



TÌM KIẾM ẢNH THEO NGỮ NGHĨA DỰA TRÊN CẤU TRÚC iR^S -TREE VÀ ONTOLOGY

Lê Thị Vĩnh Thanh^{1,3}, Lê Mạnh Thanh¹, Văn Thế Thành^{2*}

¹Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế, Việt nam

²Khoa Công nghệ thông tin, Trường Đại học Sư phạm Tp. HCM, Việt Nam

³Trường Đại học Bà Rịa – Vũng Tàu, Việt Nam

Tóm tắt. Truy xuất hình ảnh theo nội dung được thực hiện dựa trên các đặc trưng cấp thấp và ứng dụng trong nhiều lĩnh vực. Tuy nhiên, thách thức lớn của phương pháp này là khoảng cách ngữ nghĩa giữa các đặc trưng cấp thấp và các khái niệm cấp cao. Vì vậy, bài toán truy xuất hình ảnh theo tiếp cận ngữ nghĩa được quan tâm nghiên cứu nhằm nâng cao độ chính xác truy vấn. Trong bài báo này, chúng tôi đề xuất một mô hình tìm kiếm ảnh và tra cứu ngữ nghĩa dựa trên cấu trúc R^S -Tree cải tiến (iR^S -Tree) và ontology. Các cải tiến trên iR^S -Tree bao gồm: (1) phép toán thêm phần tử nhằm giảm thời gian tạo cây; (2) phép toán tách nút nhằm nâng cao độ chính xác của việc gom cụm dữ liệu. Kết quả của việc truy vấn ảnh trên iR^S -Tree là một tập ảnh tương tự và tập từ vựng thị giác. Sau đó, câu lệnh SPAQRL được tự động tạo ra từ tập từ vựng thị giác này và thực hiện truy vấn trên một ontology bán tự động nhằm trích xuất ngữ nghĩa cho hình ảnh. Thực nghiệm được thực hiện trên các tập dữ liệu ảnh: COREL, Oxford Flower-102 và CUB-200-2011. Kết quả thực nghiệm được đánh giá và so sánh với các công trình khác trên cùng một tập dữ liệu nhằm chứng minh tính hiệu quả và đúng đắn của phương pháp đề xuất.

Keywords: Từ khoá: iR^S -Tree, SBIR, clustering, image retrieval, ontology

Content-based image retrieval based on iR^S -Tree and ontology

Le Thi Vinh Thanh^{1,3}, Le Manh Thanh¹, Van The Thanh^{2*}

¹Faculty of Information Technology, University of Sciences, Hue University, Vietnam

²Faculty of Information Technology, HCMC University of Education, Vietnam

³Ba Ria–Vung Tau University, Vietnam

Abstract. Content-based image retrieval is implemented based on low-level features and applied to numerous fields. However, the major challenge of this approach is the semantic gap between low-level features and high-level concepts. Therefore, the problem of semantic-based image retrieval is attractive to improve query accuracy. In this paper, a semantic-based image retrieval model using the improved R^S -Tree structure (iR^S -Tree) and ontology was

* Liên hệ: thanhvt@hufi.edu.vn

proposed. The improvements on iR^S -Tree include: (1) improving the operation of adding elements to reduce tree building time; (2) improving the operation of splitting nodes to enhance clustering accuracy. The result of querying images on iR^S -Tree is a set of similar images and a set of visual vocabulary. Then, a SPARQL command is automatically generated from this visual vocabulary and queries a semi-automatic ontology to extract semantics for the image. Experiments were performed on three image data sets: COREL, Oxford Flower-102, and CUB-200-2011. Experimental results were evaluated and compared with recent works on the same data set to demonstrate the effectiveness and correctness of the proposed method.

Keywords: iR^S -Tree, SBIR, clustering, image retrieval, ontology

1 Giới thiệu

Ngày nay, với sự phát triển vượt bậc của các công nghệ Internet và các thiết bị điện thoại thông minh, số lượng hình ảnh kỹ thuật số đã tăng lên nhanh chóng. Do đó, để lưu trữ và tìm kiếm các hình ảnh hữu ích trong tập dữ liệu ảnh lớn, cần có một cấu trúc dữ liệu để đánh chỉ mục hình ảnh và một hệ thống truy vấn hiệu quả. Hiện nay, nhiều lĩnh vực khác nhau ứng dụng hệ thống tìm kiếm ảnh trong thực tế như hệ thống truy xuất khuôn mặt [1] giúp cảnh sát và các nhân viên an ninh tìm ra đối tượng phạm tội nhanh chóng hơn; hệ thống truy xuất hình ảnh hàng hóa [2] giúp khách hàng tìm thấy hàng hóa mà họ yêu thích và hệ thống truy xuất hình ảnh y tế [3] giúp bác sĩ chẩn đoán y tế hiệu quả hơn. Trong những năm gần đây, nhiều hệ thống tìm kiếm ảnh theo nội dung (Content-based Image Retrieval – CBIR) đã được phát triển và áp dụng cho các ứng dụng thực tế. Tuy nhiên, hệ thống CBIR tìm kiếm các tập ảnh tương tự nhau về nội dung dựa trên các đặc trưng cấp thấp như màu sắc, hình dạng, kết cấu và vị trí, nhưng chúng có thể khác nhau hoàn toàn về ngữ nghĩa [4]. Đây chính là khoảng cách ngữ nghĩa giữa ngữ nghĩa cấp cao và các đặc trưng thị giác cấp thấp của hình ảnh, một trong những thách thức lớn trong các hệ truy xuất hình ảnh dựa trên nội dung [5]. Do đó, việc phân tích và tìm kiếm ngữ nghĩa hình ảnh là một trong bài toán được quan tâm trong lĩnh vực thị giác máy tính [6–8].

Trong bài báo này, chúng tôi đề xuất một phương pháp truy vấn ảnh theo nội dung kết hợp với ngữ nghĩa sử dụng cấu trúc lưu trữ chỉ mục không gian iR^S -Tree (Improved-Region Sphere Tree) và ontology. Các đóng góp chính của bài báo bao gồm: (1) xây dựng cấu trúc dữ liệu không gian iR^S -Tree để lưu trữ các đặc trưng cấp thấp của hình ảnh và thực hiện tìm kiếm ảnh tương tự theo nội dung dựa trên cây iR^S -Tree; (2) làm giàu khung ontology đã xây dựng trên các tập dữ liệu ảnh mới bao gồm Oxford Flowers 102 và CUB-200-2011; (3) truy vấn tập ảnh tương tự trên ontology đã được xây dựng bằng câu truy vấn SPARQL tự động, đồng thời ngữ nghĩa cấp cao hình ảnh được trích xuất từ bộ từ điển.

