

# Tìm kiếm ảnh dựa vào Ontology

Nguyễn Thị Uyên Nhi<sup>1,3</sup>, Văn Thế Thành<sup>2</sup>, Lê Mạnh Thịnh<sup>3</sup>

<sup>1</sup> Khoa Thống kê – Tin học, Trường ĐH Kinh tế – ĐH Đà Nẵng

<sup>2</sup> Phòng quản lý khoa học và đào tạo sau Đại học, Trường ĐH Công nghiệp Thực phẩm TP.HCM

<sup>3</sup> Trường Đại học Khoa học, Đại học Huế

Tác giả liên hệ: Nguyễn Thị Uyên Nhi, uyennhi203@gmail.com

Ngày nhận bài: 15/04/2021, ngày sửa chữa: 01/06/2021, ngày duyệt đăng: 12/06/2021

Định danh DOI: 10.32913/mic-ict-research-vn.v2021.n1.965

**Tóm tắt:** Bài toán tìm kiếm ảnh đóng vai trò quan trọng trong mọi lĩnh vực của cuộc sống. Trong bài báo này, một phương pháp tìm kiếm ảnh dựa vào Ontology OBIR (Ontology-based Image Retrieval) được đề xuất nhằm xác định ngữ nghĩa cấp cao của hình ảnh. Ontology bán tự động được đề xuất xây dựng làm cơ sở tri thức cho tìm kiếm ảnh theo tiếp cận ngữ nghĩa, đồng thời câu truy vấn SPARQL được tự động tạo ra từ văn bản đầu vào hoặc từ lớp ngữ nghĩa của ảnh đầu vào, được xác định thông qua công cụ tìm kiếm học máy dựa trên đồ thị cụm láng giềng, để tìm kiếm trên Ontology này. Để minh chứng cho các lý thuyết đã đề xuất, chúng tôi tiến hành thực nghiệm trên bộ ảnh ImageCLEF (20.000 hình ảnh) và Stanford Dogs (20.580 hình ảnh). Kết quả thực nghiệm của phương pháp tìm kiếm ảnh dựa vào Ontology được so sánh với phương pháp tìm kiếm ảnh theo nội dung, đồng thời được so sánh với các công trình khác cùng tập dữ liệu ảnh nhằm chứng minh tính hiệu quả và đúng đắn của các đề xuất trong bài báo.

**Từ khóa:** OBIR, Ontology, SPARQL, semantic image.

---

**Title:** Ontology-based Image Retrieval

**Abstract:** Image retrieval problem plays an important role in all areas of life. In this paper, Ontology-based Image Retrieval (OBIR) is proposed to determine the high-level semantics of the image. The semi-automatic frame ontology is proposed to be built as a knowledge base for SBIR, at the same time the SPARQL query is automatically generated from the input text or the semantic class of the input image, that identified through the machine learning search engine based on the neighborhood cluster graph, to retrieval on this ontology. Experiments were performed on ImageCLEF and Stanford Dogs to demonstrate the effectiveness of the proposals. Experimental results of OBIR are compared with CBIR on the neighborhood cluster graph, and compared with other works on the same image datasets, to prove the effectiveness and correctness of the proposals in the paper.

**Keywords:** OBIR, Ontology, SPARQL, semantic image.

---

## I. GIỚI THIỆU

Ảnh số đóng một vai trò quan trọng trong mọi lĩnh vực của cuộc sống như y khoa, giáo dục, giải trí,... Tìm kiếm ảnh dựa trên nội dung CBIR (Content-based Image Retrieval) [4, 10] là phương pháp phổ biến nhất hiện tại, sử dụng các đặc trưng cấp thấp như màu sắc, kết cấu, hình dạng... và các độ đo khoảng cách để tìm tập ảnh tương tự. Tuy nhiên, luôn tồn tại một “khoảng cách ngữ nghĩa” (semantic gap) [2, 20] giữa các đặc trưng cấp thấp được máy tính trích xuất và ngữ nghĩa cấp cao được truyền đạt của người dùng. Tìm kiếm ảnh theo ngữ nghĩa SBIR (Semantic-based Image Retrieval) [15, 17] là phương pháp xác định ngữ nghĩa cấp cao của hình ảnh, nhằm giải quyết thách thức này.

Có nhiều phương pháp tìm kiếm ảnh theo ngữ nghĩa,

phổ biến nhất là dựa vào phương pháp học máy để liên kết đặc trưng cấp thấp với các nhãn lớp ngữ nghĩa, dựa vào Ontology để xác định ngữ nghĩa cấp cao. Trong phương pháp học máy, các phương pháp học sâu được sử dụng rộng rãi do tốc độ nhanh và độ chính xác cao của nó nhằm nhận dạng hình ảnh, phát hiện đối tượng, phân đoạn ngữ nghĩa như: phương pháp sử dụng mạng CNN phân loại khái niệm ngữ nghĩa cho hình ảnh nhằm chú thích các hình ảnh Web không được gắn nhãn [25], phân đoạn ngữ nghĩa với DeepLab v3 và thuật toán phân đoạn siêu pixel-dịch chuyển nhanh [29], phương pháp điều chỉnh không gian tích hợp vào mạng CNN nhằm giảm nhiễu và phân đoạn hình ảnh hiệu quả hơn [9]... Tuy nhiên, các phương pháp học sâu phát hiện đối tượng, phân đoạn ngữ nghĩa với các mô tả đặc trưng hình ảnh ở mức độ thấp, không thể chỉ ra rõ ràng mối quan hệ giữa các đối tượng hay thiếu nhận thức

ngữ cảnh vì nó thiếu các mô tả ngữ nghĩa nâng cao, do đó việc hiểu thông tin ngữ nghĩa cấp cao theo của hình ảnh theo yêu cầu người dùng trở nên khó khăn [7]. Ontology là một mô hình dữ liệu có thể biểu diễn một cách rõ ràng các khái niệm và các mối quan hệ khác nhau giữa các khái niệm. Do các tính năng định hướng máy tính và dựa trên logic, kỹ thuật Ontology đã được áp dụng rộng rãi cho mô hình hóa và phân tích thông tin như tích hợp ngữ nghĩa của cơ sở dữ liệu không đồng nhất, truy xuất và quản lý thông tin hình ảnh [8, 21]. Vì vậy, nhiều nghiên cứu nhằm tìm kiếm ngữ nghĩa ảnh trên Ontology được quan tâm như: khung khái niệm dựa trên Ontology để truy xuất hình ảnh của các hiện vật bảo tàng [21], nhận diện rủi ro trong xây dựng dựa vào ngữ nghĩa hình ảnh trên Ontology [28], sử dụng Ontology của hệ thống tìm mạch của con người để cải thiện việc phân loại các hình ảnh mô học [16]...

Trong cách tiếp cận của chúng tôi, một khung Ontology được xây dựng bán tự động cho bài toán tìm kiếm ảnh theo tiếp cận ngữ nghĩa. Tìm kiếm ảnh dựa trên Ontology OBIR (Ontology-based Image Retrieval) được chúng tôi thực hiện theo hai cách: (1) từ văn bản đầu vào của người dùng, xác định từ khóa để tìm kiếm trên Ontology và (2) từ hình ảnh đầu vào, thông qua công cụ tìm kiếm dựa trên phương pháp học máy (đồ thị cụm láng giềng) để trích xuất đặc trưng từ hình ảnh, phân lớp hình ảnh, tra cứu từ vựng thị giác (nhằm giải quyết bài toán chuyển từ một ảnh sang vec-tơ gồm các từ vựng thị giác). Sau đó tạo tự động câu lệnh SPARQL để truy vấn trên Ontology. Kết quả tìm kiếm ảnh trên Ontology là một tập ảnh tương tự cùng với ngữ nghĩa của các hình ảnh. Hệ truy vấn hình ảnh sau khi được tích hợp với Ontology sẽ cải thiện kết quả tìm kiếm cũng như độ tin cậy về mặt ngữ nghĩa cho hệ thống. Các đóng góp chính của bài báo bao gồm: (1) Đề xuất xây dựng một khung Ontology (Frame Ontology) bán tự động cho bộ dữ liệu ảnh dựa trên ngôn ngữ OWL; (2) Đề xuất mô hình tìm kiếm ảnh dựa vào Ontology; (3) Xây dựng thực nghiệm và chứng minh tính đúng đắn của phương pháp đề xuất dựa trên bộ dữ liệu ImageCLEF [12] và Stanford Dogs [14].

