng dụng CNTT vào công tác thư viện và hỗ trợ các hoạt động nghiên cứu khác.

● NỘI DUNG NGHIÊN CỨU

Nội dung nghiên cứu của đề tài là cung cấp phương tiện dễ dàng truy cập thông tin hữu dụng từ thư viện số hóa thông qua Internet và dựa trên kỹ thuật mới trong việc tìm kiếm dựa trên nội dung, ước lượng và tổng hợp các dữ liệu đa phương tiện (văn bản, hình ảnh). Một số nội dung sẽ được đầu tư nghiên cứu như:

- Chất lượng của thông tin truy cập
- Tốc độ của thông tin được truy cập xử lý, đặc biệt là ở những nơi có kết nối với Internet yếu
- Khả năng làm việc với nhiều ngôn ngữ, đặc biệt là tiếng Việt, Anh, Nga, Pháp, và Hoa
- Các kỹ thuật sẽ nghiên cứu

1. Phương tiện tìm kiếm

- Tìm kiếm cả văn bản và dữ liệu visual với tốc độ cao thông qua Internet
- Tìm kiếm chéo (cross searching) thông qua những chia cắt
- Cải tiến giao diện của cơ chế tìm kiếm.

2. Ước lượng thông tin

- Phép đo Heuristic cho việc ước lượng các thông tin thích hợp, bao gồm các phép đo cho việc chia cắt các kiểu dữ liệu (văn bản, âm thanh, hình ảnh) và sự kết hợp các loại (âm thanh hình ảnh, hoạt hình, vv.)
- Nhiều ngôn ngữ
- Chất lượng thông tin truy cập.

3. Tổng hợp thông tin

- Làm thế nào chúng ta có thể kết hợp nhiều mục tin của tìm kiếm và ước lượng dữ liệu trong một tập chọn lọc một cách thông minh?

4. Các công nghệ dự định nghiên cứu và triển khai

- Liên quan đến Internet: Web, kết nối .
- Cơ sở dữ liệu cho thông tin đa phương tiện
- Kỹ thuật trí tuệ nhân tạo (AI) cho ước lượng dữ liệu
- Đồ họa máy tính, xử lý ảnh và kỹ thuật nhận dạng các mẫu cho các dữ liệu visual
- Kỹ thuật nhận dạng ngôn ngữ
- Giám sát trong thư viện kỹ thuật số và hệ thống thông tin.

● PHƯƠNG PHÁP NGHIÊN CỨU VÀ THỰC HIỆN

- Thu nhận thông tin: Thu nhận thông tin liên quan đến đề tài nghiên cứu như các công trình, công nghệ, bài báo nghiên cứu, các phần mềm có liên quan đến đề tài
- Phân tích, tổng hợp: Sử dụng các phương pháp và mô hình trí tuệ nhân tạo, mạng neural, thuật toán di truyền, nhận dạng và xử lý ảnh, các phương pháp toán học, thống kê, khoa học thông tin
- Thiết kế và cài đặt phần mềm máy tính: Phân tích, cài đặt hệ thống phần mềm hỗ trợ tra cứu và tổng hợp thông tin trong các thư viện số.

● SẢN PHẨM CỦA ĐỀ TÀI

- Báo cáo tổng hợp các kết quả nghiên cứu
- Quy trình khai thác, tra cứu, tổng hợp thông tin đa phương tiện
- Phần mềm Tin học đáp ứng yêu cầu đề tài
- Các bài báo và công trình khoa học.

● PHƯƠNG THỨC PHỔ BIẾN KẾT QUẢ NGHIÊN CỨU, CHUYỂN GIAO SỬ DỤNG TRÊN QUI MÔ RỘNG

- Ứng dụng vào [Thư Viện Cao học](#), Đại học Khoa học Tự nhiên, ĐHQG-HCM
- Ứng dụng vào Thư viện Quốc gia Việt nam
- Ứng dụng tại các Trung tâm lưu trữ quốc gia
- Các bài báo và công trình nghiên cứu khoa học

● THỜI GIAN THỰC HIỆN ĐỀ TÀI LÀ 18 THÁNG