2 Các công trình liên quan

Trong những thập niên gần đây, nhiều công trình đã sử dụng các kỹ thuật học máy và công nghệ ontology cho bài toán truy vấn ảnh theo nội dung và theo tiếp cận ngữ nghĩa để cải thiện hiệu suất hệ thống cụ thể như sau:

Filali và cs. [9] đề xuất xây dựng từ vựng và ontology dựa trên các chú thích hình ảnh của bộ ảnh ImageCLEF. Quá trình truy xuất hình ảnh được thực hiện bằng cách tích hợp cả các đặc trưng cấp thấp và độ tương đồng về ngữ nghĩa hình ảnh. Các ontology được làm phong phú bằng các khái niệm và mối quan hệ được trích xuất từ tài nguyên từ vựng BabelNet. Tuy nhiên, hệ thống đề xuất chưa kết hợp giữa tìm kiếm hình ảnh theo đặc trưng cấp thấp với khái niệm ngữ nghĩa và truy vấn ngữ nghĩa của hình ảnh; do đó, hệ thống chưa được đánh giá và so sánh với các công trình nghiên cứu khác. Ruan và cs. [10] giới thiệu một khung để truy xuất hình ảnh theo ngữ nghĩa dựa trên ontology phụ thuộc miền. Manzoor và cs. [11] sử dụng ontology miền để truy xuất hình ảnh có liên quan đến truy vấn của người dùng. Người dùng có thể đưa ra khái niệm dưới dạng nhập văn bản hoặc có thể tự nhập hình ảnh. Gowri và Nagara [12] trình bày một hệ thống truy xuất hình ảnh được lập chỉ mục ngữ nghĩa sử dụng các đặc trưng cấp thấp của hình ảnh và ngôn ngữ ontology web. Liu và cs. [13] giới thiệu một phương pháp học ngữ nghĩa dựa trên ontology liên kết các phân loại hình ảnh với các đối tượng trong ảnh. Sulaiman và cs. [14] trình bày một khung ontology đa phương thức để mô tả ngữ nghĩa hình ảnh. Nhìn chung, các phương pháp tiếp cận tập trung vào việc cung cấp cơ chế để lập bản đồ các đặc trưng cấp thấp với các khái niệm ngữ nghĩa cấp cao để giảm khoảng cách ngữ nghĩa. Tuy nhiên, vấn đề này mới chỉ được giải quyết một phần vì các phương pháp đề xuất phụ thuộc vào ontology miền hợp lý, các dữ liệu lớn và các ngôn ngữ truy vấn. Các công trình nghiên cứu liên quan cho thấy một số hệ thống truy xuất hình ảnh dựa trên chú thích bằng cách sử dụng ontology; đồng thời, ngôn ngữ truy vấn bộ ba SPARQL được sử dụng để thực hiện truy vấn.

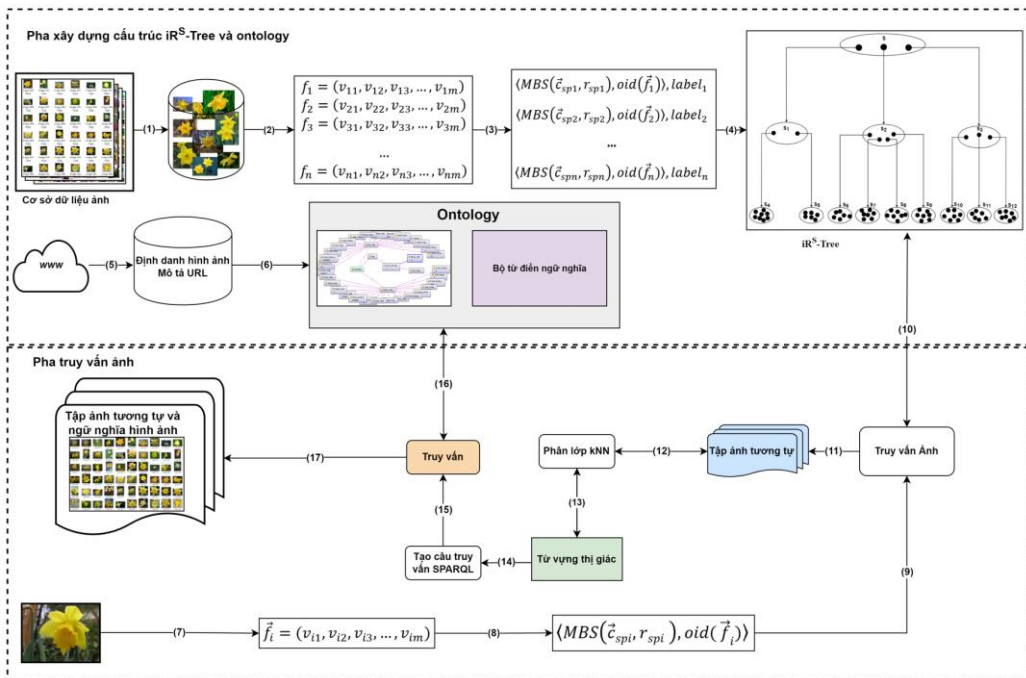
Bên cạnh đó, nhiều công trình nghiên cứu đã áp dụng các kỹ thuật trích xuất đặc trưng truyền thống và các kỹ thuật học sâu áp dụng cho các hệ truy vấn ảnh. Ahmed và cs. [15] đã trích xuất các đặc trưng màu từ ảnh RGB và sử dụng hình ảnh mức xám cho các đặc trưng cục bộ. Tuy nhiên, trong nghiên cứu này, các tác giả chỉ thực hiện lập chỉ mục trên cơ sở chủ quan và chưa đề xuất cấu trúc lập chỉ mục dữ liệu để cải thiện hiệu suất truy xuất. Nhóm tác giả đã thực nghiệm truy vấn top 10 hình ảnh với độ chính xác của giá trị 87,1% (COREL) và 98,7% (Oxford Flowers 102). Tuy nhiên, với top 40 hình ảnh, độ chính xác là 72,1% (COREL) và 71,4% (Oxford Flowers 102). Zeng và cs. [16] đã phát triển một phương pháp băm để khoanh vùng các đối tượng phân biệt chi tiết và tạo ra các mã nhị phân hiệu quả. Nhóm tác giả đã thiết kế một kiến trúc mới, thực hiện đồng thời việc huấn luyện các vùng cục bộ và tạo mã băm băm. Hai nhiệm vụ này được hiệu chỉnh và có thể bổ sung cho nhau. Bên cạnh đó, các vùng đa tỉ lệ được học bằng cách sử

dụng kích hoạt nhiều đặc trưng. Kết quả thực nghiệm truy vấn trên bộ dữ liệu ảnh CUB-200-2011 với độ chính xác 69,07% (ảnh 16 bits) và 70,1% (ảnh 64 bits). Các đánh giá thực nghiệm trên một số bộ dữ liệu ảnh chi tiết cho thấy rằng phương pháp đề xuất đạt được hiệu suất tốt. Tuy nhiên, phương pháp này tốn nhiều thời gian để huấn luyện. Đồng thời, công trình này chưa đề xuất một cấu trúc dữ liệu để lưu trữ hình ảnh nhằm cải thiện thời gian tìm kiếm ảnh.