Phần còn lại của bài báo gồm: Phần II khảo sát một số công trình liên quan đồng thời phân tích các ưu nhược điểm để chứng minh tính hiệu quả của tìm kiếm ảnh dựa trên Ontology; Phần III trình bày phương pháp xây dựng Ontology bán tự động cho dữ liệu ảnh; Phần IV mô tả hệ tìm kiếm ảnh dựa trên Ontology OnSBIR và Phần V là thực nghiệm, đánh giá so sánh; Kết luận và hướng phát triển tương lai được thể hiện trong Phần VI.

## II. CÁC CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN

Nhiều công trình nghiên cứu về Ontology cho mục đích phân tích ngữ nghĩa, chú thích và tìm kiếm ngữ nghĩa ảnh. Phương pháp tìm kiếm ảnh dựa trên Ontology đã được đề

xuất theo hai hướng chính bao gồm: tìm kiếm ảnh trên Ontology dựa trên văn bản đầu vào [5, 17, 23] và tìm kiếm ảnh dựa trên ảnh đầu vào [8, 15] nhằm liên kết các đặc trưng cấp thấp và ngữ nghĩa cấp cao của ảnh trên Ontology.

Các phương pháp tìm kiếm ảnh theo tiếp cận ngữ nghĩa dựa trên Ontology với văn bản đầu vào được đề xuất trong các nghiên cứu: Vijayarajan, V. và cộng sự (2016) [23] xây dựng một Ontology trên tập ảnh ImageCLEF gồm 20.000 hình ảnh với các chú thích đi kèm mỗi ảnh, cho bài toán SBIR. Quá trình tìm kiếm hình ảnh phụ thuộc vào việc phân tích văn phạm của chú thích để tạo thành các từ khóa mô tả nội dung hình ảnh, tuy nhiên chưa thực hiện phân lớp nội dung hình ảnh từ các đặc trưng cấp thấp để tạo các từ khóa nhằm thực hiện tìm kiếm trên Ontology. Ontology được xây dựng với công cụ Protégé, đầu ra là RDF và OWL. Hình ảnh được truy xuất trên Ontology dựa vào truy vấn SPARQL tạo tự động. Tuy nhiên, khi số lượng ảnh lớn, Ontology xây dựng thủ công trên Protégé tốn nhiều thời gian và nhân lực, đồng thời Ontology trong đề xuất chỉ dừng lại mức độ miền phân lớp và phân cấp miền, các định nghĩa, quan hệ chưa được thực hiện. Mohd K. và cộng sự (2017) [17] đã đề xuất mô hình tìm kiếm ảnh theo tiếp cận ngữ nghĩa với Ontology mở đa phương thức và Dbpedia, nhằm nâng cao hiệu quả. Ontology mở được xây dựng bằng cách sử dụng các khái niệm mô tả các đối tượng trong ảnh. Các khái niệm, phạm trù và hình ảnh được liên kết với nhau bằng các giá trị mở trong Ontology. Tuy nhiên, phương pháp này chỉ tìm kiếm trên từ khóa, không tìm kiếm từ ảnh, do đó chưa liên kết được với đặc trưng cấp thấp. Đồng thời, Ontology chỉ xây dựng trên tập ảnh nhỏ với năm phân lớp ngữ nghĩa, chưa đại diện cho các tập ảnh đa dạng trong thực tế. Nhóm nghiên cứu Bouchakwa, M. (2020) [5] đề xuất cải tiến ngữ nghĩa dựa trên thẻ (Tag) ảnh bằng cách: đầu tiên, các ý nghĩa ngữ nghĩa khác nhau được suy ra từ các truy vấn không rõ ràng của người dùng dựa trên Ontology, sau đó truy vấn ban đầu được định dạng lại bằng cách thực thi một tập hợp các quy tắc ngữ nghĩa trên Ontology. Các hình ảnh trên một tập ảnh được gom cụm theo nội dung ngữ nghĩa và lọc, xếp hạng lại các kết quả truy xuất cụm hình ảnh để sắp xếp theo mức độ thích hợp. Để đánh giá hiệu suất của phương pháp đề xuất đề xuất, thực nghiệm được tiến hành trên bộ sưu tập 25.000 hình ảnh được chia sẻ trên Flickr. Các phương pháp tìm kiếm ảnh theo tiếp cận ngữ nghĩa dựa trên Ontology kết hợp với đặc trưng thị giác được đề xuất trong các nghiên cứu: Manzoor, U. và cộng sự (2015) [15] đề xuất phương pháp tìm kiếm ảnh theo tiếp cận ngữ nghĩa dựa trên Ontology miền cụ thể, để thu thập hình ảnh có liên quan đến tìm kiếm của người dùng. Hệ thống thực hiện tìm kiếm theo hai cách: từ hình ảnh đầu vào là sự kết hợp giữa đặc trưng cấp thấp và Ontology, hoặc từ văn bản do người dùng nhập vào, thực hiện tìm kiếm trực tiếp trên

Ontology. Miền động vật có vú gồm 900 ảnh với 20 loại động vật khác nhau được sử dụng để xây dựng Ontology và thực nghiệm chứng minh tính hiệu quả của đề xuất với độ chính xác là 0,60. Tuy nhiên, Ontology chỉ xây dựng được trên miền giá trị, chưa tạo mối quan hệ phân cấp giữa các lớp cũng như khái niệm ngữ nghĩa cho miền. Tập ảnh xây dựng Ontology là nhỏ, tính kế thừa và khả năng mở rộng chưa được thể hiện trong bài báo. Nhóm nghiên cứu Filali J. (2016) [8] trình bày một hệ thống tìm kiếm hình ảnh dựa trên từ vựng thị giác và Ontology. Đề xuất cho mọi hình ảnh tìm kiếm, xây dựng từ vựng thị giác và Ontology dựa trên các chú thích hình ảnh. Quá trình tìm kiếm hình ảnh được thực hiện bằng cách tích hợp cả đặc trưng thị giác và ngữ nghĩa tương tự. Từ vựng thị giác được lấy ra từ các đặc trưng cấp thấp của hình ảnh và ảnh được tìm kiếm dựa vào khoảng cách Euclide. Các Ontology được làm phong phú thêm bởi các khái niệm và mối quan hệ được trích ra từ tài nguyên từ vựng của BabelNet, thực hiện tìm kiếm ảnh trên Ontology dựa vào độ đo ngữ nghĩa. Kết quả tìm kiếm trên đặc trưng cấp thấp và trên Ontology được giao lại để tìm ra tập ảnh tương tự về mặt nội dung cấp thấp và ngữ nghĩa cấp cao. Hệ thống thực hiện xây dựng Ontology từ 20.000 hình ảnh của tập ảnh ImageCLEF, bởi đây là tập ảnh có một danh sách phân lớp ngữ nghĩa lớn, đại diện cho hầu hết các phân lớp ảnh có thể có. Tuy nhiên, thực nghiệm chưa được đánh giá nên tính hiệu quả của hệ thống về độ chính xác chưa kiểm chứng. Gonçalves, F. M. F. và cộng sự (2018) [11] đề xuất một phương pháp tiếp cận ngữ nghĩa kết hợp CBIR, học không giám sát và kỹ thuật Ontology. Đối với mô hình ngữ nghĩa, Ontology miền đại diện được xây dựng cho các đặc tính và cấu trúc của thực vật. Một cách tiếp cận dựa trên đồ thị được đề xuất để kết hợp thông tin ngữ nghĩa và kết quả tìm kiếm nội dung trực quan cấp thấp. Phương pháp đề xuất được thực nghiệm đánh giá trên tập dữ liệu Oxford Flowers với 102 phân lớp để chứng minh tính hiệu quả. Độ chính xác khi tìm kiếm ảnh có Ontology (0,8539) cao hơn so với CBIR và học không giám sát (0,8020) chứng tỏ, sự kết hợp tìm kiếm ảnh từ đặc trưng cấp thấp với ngữ nghĩa cấp cao trên Ontology cho hiệu suất vượt trội. Tuy nhiên, Ontology này xây dựng trên miền nhỏ, chưa thực hiện định nghĩa các khái niệm miền và chưa làm rõ được mối quan hệ giữa các miền giá trị. Mazo, C. và cộng sự (2020) [16] sử dụng Ontology của hệ thống tìm mạch của con người để cải thiện việc phân loại các hình ảnh mô học với phương pháp kết hợp SVM. Phương pháp này cải thiện việc phân loại tự động các hình ảnh mô học và có khả năng nhận ra các mô biểu mô, trước đây không được phát hiện bởi bất kỳ phương pháp thị giác máy tính nào khi so sánh đồng thời, bao gồm một đề xuất của CNN có tên là HistoResNet. Zhang, M. và cộng sự (2020) [28] đề xuất một phương pháp nhận dạng tự động kết hợp phát hiện đối tượng và Ontology để xác định rủi ro