3 Hệ tìm kiếm ảnh theo tiếp cận ngữ nghĩa dựa trên iR^S -Tree và ontology

3.1 Mô hình kiến trúc hệ thống

Trong phần này, chúng tôi đề xuất một mô hình hệ tìm kiếm ảnh theo tiếp cận ngữ nghĩa (Hình 1) và đặt tên là SBIR_IRST. Trong hệ SBIR_IRST, mỗi hình ảnh được trích xuất véc-tơ đặc trưng dựa trên các đặc trưng thị giác cấp thấp như màu sắc, vị trí, hình dạng và kết cấu. Mỗi véc-tơ đặc trưng này được liên kết với một phân lớp khái niệm hình ảnh để mô tả ngữ nghĩa từ vựng thị giác. Sau đó, các véc-tơ đặc trưng này được chuyển đổi thành dạng khối cầu và được lưu trữ trên cây phân cụm cân bằng iR^S -Tree. Bên cạnh đó, một ontology được xây dựng để mô tả các khái niệm ngữ nghĩa của hình ảnh. Một từ điển đồng nghĩa bao gồm các khái niệm chung cho các từ vựng lớp và thuộc tính để mở rộng ngữ nghĩa cho ontology được xây dựng dựa vào WORDNET.



Hình 1. Mô hình truy vấn ảnh theo ngữ nghĩa SBIR_IRST

Hệ tìm kiếm bao gồm hai pha: (1) xây dựng cây phân cụm phân cấp dựa trên cấu trúc R-Tree, thực hiện trích xuất các đặc trưng của hình ảnh trong tập dữ liệu, phân đoạn hình ảnh để tạo các phân lớp khái niệm và thực hiện gom cụm dữ liệu dựa trên độ đo tương đồng Euclidean; tạo một ontology cho hình ảnh dựa trên phần mềm Protégé và tạo một cấu trúc từ điển nhằm mô tả ngữ nghĩa cho các đối tượng hình ảnh; (2) pha truy vấn ảnh thực hiện hai quá trình truy vấn: truy vấn trên cụm dữ liệu đã phân hoạch để tìm tập ảnh tương tự, từ vựng thị giác và chú thích của từng hình ảnh trong tập ảnh; truy vấn trên ontology dựa vào từ vựng thị giác để tạo ra tập ảnh tương tự theo ngữ nghĩa; đồng thời, ngữ nghĩa cấp cao của hình ảnh được truy xuất từ bộ từ điển được xây dựng.

Pha xây dựng cấu trúc phân cụm và ontology gồm năm bước như sau:

Bước 1. Trích xuất các véc-tơ đặc trưng cấp thấp và phân lớp của hình ảnh (1-2);

Bước 2. Chuyển đổi các véc-tơ đặc trưng ảnh thành dạng khối cầu (3);

Bước 3. Gom cụm dữ liệu theo phương pháp phân cụm phân cấp dựa trên cấu trúc iRS-Tree (4);

Bước 4. Chọn bộ dữ liệu ảnh từ WWW, rút trích URI và các mô tả, chú thích; tạo các dữ liệu mẫu các lớp, thuộc tính, cá thể, mô tả ngữ nghĩa cho hình ảnh (5);

Bước 5. Làm giàu ontology và tạo cấu trúc từ điển ngữ nghĩa các tập dữ liệu ảnh (6);

Pha truy vấn ảnh: bao gồm hai giai đoạn chính.

Giai đoạn 1: Truy vấn hình ảnh dựa trên cấu trúc phân cụm dữ liệu iRS-Tree

Bước 1. Trích xuất các véc-tơ đặc trưng của ảnh truy vấn và chuyển đổi thành dạng khối cầu (7-8);

Bước 2. Truy vấn tập ảnh tương tự trên cấu trúc iRS-Tree đã xây dựng (9-10-11).

Bước 3. Phân lớp ảnh truy vấn bằng phương pháp kNN (12-13)

Giai đoạn 2: Truy vấn ngữ nghĩa dựa vào ontology và từ điển

Bước 1. Tạo câu truy vấn SPARQL từ các lớp của véc-tơ từ thị giác của ảnh truy vấn (14);

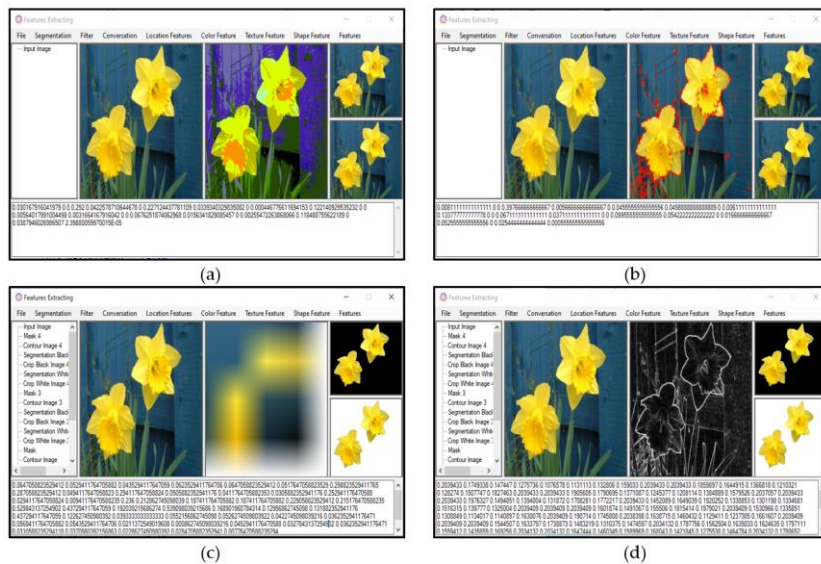
Bước 2. Truy vấn ảnh tương tự dựa trên ontology và trích xuất ngữ nghĩa dựa trên từ điển (15-16);

Bước 3. Trả về tập các hình ảnh tương tự và các mô tả ngữ nghĩa của hình ảnh (17).

3.2 Các thành phần trong hệ thống

Trích xuất đặc trưng hình ảnh

Kết quả trích xuất đặc trưng của hình ảnh có ảnh hưởng lớn đến hiệu suất hệ thống truy vấn ảnh bởi vì các đặc trưng này được sử dụng để tính toán độ đo tương đồng giữa các hình ảnh trong tập cơ sở dữ liệu. Việc sử dụng một loại đặc trưng duy nhất để thực hiện truy vấn sẽ làm giảm hiệu suất của hệ thống. Do đó, nhiều công trình đã thực hiện kết hợp các loại đặc trưng khác nhau nhằm nâng cao độ chính xác của tập ảnh tương tự [17]. Các đặc trưng cấp thấp được sử dụng phổ biến trong các hệ truy vấn ảnh theo nội dung bao gồm: màu sắc, kết cấu, vị trí và hình dạng [18]. Trong bài báo này, chúng tôi sử dụng kết hợp các đặc trưng bao gồm: đặc trưng màu dựa trên hệ màu MPEG7 (Hình 2a); đặc trưng vị trí theo thuật toán ShiTomasi (Hình 2b); đặc trưng hình dạng dựa trên phép toán Maxpooling và bộ dò cạnh Sobel (Hình 2c); đặc trưng kết cấu dựa trên HOG và phép lọc Sobel (Hình 2d). Số đặc trưng của hình ảnh được trích xuất là 242 chiều.



Hình 2. Minh họa trích xuất đặc trưng hình ảnh

Xây dựng ontology

Trong phần này, một ontology miền cho bộ ảnh COREL, OxfordFlower-120 và CUB-200-2011 được xây dựng bán tự động dựa trên ngôn ngữ bộ ba RDF để mô tả chú thích ngữ nghĩa của các hình ảnh. Trên cơ sở đó, quá trình truy vấn ảnh tương tự theo ngữ nghĩa được thực hiện bằng ngôn ngữ SPARQL trên ontology đã xây dựng. Ontology được kế thừa từ một khung ontology nền tảng của bộ ảnh ImageCLEF [19] và được thực hiện bổ sung theo các bước như sau:

Bước 1. Chuẩn bị các mẫu dữ liệu cho các bộ ảnh bao gồm: các phân lớp, thực hiện phân cấp cây phân lớp, các thuộc tính của phân lớp, các thuộc tính của hình ảnh, các định nghĩa của các phân lớp.

Bước 2. Bổ sung tự động các cá thể và các mô tả dữ liệu ảnh vào ontology nền tảng đã xây dựng:

- Xác định các phân lớp trong khung ontology [19] và đối sánh với các phân lớp của tập ảnh muốn bổ sung, nếu trùng thì không cần bổ sung phân lớp; nếu là phân lớp mới thì xác định thứ bậc để bổ sung: nếu là bậc con thì thêm vào phân lớp cha có sẵn, hoặc thêm vào thành một lớp cha mới;
- Bổ sung các định nghĩa cho các phân lớp mới;
- Các cá thể khi được thêm vào được phân lớp và tự động bổ sung vào theo từng phân lớp trên ontology;
- Bổ sung các thuộc tính của phân lớp mới và các thuộc tính cho hình ảnh.

Một kết quả bổ sung dữ liệu cho ontology với bộ ảnh Oxford Flowers được thực hiện trên phần mềm protégé được minh họa trên Hình 3–5. Lúc này, các phân lớp ảnh mới trong bộ OxfordFlower-102 là phân lớp con của phân lớp Flower có sẵn trong khung ontology.

<p>WINDFLOWER URI: http://sbir-hcm.vn/WINDFLOWER Superclasses: WINDFLOWER SubClassOf FLOWER Annotations: anoFilename "../Dictionary/WINDFLOWER.txt" anoURI "http://sbir-hcm.vn/WINDFLOWER" anoDescription "any woodland plant of the genus Anemone grown for its beautiful flowers and whorls of dissected leaves"</p>	<p>IMGimage_1218 URI: http://sbir-hcm.vn/IMGimage_1218 Object property assertions: IMGimage_1218 opWINDFLOWER inWINDFLOWER Annotations: imgName "image_1218.jpg" imgListclass "Windflower" imgFilename "../_ImageDBs/OxfordFlowers17/Annotations/Windflower/image_1218.xml" imgPath "../_ImageDBs/OxfordFlowers17/Images/Windflower/image_1218.jpg" imgURI "http://sbir-hcm.vn/IMGimage_1218"</p>
---	--

Hình 3. Chú thích ảnh được thể hiện trên phần mềm Protégé

```

1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
2 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
3 @prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
4 @prefix owl: <http://www.w3.org/2002/07/owl#>.
5 @prefix xml: <http://www.w3.org/XML/1998/namespace>.
6 @prefix sbir: <http://sbir-hcm.vn/>.
7
8 [sbir:BLUEBELL sbir:anoDescription "perennial of northern hemisphere with slender stems and bell-shaped bl
9 sbir:anoFilename "../Dictionary/BLUEBELL.txt";
10 sbir:anoURI "http://sbir-hcm.vn/BLUEBELL";
11 a owl:Class;
12 rdfs:subClassOf sbir:FLOWER.
13 [sbir:BUTTERCUP sbir:anoDescription "any of various plants of the genus Ranunculus";
14 sbir:anoFilename "../Dictionary/BUTTERCUP.txt";
15 sbir:anoURI "http://sbir-hcm.vn/BUTTERCUP";
16 a owl:Class;
17 rdfs:subClassOf sbir:FLOWER.
18 [<http://sbir-hcm.vn/COLTS'FOOT> sbir:anoDescription "perennial herb with large rounded leaves resembling
19 sbir:anoFilename "../Dictionary/COLTS'FOOT.txt";
20 sbir:anoURI "http://sbir-hcm.vn/COLTS'FOOT";
21 a owl:Class;
22 rdfs:subClassOf sbir:FLOWER.
23 [sbir:COWSLIP sbir:anoDescription "early spring flower common in British isles having fragrant yellow or s
24 sbir:anoFilename "../Dictionary/COWSLIP.txt";
25 sbir:anoURI "http://sbir-hcm.vn/COWSLIP";
26 a owl:Class;

```

Hình 4. Một ví dụ ontology dưới dạng ngôn ngữ RDF

Một ontology cho hình ảnh được tạo thành công trên Protégé có thể được lưu trữ dưới dạng tệp RDF. Các hình ảnh được chú thích với các mô tả ngữ nghĩa lưu trữ cùng với tập dữ liệu. Để thực hiện truy vấn ngữ nghĩa hình ảnh trên ontology được tạo và câu lệnh SPARQL được sử dụng.

```

Tư vựng thị giác Buttercup Dandelion
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX sbir: <http://sbir-hcm.vn/>
SELECT DISTINCT ?imgName
WHERE{ (?IMG sbir:imgName ?imgName. ?IMG sbir:opBUTTERCUP sbirinBUTTERCUP . sbirinBUTTERCUP rdf:type owl:NamedIndividual . sbirinBUTTERCUP rdf:type sbir:BUTTERCUP .) UNION (?IMG sbir:imgName ?imgName. ?IMG sbir:opDANDELION sbirinDANDELION . sbirinDANDELION rdf:type owl:NamedIndividual . sbirinDANDELION rdf:type sbir:DANDELION .)}
    
```

Hình 5. Minh họa câu lệnh SPAQRL được tạo tự động

Xây dựng bộ từ điển ngữ nghĩa cho hình ảnh

Một từ điển đồng nghĩa bao gồm các khái niệm chung cho các từ vựng lớp và thuộc tính để mở rộng ngữ nghĩa cho ontology. Quy trình xây dựng bộ từ điển ngữ nghĩa cho hình ảnh được kế thừa từ công trình [20]. Quy trình xây dựng từ điển gồm ba bước: (1) với mỗi ảnh của bộ dữ liệu ảnh, thực hiện lấy phân lớp hình ảnh để tạo tập từ vựng mục tiêu; (2) thực hiện kết xuất tự động các khái niệm là danh từ và đồng nghĩa với từ vựng mục tiêu dựa vào WORDNET theo cấu trúc: danh từ, từ khóa, định nghĩa cho từ vựng; (3) tạo bộ dữ liệu từ điển lưu trữ các khái niệm ngữ nghĩa cấp cao của hình ảnh.