trong quá trình xây dựng và ngăn ngừa tai nạn xây dựng. Đầu tiên, mạng CNN được dùng để trích xuất thông tin ngữ nghĩa cấp thấp từ các phần tử ngữ cảnh và các thuộc tính quan hệ không gian phần tử từ hình ảnh xuất từ video giám sát. Tiếp đến, Ontology được thiết lập trong phạm vi của một ngữ cảnh và ngôn ngữ logic của Ontology được dùng để chuyển đổi thông tin ngữ nghĩa cấp thấp của hình ảnh thành ngữ nghĩa cấp cao. Thứ ba, các quy tắc rủi ro được chuyển thành các quy tắc Ontology, và các tình huống rủi ro cao có thể phát sinh thực tế được xác định bởi một công cụ suy luận Pellet. Cuối cùng, một cảnh đào hồ móng được lấy làm ví dụ, và kết quả thử nghiệm được sử dụng để xác minh tính khả thi và hiệu quả của phương pháp đề xuất. Phương pháp đề xuất có thể được dùng để nâng cao hiệu quả quản lý an toàn xây dựng. Các nghiên cứu về tìm kiếm ảnh theo tiếp cận ngữ nghĩa được khảo sát, phân tích cho thấy tính hiệu quả của phương pháp tìm kiếm ảnh dựa trên Ontology. Hệ tìm kiếm ảnh trên Ontology cho độ chính xác cao hơn so với tìm kiếm thông thường dựa trên nội dung. Tuy nhiên, việc chỉ sử dụng văn bản đầu vào để truy vấn ngữ nghĩa trên Ontology cho kết quả không tốt bằng sự kết hợp giữa ngữ nghĩa trên Ontology và đặc trưng cấp thấp mô tả nội dung ảnh. Với sự kết hợp này, kết quả truy vấn không chỉ tương tự về nội dung thị giác, mà còn tương tự về mặt ngữ nghĩa. Ontology hiệu quả khi xây dựng trên tập dữ liệu lớn, có nhiều phân lớp đại diện cho nhiều đối tượng trong thực tế như bộ ảnh ImageCLEF, Oxford Flowers,... Từ các ưu nhược điểm của các công trình nghiên cứu liên quan, chúng tôi đề xuất phương pháp tìm kiếm ảnh tiếp cận ngữ nghĩa dựa trên Ontology và nội dung cấp thấp của hình ảnh, sử dụng tập ảnh ImageCLEF là tập ảnh mục tiêu để xây dựng khung Ontology. Đồng thời đề xuất phương pháp bổ sung dữ liệu để làm phong phú thêm ngữ nghĩa cho khung Ontology này.

### III. PHƯƠNG PHÁP XÂY DỰNG ONTOLOGY CHO DỮ LIỆU ẢNH

Dữ liệu ảnh luôn tăng trưởng không ngừng, có sự biến đổi thường xuyên của miền và mối quan hệ cấp bậc trong mỗi bộ ảnh, do đó xây dựng Ontology cho hình ảnh là một nhiệm vụ khó khăn, đòi hỏi nguồn nhân lực và thời gian lớn. Xây dựng Ontology phải đảm bảo việc bổ sung dữ liệu đúng để làm phong phú ngữ nghĩa hình ảnh. Chúng tôi đề xuất xây dựng một khung ontology bán tự động có khả năng bổ sung dữ liệu cho bài toán tìm kiếm ảnh theo tiếp cận ngữ nghĩa.

#### 1. Mô hình xây dựng Ontology bán tự động cho tập dữ liệu ảnh

Một mô hình xây dựng khung Ontology (Frame Ontology) bán tự động cho ảnh được đề xuất trong Hình 1, sử

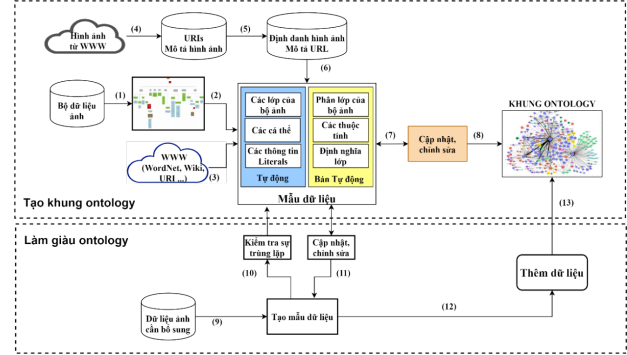
dụng tập ảnh ImageCLEF, là bộ ảnh phong phú về miền (276 lớp) làm nền tảng, kế thừa các hệ thống phân cấp ảnh, có khả năng mở rộng miền và mô tả tốt nhất các quan hệ của ảnh với phân lớp của nó. Khung Ontology bao gồm các thành phần: lớp, phân cấp lớp, các thuộc tính dữ liệu, thuộc tính quan hệ, các thông tin literals cho lớp và hình ảnh, từ điển đồng nghĩa. Các mẫu dữ liệu được tạo ra dựa trên các thành phần này (Hình 2), làm dữ liệu đầu vào cho quá trình xây dựng khung Ontology bán tự động. Mô hình tạo khung Ontology bán tự động bao gồm các bước như sau:

- **Bước 1.** Xác định tập ảnh, kế thừa lớp, phân cấp lớp (1);
- **Bước 2.** Xây dựng mẫu dữ liệu cho các thành phần của khung Ontology, được tạo tập ảnh (2) và từ thông tin WWW (3):  
+ Các lớp, các cá thể và thông tin Literals được tạo tự động từ tập ảnh;  
+ Phân lớp, các thuộc tính, từ điển được tạo bán tự động từ tập ảnh và WWW;
- **Bước 3.** Ảnh từ WWW, kết xuất các định danh tài nguyên và mô tả (4), các định danh ảnh và URL (5), bổ sung vào các dữ liệu đầu vào (6) cho quá trình tạo Ontology;
- **Bước 4.** Các dữ liệu được tạo từ bước 3 và bước 4 được cập nhật, chỉnh sửa với sự tham gia của chuyên gia (7) và tạo khung Ontology bán tự động (8);

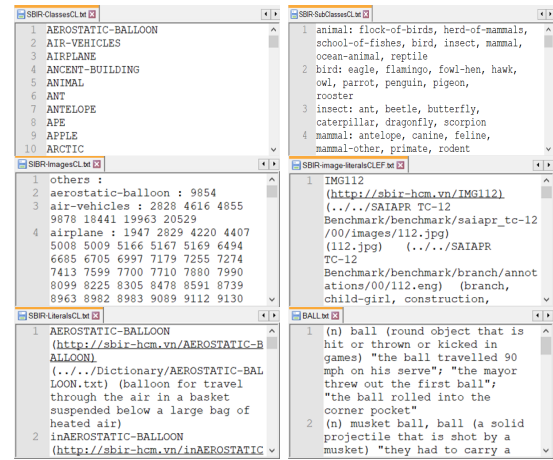
Trong mô hình xây dựng khung Ontology bán tự động, chuyên gia đóng vai trò quan trọng nhằm đảm bảo tính đúng đắn, tin cậy cho Ontology do các dữ liệu được bổ sung thực hiện theo qui trình xử lý dữ liệu, làm sạch dữ liệu, kiểm chứng, kiểm thử dữ liệu,... Bước tham gia chỉnh sửa của chuyên gia được thực hiện thủ công dựa trên các thao tác tự động trước đó như sau:

- + Tạo mẫu các phân lớp của bộ ảnh (SubClass) từ mẫu tập tin lớp (Class), kế thừa từ tập ảnh và từ WWW như cây phân cấp lớp (Hierarchy tree) của ImageNet [13], Ontology (taxonomy) của Dbpedia [6];...
- + Tạo mẫu các định nghĩa lớp hay từ điển đồng nghĩa dựa vào mạng từ WORDNET. Từ WORDNET, tự động lấy về định nghĩa tương ứng của các lớp, sau đó, chuyên gia lọc lại các định nghĩa cho các danh từ và các đồng nghĩa của lớp.
- + Tạo mẫu các thuộc tính quan hệ giữa lớp và các thể hiện (instances) của nó, xác định hình ảnh đại diện phù hợp cho lớp và định nghĩa chính của lớp tùy theo ngữ cảnh của bộ ảnh.