Mỗi khái niệm được lưu bằng tập tin chứa tên các lớp hay thuộc tính (Hình 6), ví dụ về từ vựng "COWSLIP" được định nghĩa.

```

COWSLIP - Notepad
File Edit Format View Help
(n) cowslip, paigle, Primula veris (early spring flower common in British isles having fragrant yellow or sometimes purple flowers)
(n) marsh marigold, kingcup, meadow bright, May blob, cowslip, water dragon, Caltha palustris (swamp plant of Europe and North America having bright yellow flowers resembling buttercups)
Ln 5, Col 32 100% Windows (CRLF) UTF-8
    
```

Hình 6. Một ví dụ về khái niệm từ vựng trong từ điển ontology

Trên Hình 6, từ vựng mục tiêu "COWSLIP" được trích xuất tự động trên WORDNET để lấy các định nghĩa và các từ đồng nghĩa: là các danh từ (n); từ khóa: "cowslip", "paigle" ...; khái niệm; các ví dụ để làm rõ ngữ nghĩa theo ngữ cảnh của từ vựng.

Mô tả cấu trúc cây phân cụm iRS-Tree

R^S -Tree [21] được xây dựng dựa trên cấu trúc SS-Tree [22]. Cách tính véc-tơ tâm của khối cầu nút lá và nút trong trên R^S -Tree tương tự với SS-Tree. Các cải tiến trên R^S -Tree bao gồm: (1) tạo một khối cầu không gian để lưu trữ véc-tơ đặc trưng của hình ảnh; (2) cải tiến thuật toán tách nút dựa trên một độ đo sai biệt giữa các phần tử; (3) đề xuất một ngưỡng \otimes để gom cụm các dữ liệu tương tự; (4) kết hợp truy vấn láng giềng gần nhất và truy vấn vùng không gian.

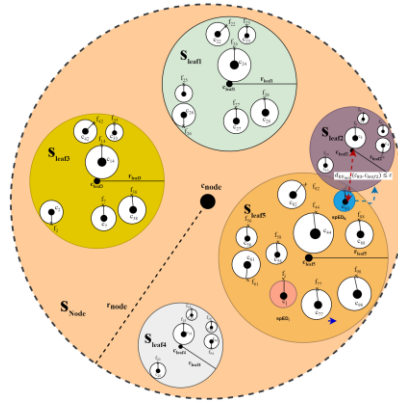
R^S -Tree là cây đa nhánh cân bằng ứng dụng cho bài toán tìm kiếm ảnh tương tự. Việc gom nhóm dữ liệu được thực hiện trên từng nút của cây R^S -Tree dựa vào độ đo tương tự giữa các véc-tơ đặc trưng ảnh ngưỡng θ cho trước nhằm tạo ra một cây đa nhánh cân bằng để giảm thời gian tìm kiếm. Cây R^S -Tree là cây phân hoạch dữ liệu không gian bao gồm: một nút gốc, một tập nút trong và một tập nút lá. Mỗi nút trong S_{node} trên cây là một khối cầu MBS (Minimum Bounding Sphere) bao phủ tất cả khối cầu các nút thuộc nhánh cây con. Mỗi nút lá S_{leaf} trên cây gồm một tập các thực thể; mỗi thực thể $spED$ là một khối cầu chứa không gian đối tượng dữ liệu và định danh đối tượng.

Việc tách nút thường xuyên sẽ làm chậm quá trình tạo cây. Do đó, để nâng cao hiệu quả gom cụm và cải thiện thời gian tạo cây, chúng tôi xây dựng cấu trúc iR^S -Tree, được cải tiến từ R^S -Tree. Các cải tiến trong cấu trúc iR^S -Tree bao gồm:

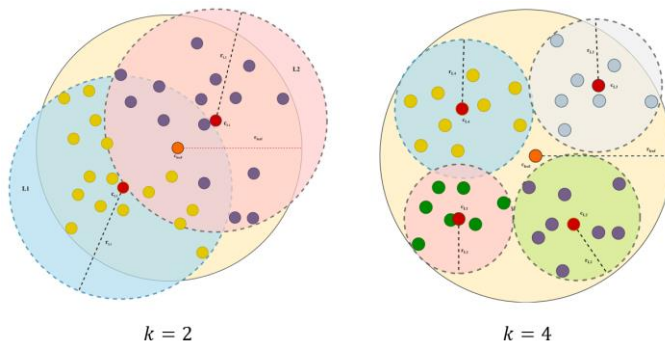
Thứ nhất, việc thêm một phần tử $spED_i$ vào iR^S -Tree được thực hiện theo các nguyên tắc tương tự R^S -Tree. Tuy nhiên, khi gặp một nút lá đầy, thay vì thực hiện tách nút, một phần tử xa tâm nhất của nút lá hiện hành nhất, ký hiệu là $spED_k$, được chọn để thêm vào một nút lá khác có cùng cha. Nút lá được chọn để thêm vào thỏa mãn hai tiêu chí sau: (1) nút lá được chọn có số phần tử nhỏ hơn M ; (2) khoảng cách Euclidean từ tâm của phần tử $spED_k$ đến tâm của phần tử nút lá được chọn nhỏ hơn hoặc bằng \otimes . Trường hợp có nhiều nút lá thỏa mãn hai tiêu chí trên nút lá có khoảng cách Euclidean ngắn nhất sẽ được chọn. Ngược lại, nếu không tồn tại nút lá nào thỏa mãn hai tiêu chí này thì thực hiện tách nút. Quá trình thêm phần tử iR^S -Tree vào được minh họa trên Hình 7.

Hình 7 mô tả quá trình thêm phần tử $spED_i$ vào cây iR^S -Tree. S_{leaf5} là nút lá được lựa chọn để thêm phần tử $spED_i$. Trong trường hợp nút lá S_{leaf5} đầy, phần tử $spED_k$ là phần tử xa nhất được chọn để thêm vào nút lá gần nhất S_{leaf2} .