Ontology được tạo bán tự động với sự tham gia chỉnh sửa của chuyên gia có các ưu điểm là giảm chi phí, nhân lực nhưng vẫn đảm bảo tính đúng đắn và tin cậy. Trong khi đó, việc tạo Ontology thủ công trên protégé là không khả



Hình 1. Mô hình xây dựng và bổ sung dữ liệu cho một khung Ontology



Hình 2. Các dữ liệu mẫu cho khung Ontology bán tự động

thi khi số lượng ảnh nhiều, chi phí và nhân lực rất lớn, còn tạo Ontology tự động thì chi phí và nhân lực được giảm đi rất nhiều nhưng không đảm bảo tính chính xác và tin cậy. Các thuật toán để xây dựng Ontology bao gồm tạo các lớp, lớp phân cấp, các cá thể được mô tả như sau:

### Thuật toán 1: Tạo lớp cho Ontology (COC).

```

1  Đầu vào: C = {classi, i = 1..N}, Ontology;
2  Đầu ra: Ontology;
3  begin
4      foreach class in C do
5          Sub = "sbir:" + Class;
6          Pre = "rdf:type";
7          Obj = "owl:" + "Class";
8          Ontology = Ontology ∪ Triple(Sub, Pre, Obj);
9      end
10     return Ontology;
11 end
  
```

## 2. Bổ sung dữ liệu ảnh cho khung Ontology

Bổ sung dữ liệu cho khung Ontology là bổ sung các mô tả ngữ nghĩa và mở rộng cấu trúc của Ontology, làm cho



### Thuật toán 2: Tạo phân cấp lớp cho Ontology (COCS).

```

1 Đầu vào: superClass, subClass, Ontology;
2 Đầu ra: Ontology;
3 begin
4   foreach class in C do
5     Sub = "sbir:" + subClass;
6     Pre = "rdfs:subClassOf ";
7     Obj = "owl:" + " superClass ";
8     Ontology = Ontology  $\cup$  Triple(Sub, Pre, Obj);
9   end
10  return Ontology;
11 end

```

### Thuật toán 3: Tạo các cá thể cho Ontology (CIC).

```

1 Đầu vào: C = {classi, i=1..N}, Individual, Ontology;
2 Đầu ra: Ontology;
3 begin
4   foreach class in C do
5     Sub = "sbir:" + Individual.ElementAt(i);
6     Pre = "rdf:type";
7     Obj = "owl:NamedIndividual";
8     Ontology = Ontology  $\cup$  Triple(Sub, Pre, Obj);
9   end
10  return Ontology;
11 end

```

nó chứa nhiều thông tin hơn, giàu ngữ nghĩa hơn. Các lớp, phân cấp lớp, các thuộc tính, quan hệ, các cá thể và các mô tả ngữ nghĩa được bổ sung bán tự động cho khung Ontology ban đầu. Việc bổ sung này đảm bảo các lớp, phân cấp lớp, các thuộc tính, các định nghĩa trong từ điển và các cá thể mới có thể được thêm vào khung Ontology mà không phá vỡ cấu trúc của nó. Mô hình bổ sung dữ liệu cho khung Ontology được đề xuất trong Hình 1. Mô hình này bao gồm hai giai đoạn, giai đoạn tạo khung Ontology (từ bước 1-8) và giai đoạn bổ sung cho khung Ontology. Quá trình bổ sung dữ liệu cho Ontology được thực hiện như sau:

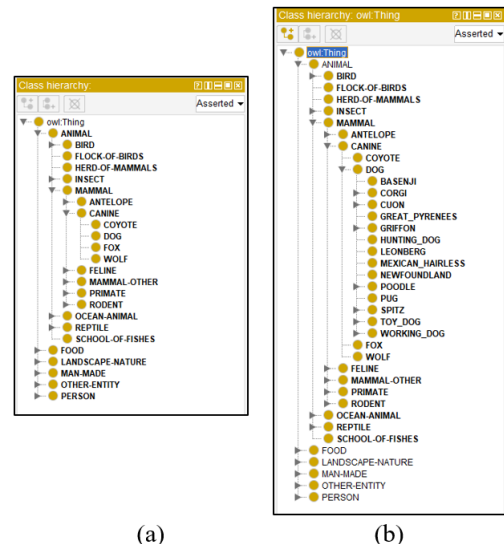
- **Bước 1.** Xác định tập ảnh ảnh để bổ sung cho một khung Ontology đã xây dựng;
- **Bước 2.** Dựa vào mẫu dữ liệu được tạo từ các thành phần của khung Ontology, tạo mẫu dữ liệu để bổ sung bán tự động cho khung Ontology (9);
- **Bước 3.** Kiểm tra tự động sự trùng lặp về lớp, cá thể, thuộc tính... giữa mẫu dữ liệu cần bổ sung vào mẫu dữ liệu của khung Ontology (10).
- **Bước 4.** Cập nhật, chỉnh sửa mẫu dữ liệu với sự tham gia của chuyên gia (11);
- **Bước 5.** Dữ liệu được thêm (12) vào khung Ontology để làm phong phú thêm cho ngữ nghĩa các tập ảnh (13).

Ứng dụng tạo khung Ontology và bổ sung dữ liệu SBIR-

Ontology được xây dựng trên mô hình đã đề xuất, trên nền tảng dotNET Framework 4.8, ngôn ngữ lập trình C#, lưu trữ bằng Notation 3 syntax (N3). Khung Ontology được tạo từ tập dữ liệu ảnh ImageCLEF (20.000 ảnh) và bổ sung thêm với tập ảnh Stanford Dogs (20.580 ảnh), lưu trữ tại tập tin SBIR-Ontology.n3 (Hình 3).

Hình 3. Ontology theo định dạng N3

Ví dụ của quá trình bổ sung dữ liệu được thể hiện trong Hình 4, cho thấy phân cấp trước và sau khi bổ sung vào khung Ontology. Ban đầu (Hình 4a) phân lớp của khung Ontology có lớp DOG và lớp này không có các lớp con, sau đó, khi bổ sung dữ liệu từ tập ảnh các loại chó của Stanford Dogs, thì phân lớp này tạo thêm các con (Hình 4b). Việc bổ sung này đảm bảo các lớp, phân cấp lớp, các thuộc tính, các định nghĩa trong từ điển và các cá thể mới có thể được thêm vào khung Ontology mà không phá vỡ cấu trúc ban đầu.



Hình 4. Trực quan phân cấp lớp trước (a) và sau (b) khi bổ sung dữ liệu vào Ontology

## IV. HỆ TÌM KIẾM ẢNH DỰA TRÊN ONTOLOGY

## 1. Mô hình tìm kiếm ảnh

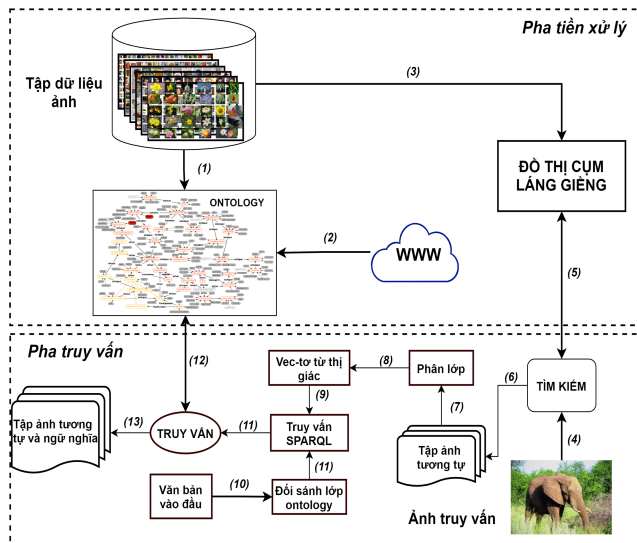
Một hệ tìm kiếm theo ngữ nghĩa dựa trên Ontology, onSBIR, được xây dựng bao gồm 2 pha được mô tả trong Hình 5 như sau:

### • Pha tiền xử lý:

- (1) Các hình ảnh trong tập ảnh được tổ chức lưu trữ trên cấu trúc đồ thị cụm láng giềng;
- (2) Xây dựng, làm giàu một Ontology bán tự động dựa trên ngôn ngữ bộ ba RDF từ tập ảnh và từ WWW;

### • Pha tìm kiếm:

- (1) Với một ảnh đầu vào, hệ thống thực hiện tìm kiếm trên đồ thị cụm láng giềng lấy ra tập các ảnh tương tự theo nội dung được sắp xếp theo độ đo, phân lớp tập ảnh này với thuật toán k-NN để tìm tập từ vựng thị giác; Với đầu vào là một đoạn văn bản, hệ thống thực hiện tìm các từ khóa phù hợp với các lớp trong Ontology;
- (2) Tự động tạo câu truy vấn SPARQL để tìm kiếm ảnh trên Ontology;
- (3) Kết xuất tập ảnh tương tự theo ngữ nghĩa trên Ontology.



Hình 5. Mô hình hệ tìm kiếm theo tiếp cận ngữ nghĩa onSBIR

## 2. Các thành phần của hệ truy vấn onSBIR

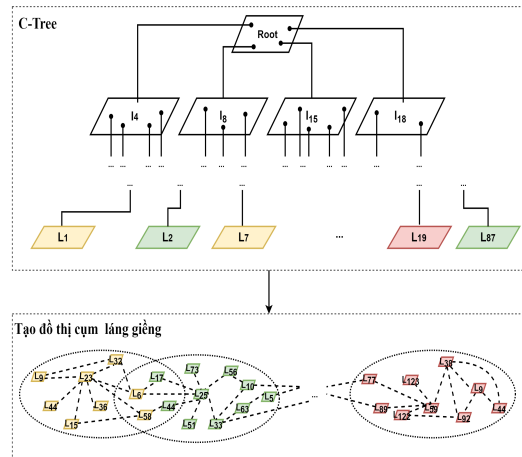
### a) Trích xuất đặc trưng cấp thấp

Độ chính xác của tìm kiếm ảnh phụ thuộc vào sự lựa chọn các đặc trưng cấp thấp, do đó, để phát huy các đặc trưng trong hệ thống tìm kiếm ảnh, chúng tôi sử dụng các đặc trưng cấp thấp về màu sắc, kết cấu, hình dạng để tạo thành bộ mô tả đặc trưng kết hợp màu sắc được trích xuất dựa trên bộ mô tả màu chủ đạo DCD của MPEG-7, đặc trưng kết cấu

được trích xuất dựa vào độ tương phản, phép lọc tần số cao, phép lọc Sobel, phép lọc Gaussian và phương pháp LoG, đặc trưng hình dạng dựa trên phương pháp Laplacian... Một vec-tơ đặc trưng cấp thấp gồm 81 chiều được trích xuất cho bài toán tìm kiếm ảnh trong bài báo này.

### b) Đồ thị cụm láng giềng

Đồ thị cụm láng giềng được cải tiến từ cấu trúc cây phân cụm cân bằng C-Tree [18, 19] đã được chúng tôi xây dựng. C-Tree có thể lưu trữ được dữ liệu lớn, hiệu quả cho bài toán tìm kiếm ảnh với thời gian tìm kiếm nhanh, độ chính xác khá cao. Tuy nhiên, nhược điểm chính của C-Tree là mỗi lần tách nút, các phần tử tương tự nhau có thể bị tách thành các nút khác nhau, trong trường hợp xấu nhất, các nút này có thể nằm trên các nhánh riêng biệt. Do đó, quá trình tìm kiếm ảnh trên cây C-Tree có thể sẽ bỏ sót các phần tử tương tự đã bị chuyển nhánh. Điều này làm ảnh hưởng đến hiệu suất tìm kiếm trên C-Tree. Vì vậy, đồ thị gom cụm láng giềng (Hình 6), là sự kết hợp giữa đồ thị gom cụm và cây C-Tree, được đề xuất như sau: mỗi nút lá trên cây C-Tree trong quá trình huấn luyện sẽ tìm được các cụm láng giềng lân cận theo ba mức: mức một nếu khoảng cách giữa hai tâm nút bé hơn  $\theta$ , mức hai nếu lớp đại diện của hai nút trùng nhau và mức ba nếu lớp đại diện của hai nút có quan hệ phân lớp cha-con. Các mức láng giềng được đánh dấu để tạo thành một đồ thị cụm liên quan với nhau về độ đo và quan hệ cấp bậc (cha-con). Đồ thị gom cụm láng giềng cho phép thời gian tìm kiếm nhanh hơn và chi phí bộ nhớ ít hơn so với đồ thị gom cụm truyền thống.



Hình 6. Đồ thị gom cụm láng giềng

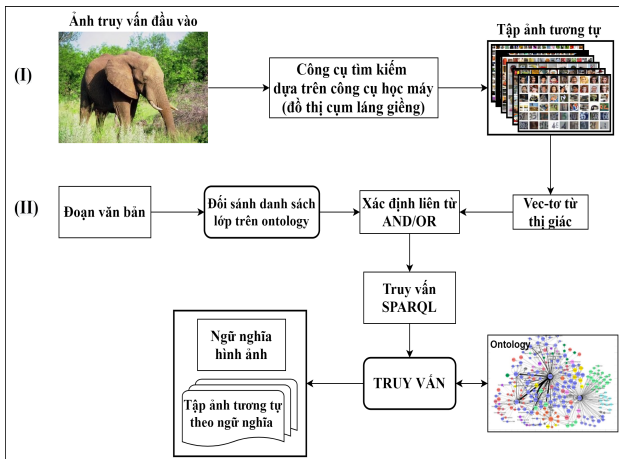
### c) Tìm kiếm ảnh dựa trên Ontology

Tìm kiếm ảnh theo tiếp cận ngữ nghĩa dựa trên Ontology được thực hiện theo hai cách: tìm kiếm dựa trên ảnh đầu vào và tìm kiếm dựa đoạn văn bản, mô tả trong Hình 7.

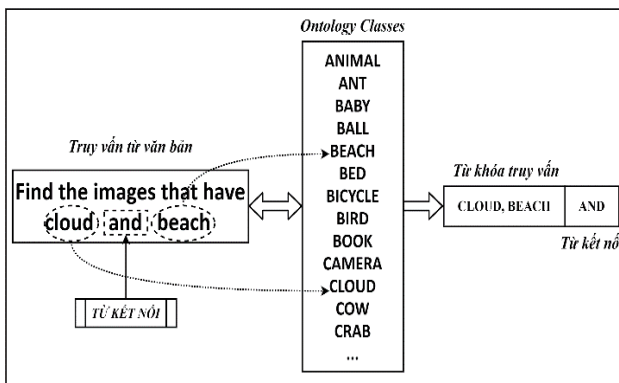
- (I) **Tìm kiếm từ ảnh đầu vào:** ảnh đầu vào có thể chứa một đối tượng hoặc nhiều đối tượng, dựa vào đồ thị

cụm láng giềng để tìm tập ảnh tương tự. Sau đó thực hiện thuật toán phân lớp k-NN để tìm tập từ vựng thị giác, là vec-tơ chứa một hay nhiều lớp ngữ nghĩa của ảnh và tự động tạo câu lệnh SPARQL (AND hoặc OR), từ đó truy vấn trên Ontology để tìm tập các ảnh tương tự và ngữ nghĩa của nó;

(II) **Tìm kiếm từ văn bản đầu vào:** Từ một đoạn văn bản đầu vào, thực hiện đối sánh danh sách các lớp Ontology để tìm từ khóa truy vấn và từ kết nối AND hoặc OR trong ngữ nghĩa đoạn văn bản, từ đó tự động tạo câu lệnh SPARQL để thực hiện tìm kiếm trên Ontology, truy xuất tập các ảnh tương tự và ngữ nghĩa của tập ảnh. Hình 8 là một ví dụ về phân tích đoạn văn bản nhằm tìm các từ khóa cho câu truy vấn SPARQL.