Thứ hai, khi một nút lá bị tràn, phương pháp k-Means được sử dụng để thực hiện tách nút thành k cụm ($k \geq 2$). Việc này nhằm tối ưu hóa độ tương tự của các đối tượng không gian trong một cụm, đồng thời giảm thiểu độ tương tự giữa các cụm khác nhau. Các phần tử trong nút tràn sẽ được phân hoạch thành k cụm theo độ đo tương đồng (Hình 8).



Hình 7. Minh họa việc thêm một phần tử vào cây



Hình 8. Phép tách nút trong iR^S -Tree bằng phương pháp k-Means

Cải tiến phép toán thêm phần tử vào cây iRS -Tree

Khi thêm phần tử $spED$ vào cây iR^S -Tree, thuật toán chèn phần tử vào cây được thực hiện theo các nguyên tắc của R^S -Tree. Tuy nhiên, khi gặp một nút tràn thay vì tách nút thì phần tử xa tâm nhất của nút tràn được lựa chọn để thực hiện thêm vào một nút phù hợp có cùng cha với nút lá hiện hành; trong trường hợp không chọn được nút lá thỏa mãn tiêu chí để thêm vào thì thực hiện tách nút. Thuật toán thêm phần tử vào cây iR^S -Tree được trình bày như sau:

Thuật toán 1. Thêm phần tử vào cây

Input: Nút S_N , ngưỡng M , ngưỡng θ và phần tử dữ liệu $spED$

Output: Cây iR^S -Tree sau khi thêm phần tử

Function: $INspEDiRST(S_N, spED)$

```

1  Begin
2  If  $S_N$  là nút lá then
3    If số phần tử của  $S_N$  nhỏ hơn  $M$  then
4      Tính khoảng cách Euclidean  $d = Euclidean(spED.\vec{c}, S_N.\vec{c}) + spED.r$ ;
5      If  $d \leq \theta$  then
6        Thêm phần tử  $spED$  vào  $S_N$ ;
7      Else
8        Tạo nút lá mới để lưu  $spED$ ;
9    Else
10     Đặt  $S = S_N \cup \{spED\}$ ;
11     Chọn phần tử  $spED_k$  thuộc  $S$  thỏa  $d = Max_{k=1..count(S)}\{d_{EU}(spED_k.\vec{c}, S_N.\vec{c})\}$ ;
12     Chọn nút lá  $S_t$  cùng cha với nút  $S_N$  thỏa điều kiện;
13      $S_k$  có số phần tử nhỏ hơn  $M$ ;
14      $d = Min_{t=1..count(S_N.parent)}\{d_{EU}(spED_k.\vec{c}, S_t.\vec{c})\}$ ;
15      $d \leq \theta$ ;
16     Thêm phần tử  $spED_k$  vào nút lá  $S_t$ ;
17     If không tồn tại nút lá thỏa điều kiện then
18       Thực hiện tách nút;
19  End

```

Gọi n là số phần tử của tập dữ liệu và M là số phần tử tối đa trong một nút của iR^S -Tree. Thuật toán INspEDiRST lần lượt thực hiện duyệt từ nút gốc đến nút lá; mỗi lần duyệt qua M phần tử; trong mỗi lần duyệt, Thuật toán INspEDiRST thực hiện phép cập nhật tâm và tách nút từ nút lá đến nút gốc. M là hằng số. Do đó, Thuật toán INspEDiRST có độ phức tạp là $O((\log n)^3)$.

Cải tiến phép toán tách nút trên cây iRS-Tree

Cây iRS-Tree là một cấu trúc cây cân bằng động phát triển theo chiều từ lá dưới lên gốc. Việc nút N_i bị tràn được xử lý bằng cách phân chia $M + 1$ phần tử vào k nút ($k \geq 2$). Thuật toán k-Means được sử dụng để thực việc phân hoạch $M + 1$ phần tử thành k nhóm. Sau đó mỗi nhóm sẽ được gán vào mỗi nút lá trên cây iRS-Tree có cùng cha với nút lá được tách. Khi quá trình phân tách xảy ra lan truyền đến nút trong thì thực hiện tách nút trong. Quá trình này có thể lan truyền đến nút gốc; khi nút gốc đầy thì một nút gốc mới sẽ được tạo ra và cây iRS-Tree phát triển thêm một bậc. Việc lựa chọn thuật toán k-Means vì ba lý do sau: (1) hiệu quả về thời gian thực thi; (2) độ chính xác cao; (3) không phụ thuộc vào thứ tự của các đối tượng. Thuật toán tách nút lá được trình bày như sau:

Thuật toán 2. Tách nút trên cây

Input: Nút lá cần tách S_L

Output: k nút sau khi được tách ($k \geq 2$)

Function: **KMSplitLeafiRST**(S_L)

- 1 **Begin**
- 2 Phân chia tập $M + 1$ phần tử thành k nút sử dụng phương pháp K-Means;
- 3 Khởi tạo số cụm với $k = 2$;
- 4 Áp dụng k-Means trên tập $M+1$ phần tử để phân chia thành k cụm;
- 5 Tính toán vùng diện tích chồng lấp giữa k vùng không gian;
- 6 $Max = \text{OverlapAera}(k)$;
- 7 $k_{opt} = k$;
- 8 For $k = 3$ to k_{max}
- 9 Áp dụng k-Means trên tập $M+1$ phần tử để phân chia thành k cụm;
- 10 Tính toán $\text{OverlapAera}(k)$;
- 11 If $\text{OverlapAera}(k) < Max$ then
- 12 $Max = \text{OverlapAera}(k)$;
- 13 $k_{opt} = k$;

- 14 For $k = 1$ to k_{opt}
- 14 Gán các phần tử của k cụm vào k nút.
- 15 **End**

Gọi n là số phần tử của tập dữ liệu và M là số phần tử tối đa trong một nút iR^S -Tree. Khi thực hiện tách nút, trong trường hợp xấu nhất, Thuật toán $KMSplitLeafiRST$ phải tách từ nút lá đến nút gốc. Trong mỗi lần tách nút, Thuật toán $KMSplitLeafiRST$ phải thực hiện M phép so sánh để phân bố về k -cụm. Mặt khác, trong mỗi lần thực hiện tách nút và trong trường hợp xấu nhất, Thuật toán $KMSplitLeafiRST$ phải gọi đệ quy từ nút lá đến nút gốc. M là hằng số. Do đó, độ phức tạp của Thuật toán $KMSplitLeafiRST$ là $O((\log n)^2)$.

3.3 Thuật toán tìm kiếm ảnh tương tự theo tiếp cận ngữ nghĩa

Thuật toán tìm kiếm ảnh theo nội dung dựa trên cấu trúc iRS -Tree

Từ cây phân cụm dữ liệu iR^S -Tree đã xây dựng, một thuật toán tra cứu ảnh tương tự theo nội dung dựa trên cây iR^S -Tree được đề xuất. Quá trình tìm kiếm ảnh tương tự được thực hiện trên cây iR^S -Tree và được mô tả như sau:

Thuật toán 3. Truy vấn ảnh trên cây iR^S -Tree

Input: vec-tơ đặc trưng $spED$ của ảnh truy vấn I_Q , cây iR^S -Tree.