Hình 7. Các cách tìm kiếm ảnh trên Ontology



Hình 8. Một ví dụ về phân tích văn bản tìm kiếm

Thuật toán phân tích văn bản text để tìm các từ khóa cho truy vấn SPRQL được thực hiện như sau:

Thuật toán tự động tạo câu truy vấn SPRQL được thực hiện như sau:

**Thuật toán 4:** Trích xuất từ khóa tìm kiếm từ text (GLCT).

```

1 Đầu vào: Text và danh sách lớp LClass
2 Đầu ra: danh sách từ khóa ListClass
3 begin
4   text=txtQuery.Text;
5   Classes=text.Split;
6   strClasses=LClass.Split;
7   foreach (cla ∈ Class) do
8     If (cla ∈ strClass) then
9       ListClass.Add(cla);
10      CSP(ListClass);
11    end
12  end
13 end

```

**Thuật toán 5:** Tự động tạo câu truy vấn SPARQL (CSP)

```

1 Đầu vào: tập từ vựng thị giác W
2 Đầu ra: câu lệnh SPARQL
3 begin
4   SPARQL= ∅;
5   n = W.count;
6   SELECT DISTINCT ?Im
7   WHERE {";
8   For (i=0..n) do
9     SPARQL+="<subject > : W(i)+ "rdf:type" +
10    <object >: +?Img'+ "UNION/AND";
11   end
12   SPARQL+=}";
13   Return SPARQL;
14 end

```

## V. THỰC NGHIỆM VÀ ĐÁNH GIÁ

### 1. Môi trường thực nghiệm

Hệ truy vấn OnSBIR được xây dựng dựa trên nền tảng dotNET Framework 4.8, ngôn ngữ lập trình C#. Cấu hình máy tính của thực nghiệm: Intel(R) Core™ i7-8750H, CPU 2,70GHz, RAM 8GB. Tập dữ liệu sử dụng trong thực nghiệm là ImageCLEF và StanfordDogs, được mô tả trong Bảng I.

Tập ảnh ImageCLEF bao gồm 20.000 ảnh đa đối tượng, mỗi ảnh có nhiều đối tượng với nhiều chủ đề khác nhau. Tập ảnh này có 276 lớp khác nhau và được trộn lẫn với nhau trong 39 thư mục. Tập ảnh Stanford Dogs gồm 20.580 ảnh, thuộc hệ thống ImageNet, do nhóm nghiên cứu A. Khosla xây dựng vào năm 2011. Đây là một bộ ảnh chỉ có một đối tượng, bao gồm 120 loại chó trên thế giới.

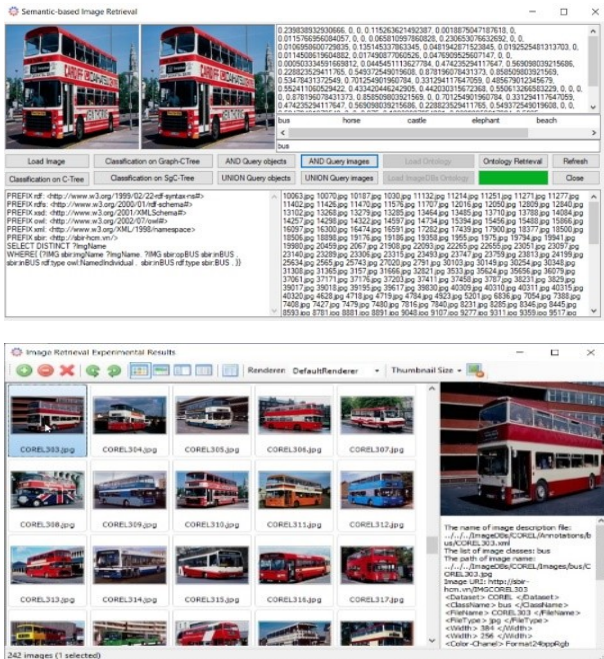


Bảng I  
CÁC TẬP DỮ LIỆU ẢNH ĐƯỢC THỰC NGHIỆM

| STT | Tên tập ảnh   | Số lượng ảnh | Số thư mục ảnh | Số chủ đề ảnh | Kích thước |
|-----|---------------|--------------|----------------|---------------|------------|
| 1   | ImageCLEF     | 20.000       | 39             | 276           | 1,6 GB     |
| 2   | Stanford Dogs | 20.580       | 120            | 120           | 778 MB     |

## 2. Ứng dụng

Với một ảnh đầu vào, đặc trưng ảnh được trích xuất để tìm tập ảnh tương tự dựa trên đồ thị cụm láng giềng, phân lớp ảnh để tìm tập từ vựng thị giác, sau đó tự động tạo câu lệnh SPARQL để tìm danh sách các ảnh có cùng ngữ nghĩa với ảnh đầu vào, được thể hiện trong Hình 9. Đồng thời, khái niệm ngữ nghĩa của phân lớp được trích xuất từ tập từ điển đồng nghĩa của Ontology (Hình 10).



Hình 9. Tìm kiếm ảnh trên Ontology từ ảnh đầu vào

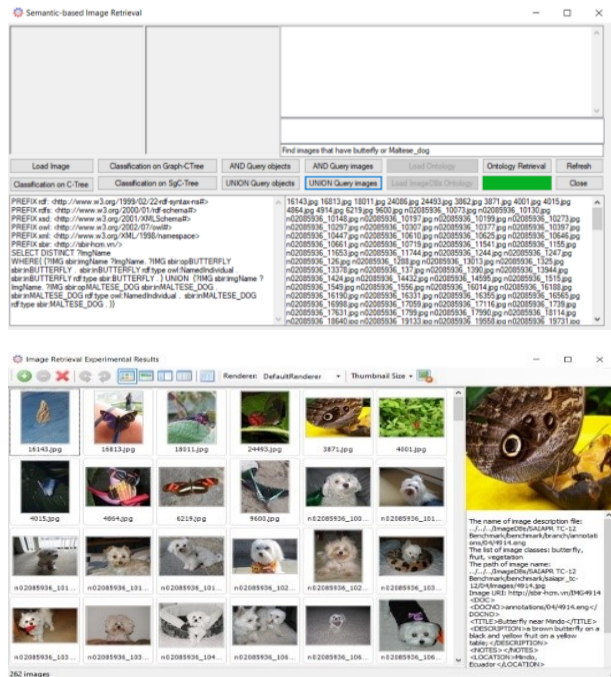
Với đầu vào là một đoạn văn bản, hệ thống đối sánh với các lớp có trong Ontology và từ vựng kết nối (AND/OR) để tìm các từ khóa và tạo câu truy vấn SPARQL tự động để tìm kiếm trên Ontology tập các hình ảnh tương tự về ngữ nghĩa (Hình 11).

## 3. Đánh giá kết quả

Để đánh giá hiệu quả tìm kiếm ảnh trên Ontology của hệ tìm kiếm onSBIR, các giá trị như độ chính xác (precision), độ phủ (recall), độ dung hòa (F-measure) và thời gian tìm



Hình 10. Khái niệm ngữ nghĩa cho phân lớp



Hình 11. Tìm kiếm ảnh trên Ontology từ văn bản đầu vào

kiếm được tính toán. Bảng II là hiệu suất tìm kiếm trên Ontology của các tập ảnh ImageCLEF và Stanford Dogs. Kết quả tìm kiếm ảnh dựa trên Ontology được so sánh với phương pháp tìm kiếm ảnh dựa vào đồ thị cụm láng giềng (Bảng III). Bảng III cho thấy độ chính xác, độ phủ của hệ thống tìm kiếm ảnh dựa trên Ontology tốt hơn so với hệ thống tìm kiếm ảnh theo nội dung, tuy nhiên thời gian tìm kiếm lâu hơn, do OnSBIR kết hợp tìm kiếm trên cả đồ thị cụm láng giềng và Ontology. Điều này chứng tỏ, phương pháp tìm kiếm ảnh trên Ontology nâng cao độ chính xác và độ phủ cho tìm kiếm.

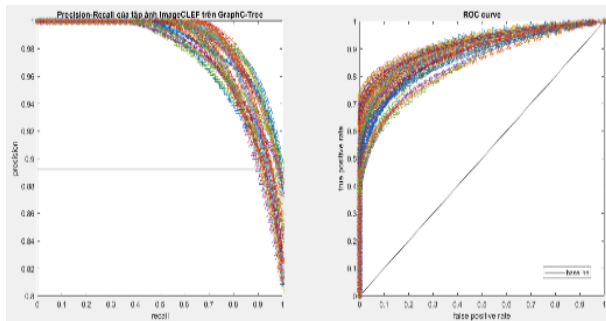


Bảng II  
HIỆU SUẤT TÌM KIẾM TRÊN ONTOLOGY

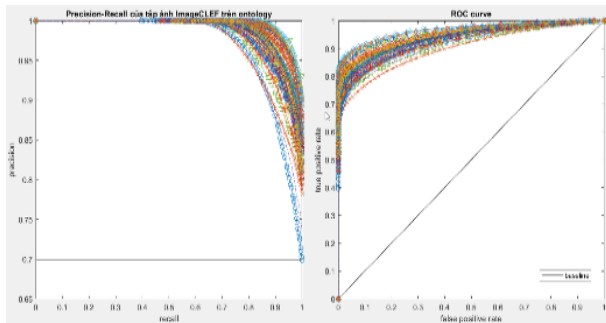
| Tập ảnh       | Độ chính xác trung bình | Độ phủ trung bình | Độ dung hòa trung bình | Thời gian truy vấn trung bình (ms) |
|---------------|-------------------------|-------------------|------------------------|------------------------------------|
| ImageCLEF     | 0,932574                | 0,916225          | 0,926373               | 248,5511                           |
| Stanford Dogs | 0,87395                 | 0,865369          | 0,869612               | 284,3384                           |

Bảng III  
SO SÁNH HIỆU SUẤT TÌM KIẾM ẢNH

| Tập ảnh       | Phương pháp           | Độ chính xác trung bình | Độ phủ trung bình | Thời gian truy vấn trung bình (ms) |
|---------------|-----------------------|-------------------------|-------------------|------------------------------------|
| ImageCLEF     | OnSBIR                | <b>0,932574</b>         | <b>0,916225</b>   | 248,5511                           |
|               | Đồ thị cụm láng giềng | 0,839814                | 0,780735          | <b>239,9458</b>                    |
| Stanford Dogs | OnSBIR                | <b>0,87395</b>          | <b>0,865369</b>   | 284,3384                           |
|               | Đồ thị cụm láng giềng | 0,826416                | 0,513825          | <b>261,8991</b>                    |



(a)



(b)

Hình 12. Hiệu suất trên đồ thị cụm láng giềng (a) và Ontology (b) của tập ảnh ImageCLEF

Ngoài ra, các đồ thị Precision-Recall và ROC dựa trên Ontology và dựa trên đồ thị cụm láng giềng (Hình 12, Hình 13) được biểu thị. Mỗi đường cong trên đồ thị Precision – Recall là một thư mục ảnh. Diện tích dưới đường cong càng lớn, độ chính xác truy vấn càng cao. Đồng thời, đường cong ROC cho biết tỷ lệ kết quả truy vấn đúng và sai, diện tích dưới đường cong ROC càng lớn, tỷ lệ truy vấn đúng càng cao. Kết quả thực nghiệm được thể hiện trên các đồ thị cho thấy, diện tích dưới đường cong đồ thị Precision-

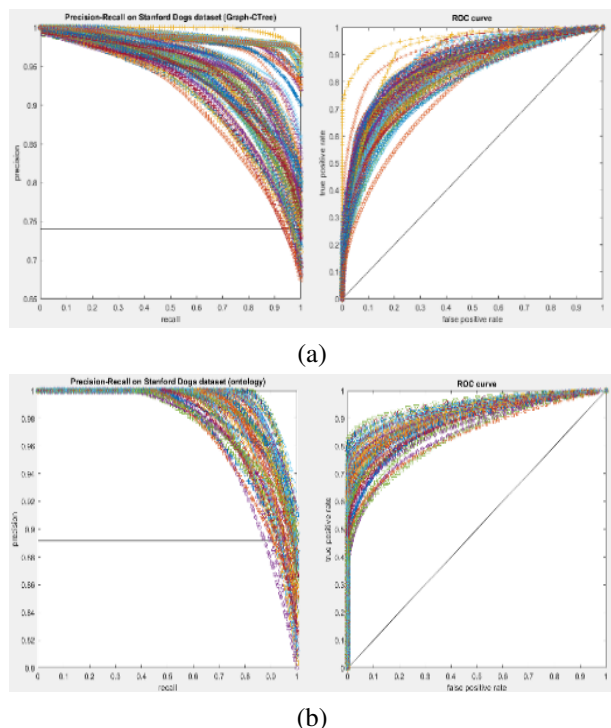
Bảng IV  
SO SÁNH ĐỘ CHÍNH XÁC TÌM KIẾM ẢNH GIỮA CÁC PHƯƠNG PHÁP TRÊN TẬP ẢNH IMAGECLEF

| Phương pháp                                                       | Năm  | Độ chính xác trung bình |
|-------------------------------------------------------------------|------|-------------------------|
| Mô hình Ontology cho ảnh O-V-A[23]                                | 2016 | 0,46                    |
| Đồ thị mẫu cho Ontology hình ảnh[1]                               | 2017 | 0,3513                  |
| Phương pháp HDLA (hybrid deep learning architecture)[3]           | 2018 | 0,797                   |
| Phương pháp SDCH (Semantic Deep Cross-modal Hashing)[27]          | 2019 | 0,803                   |
| Phương pháp CPAH (Consistency Preserving Adversarial Hashing)[26] | 2020 | 0,8324                  |
| <b>OnSBIR</b>                                                     |      | <b>0,932574</b>         |

Recall và ROC của OnSBIR đều lớn hơn so với kết quả của đồ thị cụm láng giềng, chứng tỏ độ chính xác và kết quả truy vấn đúng đều tốt hơn. Từ các kết luận này, có thể thấy tính đúng đắn của các phương pháp tìm kiếm ảnh dựa trên Ontology, nhằm nâng cao hiệu quả tìm kiếm ảnh.

Chúng tôi thực hiện so sánh kết quả tìm kiếm ảnh trên Ontology (onSBIR) với các phương pháp khác trên cùng tập ảnh ImageCLEF. Bảng IV cho thấy, độ chính xác tìm kiếm ảnh dựa trên Ontology được đề xuất trong bài báo vượt trội hơn các phương pháp khác trên tập ảnh ImageCLEF.

Chúng tôi so sánh với các phương pháp có sử dụng Ontology: Vijayarajan V. và cộng sự (2016) [23] chỉ thực hiện tìm kiếm trên Ontology dựa trên văn bản/từ khóa, do đó, hiệu suất tìm kiếm chưa cao; Nhóm nghiên cứu Allani O. (2017) [1] thực hiện kết hợp giữa đặc trưng cấp thấp trên đồ thị và ngữ nghĩa cấp cao trên Ontology, tuy



Hình 13. Hiệu suất trên đồ thị cụm láng giềng (a) và Ontology (b) của tập ảnh Stanford Dogs

nhien, việc xây dựng Ontology hoàn toàn tự động, không có sự chỉnh sửa, cập nhật của chuyên gia, do đó, hiệu suất tìm kiếm là không cao. Ngoài ra, các phương pháp tìm kiếm ngữ nghĩa ảnh dựa vào các kỹ thuật học máy hiện đại cũng được so sánh với phương pháp của chúng tôi: Phương pháp HDLA (Hybrid Deep Learning Architecture) mô hình hóa các mối tương quan bậc cao giữa các từ trực quan nhằm giảm khoảng cách ngữ nghĩa trong tìm kiếm ảnh [3]; Phương pháp SDCH (Semantic Deep Cross-modal Hashing) sử dụng mạng CNN để trích xuất đặc trưng và hàm băm sâu để lấy ngữ nghĩa ảnh [27]; Phương pháp CPAH (Consistency Preserving Adversarial Hashing) nhằm khai thác tính nhất quán về ngữ nghĩa cho tìm kiếm ảnh, các đặc trưng được trích xuất dựa trên mạng CNN[26]. Từ các kết quả so sánh chứng tỏ, phương pháp tìm kiếm ảnh dựa trên Ontology được đề xuất trong bài báo là đúng đắn và hiệu quả.

## VI. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong bài báo này, một khung Ontology bán tự động được xây dựng và bổ sung để làm phong phú thêm ngữ nghĩa. Bài toán tìm kiếm trên Ontology được thực hiện theo hai cách: tìm kiếm dựa trên ảnh đầu vào và tìm kiếm từ văn bản. Câu truy vấn SPARQL được tạo ra tự động để tìm kiếm trên Ontology. Hệ tìm kiếm ảnh theo ngữ nghĩa dựa trên Ontology (OnSBIR) được xây dựng. Thực nghiệm

được triển khai trên tập ảnh ImageCLEF và Stanford Dogs cho độ chính xác cao, lần lượt là 0,932574 và 0,87395. So sánh với phương pháp tìm kiếm ảnh theo nội dung dựa trên đồ thị cụm láng giềng cho thấy hiệu suất tìm kiếm ảnh dựa trên Ontology là vượt trội. Đồng thời, chúng tôi tiến hành so sánh phương pháp đề xuất với các phương pháp khác trên cùng tập ảnh ImageCLEF cũng cho độ chính xác tốt hơn. Điều này chứng tỏ, các đề xuất của chúng tôi trong bài báo là đúng đắn và hiệu quả. Tuy nhiên, phương pháp đề xuất của chúng tôi, Ontology chưa thực hiện mối quan hệ ngữ nghĩa trong ảnh, vì vậy, trong định hướng tương lai, chúng tôi bổ sung nhiều tập ảnh hơn, làm giàu ngữ nghĩa Ontology với các mối quan hệ trong ảnh, xác định vị trí đối tượng ảnh.