Output: tập ảnh tương tự SI , tập từ vựng thị giác của hình ảnh Y_{vwv}

Function: $iRSTIR(S_{Nr}, spED)$

- 1 **Begin**
- 2 **if** S_{Nr} không tồn tại **then**
- 3 trả về giá trị *null*;
- 4 **else**
- 5 **if** S_{Nr} không phải là nút lá **then**
- 6 Chọn nhánh S_{Nk} có độ đo tương tự Euclidean ngắn nhất;
- 7 *iRSTIR* (S_{Nk}, S_{spED});
- 8 **else**

- 9 Trả về tập ảnh tương tự SI của ảnh I_Q ;
- 10 Thực hiện phân lớp KNN để tạo từ vựng thị giác Y_{vwv} cho ảnh I_Q ;
- 11 Trả về kết quả $\{SI, Y_{vwv}\}$;
- 12 **End**

Gọi n là số phần tử của tập dữ liệu và M là số phần tử tối đa trong một nút của iR^S -Tree. Thuật toán $iRSTIR$ lần lượt duyệt qua các nút từ gốc đến lá. Hơn nữa, vì cây iR^S -Tree là cây cân bằng nên thuật toán $iRSTIR$ duyệt qua chiều cao h của cây. Mỗi lần duyệt, thuật toán $iRSTIR$ phải so sánh với M phần tử của mỗi nút. M là hằng số. Do đó, độ phức tạp của thuật toán $iRSTIR$ là $O(\log n)$.

Thuật toán tra cứu ảnh theo tiếp cận ngữ nghĩa dựa trên ontology

Với mỗi ảnh đầu vào, tập từ vựng thị giác được trích xuất bằng phương pháp KNN dựa trên tập ảnh tương tự theo nội dung. Dựa vào tập từ vựng thị giác, hệ thống tự động tạo ra câu lệnh SPARQL và thực hiện truy vấn trên ontology đã xây dựng. Kết quả truy vấn là một tập các URI cùng ngữ nghĩa hình ảnh và metadata của bộ dữ liệu ảnh tương tự. Thuật toán trích xuất ngữ nghĩa hình ảnh được trình bày như sau:

Thuật toán. Trích xuất ngữ nghĩa hình ảnh

Input: Tập từ vựng thị giác Y_{vwv} của ảnh I

Output: Tập ảnh tương tự \mathbb{C}_{SI} và các mô tả ngữ nghĩa \mathcal{M}_{SI}

Function: OnSBIR(string [] Y_{vwv})

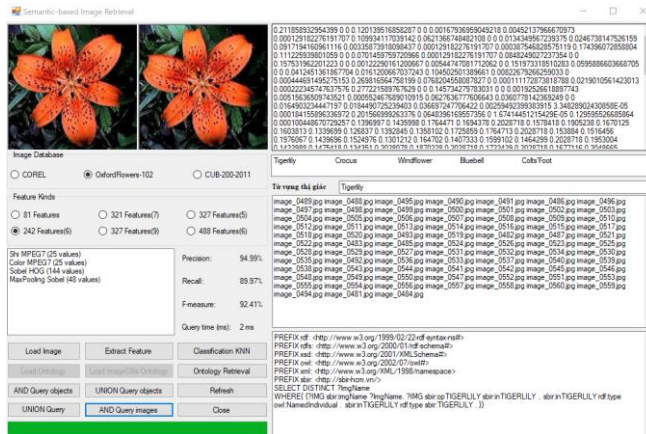
- 1 **Begin**
- 2 Khởi tạo tập ảnh tương tự $\mathbb{C}_{SI} = \emptyset$;
- 3 Khởi tạo tập mô tả ngữ nghĩa $\mathcal{M}_{SI} = \emptyset$;
- 4 $imgSPAQL = CreateSPAQL(Y_i)$;
- 5 $\{\mathbb{C}_{SI}, \mathcal{M}_{SI}\} = ExecuteQuery(imgSPAQL, Ontology)$;
- 6 Trả về tập ảnh tương tự và các mô tả ngữ nghĩa của ảnh I $\{\mathbb{C}_{SI}, \mathcal{M}_{SI}\}$;
- 7 **End**

4 Kết quả thực nghiệm

4.1 Môi trường và ứng dụng thực nghiệm

Pha tiền xử lý được thực hiện trên máy PC CPU 2,3 GHz 8-core 9th-generation Intel Core i9, 16 GB, 2666 MHz memory, 1 TB flash storage. Pha tìm kiếm được thực nghiệm trên máy PC CPU Intel Core i7-6500U CPU @ 2,50 GHz, 8 GB RAM, hệ điều hành Windows 10 Pro 64 bit.

Trong bài báo này, chúng tôi tiến hành thực nghiệm trên ba bộ ảnh COREL 1k, Oxford Flowers 102 và CUB-200-2011 với các chủ đề khác nhau. Bộ ảnh COREL 1k có 1000 ảnh bao gồm 10 phân lớp. Bộ ảnh Oxford Flowers 102 có 8189 ảnh được chia thành 102 phân lớp. Bộ ảnh CUB-200-2011 gồm 11788 ảnh được chia thành 200 phân lớp. Ứng dụng thực nghiệm được minh họa trên Hình 9–11.



Hình 9. Giao diện tạo ontology và truy vấn SPARQL



Hình 10. Một kết quả truy vấn dựa trên ontology trên bộ ảnh OxfordFlowers-102



Hình 11. Trích xuất chú thích ảnh dựa trên từ điển

4.2 Kết quả thực nghiệm

Trong bài báo này, các tham số M, m, N, θ được tiến hành thực nghiệm để lựa chọn giá trị nhằm nâng cao độ chính xác truy vấn. Gọi $nMax_{inclass}$ là số phần tử tối đa trong một phân lớp. Chúng tôi thực nghiệm chọn $M \in [nMax_{inclass} - \alpha, nMax_{inclass} + \alpha]$, α là hằng số; $N \in [2; M]$. Ngưỡng θ được sử dụng để đánh giá độ tương tự các phần tử thuộc một cụm. Số phần tử của mỗi phân lớp xấp xỉ 10% của bộ dữ liệu. Do đó, chúng tôi thực nghiệm giá trị $\theta \in [0.1 - \mu, 0.1 + \mu]$, $\mu = 0,005$ là hệ số học. Thông qua quá trình thực nghiệm, các tham số tối ưu được lựa chọn sao cho hệ thống đạt được độ chính xác tốt nhất được trình bày trong Bảng 1.