## LỜI CẢM ƠN

Chúng tôi xin trân trọng cảm ơn Khoa Công nghệ thông tin – Đại học Khoa học- Đại học Huế, nhóm nghiên cứu SBIR-HCM đã góp ý chuyên môn cho nghiên cứu này. Chúng tôi xin trân trọng cảm ơn Trường Đại học Kinh tế – Đại học Đà Nẵng đã tạo điều kiện về cơ sở vật chất giúp chúng tôi hoàn thành bài nghiên cứu này.

## TÀI LIỆU THAM KHẢO

- [1] Allani O., Zghal H. B., Mellouli N., Akdag H., "Pattern graph-based image retrieval system combining semantic and visual features," *Multimedia Tools and Applications*, vol. 76(19), pp. 20287-20316, 2017.
- [2] Alzubaidi M. A., "A new strategy for bridging the semantic gap in image retrieval," *International Journal of Computational Science and Engineering*, vol. 14(1), pp. 27-43, 2017.
- [3] Arun K. S., Govindan V. K., "A hybrid deep learning architecture for latent topic-based image retrieval," *Data Science and Engineering*, vol. 3(2), pp. 166-195, 2018.
- [4] Bella M. I. T., Vasuki A., "An efficient image retrieval framework using fused information feature," *Computers & Electrical Engineering*, vol. 75, pp. 46-60, 2019.
- [5] Bouchakwa M., Ayadi Y., Amous I., "Multi-level diversification approach of semantic-based image retrieval results," *Progress in Artificial Intelligence*, vol. 9(1), pp. 1-30, 2020.
- [6] Dbpedia ontology, <https://dbpedia.org/ontology/>, last accessed 2021/03/20.
- [7] Fang W., Ding L., Zhong B., Love P. E., & Luo, H., "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach," *Advanced Engineering Informatics*, vol. 37, pp. 139-149, 2018.
- [8] Filali J., Zghal H., & Martinet J., "Towards Visual Vocabulary and Ontology-based Image Retrieval System," *In International Conference on Agents and Artificial Intelligence*, vol. 2, pp. 560-565, 2016.
- [9] Jia F., Liu J., Tai X. C., "A regularized convolutional neural network for semantic image segmentation," *Analysis and Applications*, vol. 19(01), pp. 147-165, 2021.
- [10] Garg M., Dhiman G., "A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants," *Neural Computing and Applications*, pp. 1-18, 2020.

- [11] Gonçalves F. M. F., Guilherme I. R., Pedronette D. C. G., "Semantic guided interactive image retrieval for plant identification," *Expert Systems with Applications*, vol. 91, pp.12-26, 2018.
- [12] ImageCLEF Homepage, <https://www.imageclef.org/>, last accessed 2021/03/21.
- [13] ImageNET Hierarchy, <https://observablehq.com/@mbostock/imagenet-hierarchy>, last accessed 2021/03/20.
- [14] Khosla A., Jayadevaprakash N., Yao B., Li F. F., "Novel dataset for fine-grained image categorization: Stanford dogs," In: *Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC)*, vol. 2, no. 1, 2011.
- [15] Manzoor U., Balubaid M. A., Zafar B., Umar H., Khan M. S., "Semantic image retrieval: An ontology based approach," *International Journal of Advanced Research in Artificial Intelligence*, vol. 4(4), pp. 1-8, 2015.
- [16] Mazo C., Alegre E., Trujillo M., "Using an ontology of the human cardiovascular system to improve the classification of histological images," *Scientific Reports*, vol. 10(1), pp. 1-14, 2020.
- [17] Mohd K., Yanti I.A., Mohd N., Shahrul A., Bloechle M., "Semantic text-based image retrieval with multi-modality ontology and DBpedia," *The Electronic Library*, vol. 35, no. 6, pp. 1191-1214, 2017.
- [18] Nhi, N. T. U., Van T. T., Le T. M., "A self-balanced clustering tree for Semantic-based image retrieval," *Journal of Computer Science and Cybernetics*, vol. 36(1), pp. 49-67, 2020.
- [19] Nhi N. T. U., Van T. T., Le T. M., "Semantic-Based Image Retrieval Using Balanced Clustering Tree," In *WorldCIST*, vol.2, pp. 416-427, 2021.
- [20] Shati N. M., Khalid Ibrahim N., Hasan T. M., "A review of image retrieval based on ontology model," *Journal of Al-Qadisiyah for computer science and mathematics*, vol. 12(1), pp.10, 2020.
- [21] Sharma M. K., Siddiqui T. J., "An ontology based framework for retrieval of museum artifacts," *Procedia Comput. Sci.*, vol. 84, pp. 169-176, 2016.
- [22] Singh V., Gupta R., "Novel framework of semantic based image retrieval by convoluted features with non-linear mapping in cyberspace," *International Journal of Recent Technology and Engineering*, pp. 2277-3878, 2019.
- [23] Vijayarajan V., Dinakaran M., Tejaswin P., Lohani M., "A generic framework for ontology-based information retrieval and image retrieval in web data," *Human-centric Computing and Information Sciences*, vol 6(1), pp. 18-27, 2016.
- [24] WORDNET Homepage, <https://wordnet.princeton.edu/>, last accessed 2021/03/17
- [25] Xu H., Huang C., Wang D., "Enhancing semantic image retrieval with limited labeled examples via deep learning," *Knowledge-Based Systems*, vol. 163, pp. 252-266, 2019.
- [26] Xie D., Deng C., Li C., Liu X., Tao D., "Multi-task consistency-preserving adversarial hashing for cross-modal retrieval," *IEEE Transactions on Image Processing*, vol. 29, pp. 3626-3637, 2020.
- [27] Yan C., Bai X., Wang S., Zhou J., Hancock E. R., "Cross-modal hashing with semantic deep embedding," *Neurocomputing*, vol. 337, pp. 58-66, 2019.
- [28] Zhang M., Zhu M., Zhao X., "Recognition of High-Risk Scenarios in Building Construction Based on Image Semantics," *Journal of Computing in Civil Engineering*, vol. 34(4), 2020.
- [29] Zhang S., Ma Z., Zhang G., Lei T., Zhang R., Cui Y., "Semantic image segmentation with deep convolutional neural networks and quick shift," *Symmetry*, vol. 12(3), pp. 427-439, 2020.

## SƠ LƯỢC VỀ TÁC GIẢ

### Nguyễn Thị Uyên Nhi



Sinh năm 1985, nhận bằng cử nhân và thạc sĩ chuyên ngành Khoa học máy tính và kỹ thuật tính toán, tại trường Đại học tổng hợp kỹ thuật Volgograd, Liên bang Nga, lần lượt vào các năm 2008, 2010. Hiện đang là NCS ngành Khoa học máy tính tại Trường đại học Khoa học, Đại học Huế. Lĩnh vực nghiên cứu: xử lý ảnh, tìm kiếm ảnh, cơ sở dữ liệu mờ.

Email: nhintu@due.edu.vn

### Văn Thế Thành



Sinh năm 1979, tốt nghiệp chuyên ngành Toán tin tại Đại học Khoa học Tự nhiên - Đại học Quốc gia TP.HCM vào năm 2001, nhận bằng Thạc sĩ Khoa học Máy tính tại Đại học Quốc gia TP.HCM vào năm 2008. Năm 2016, nhận bằng Tiến sĩ Khoa học Máy tính tại Đại học Khoa học, Đại học Huế. Lĩnh vực nghiên cứu: xử lý ảnh, khai thác dữ liệu ảnh và tìm kiếm ảnh.

Email: thanhvt@hufi.edu.vn

### Lê Mạnh Thạnh



Sinh năm 1953, nhận bằng Tiến sĩ ngành khoa học máy tính tại Đại học Budapest (ELTE), Hungary vào năm 1994. Ông trở thành Phó giáo sư tại trường Đại học Khoa học, Đại học Huế, Việt Nam, vào năm 2004. Lĩnh vực nghiên cứu: cơ sở dữ liệu, cơ sở tri thức và lập trình logic.

Email: lmthanh@hueuni.edu.vn