Bảng 1. Mô tả các tham số thực nghiệm xây dựng cây iRS-Tree

Tham số	COREL	Oxford Flowers 102	CUB-200-2011
M	120	80	70
m	1	1	1
N	20	25	30
θ	0,074	0,098	0,085
<i>Topk</i>	90	50	45
Thời gian xây dựng iRS -Tree (giờ)	0,39	2,55	3,15

Việc lựa chọn tham số trong bài báo phụ thuộc vào tập dữ liệu bao gồm: (1) số các phần tử dữ liệu mẫu trong một phân lớp; (2) số lượng phần tử trong tập dữ liệu ảnh. Số phần tử trong một phân lớp càng lớn thì tham số M càng lớn. Số lượng phần tử trong tập dữ liệu lớn thì N càng lớn nhằm hạn chế chiều cao của cây giúp giảm thời gian tìm kiếm. Ngưỡng θ phụ thuộc vào độ

tương tự của các phần tử trong một phân lớp, nếu các phần tử trong một phân lớp càng gần nhau thì ngưỡng θ càng bé. $Topk$ càng lớn độ chính xác càng thấp, độ phủ càng cao.

Để đánh giá hiệu quả của phương pháp tìm kiếm ảnh, trong phần thực nghiệm chúng tôi đánh giá các giá trị gồm: độ chính xác (precision), độ phủ (recall) và độ đo dung hòa F-measure.

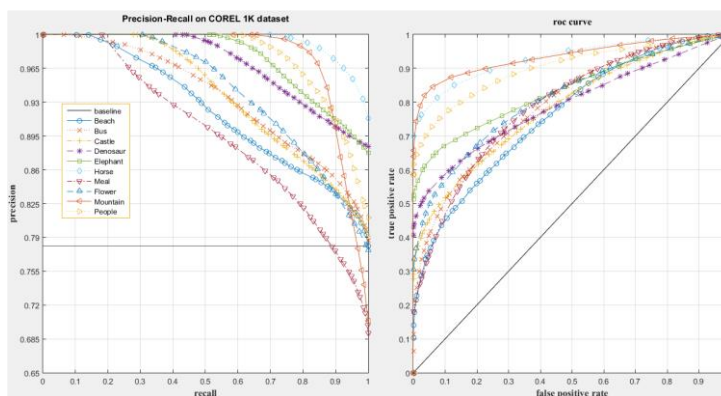
Ngoài ra, để đánh giá độ chính xác của phương pháp, chúng tôi sử dụng độ chính xác trung bình MAP (Mean Average Precision). MAP được tính theo công thức như sau:

$$MAP = \frac{\sum_{i=1}^n P_i}{n}$$

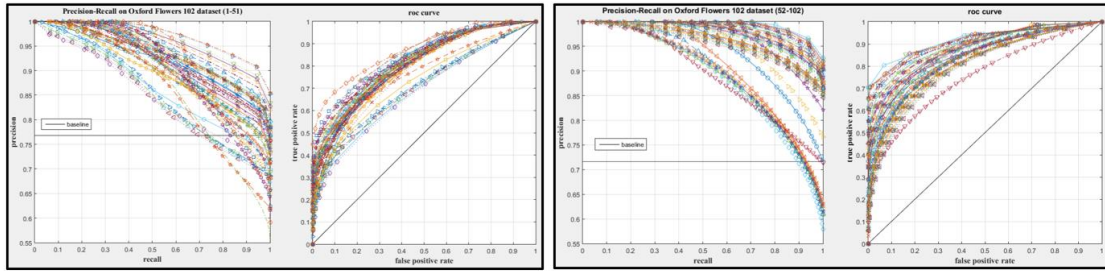
trong đó P_i là độ chính xác của mỗi truy vấn và n là số lượng truy vấn được thực hiện.

Kết quả thực nghiệm được trình bày trên Hình 12–14. Mỗi đường cong trên đồ thị mô tả kết quả truy vấn từ một chủ đề ảnh trong bộ dữ liệu COREL, OxfordFlower-102 và CUB-200-2011; mỗi điểm trên đường cong là một hình ảnh theo từng chủ đề. Dựa vào các hình ảnh có thể thấy diện tích AUC dưới đường cong Precision-Recall của tập ảnh COREL là lớn nhất (Hình 12), có nghĩa là độ chính xác cao nhất. Trong khi đó, AUC của bộ ảnh CUB-200-2011 nhỏ hơn; các đường cong gần đường baseline, nên độ chính xác cũng thấp hơn. Điều này cho thấy khi số lượng lớp ảnh lớn, việc tìm kiếm chính xác hình ảnh cũng khó khăn hơn.

Đồng thời, đường cong tương ứng trong đồ thị ROC cho biết tỷ lệ kết quả truy vấn đúng và sai, nghĩa là diện tích dưới đường cong này đánh giá tính đúng đắn của các kết quả truy vấn. Diện tích AUC dưới đường cong của đồ thị ROC của các tập ảnh đều nằm trên đường baseline, cho thấy kết quả phân loại trong bài báo của chúng tôi là đúng. Trong đó, các điểm trên đường cong của tập ảnh COREL nằm gần điểm gốc (0,1) bên trái của đồ thị hơn các đồ thị khác, cho thấy kết quả phân loại của COREL là tốt nhất. Do đó, với các tập ảnh có số lượng ảnh, số lượng lớp ảnh ít thì việc tìm kiếm và phân loại sẽ dễ dàng và tốt hơn các tập ảnh khác.



Hình 12. Precision-Recall và đường cong ROC của bộ dữ liệu COREL

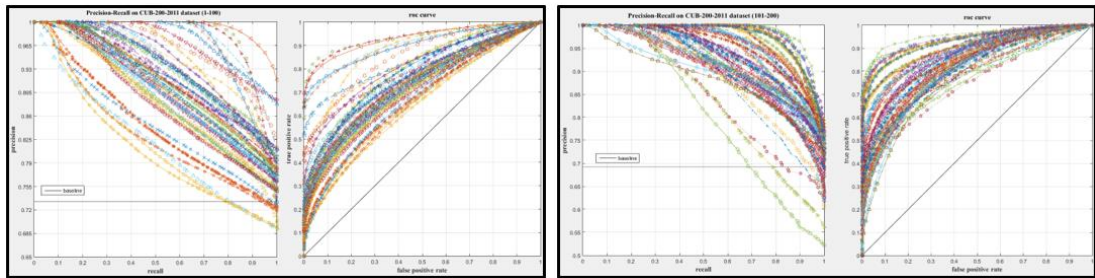


(a)

(b)

Hình 13. Precision-Recall và đường cong ROC của bộ dữ liệu OxfordFlowers-102

a) Chủ đề 1–51, (b) Chủ đề 52–102



(a)

(b)

Hình 14. Precision-Recall và đường cong ROC của bộ dữ liệu CUB-200-2011

(a) Chủ đề 1–100, (b) Chủ đề 101–200

Hiệu suất truy vấn của thuật toán tìm kiếm ảnh tương tự dựa trên cấu trúc iR^S -Tree được trình bày trong Bảng 2. Để đánh giá hiệu suất của mô hình đề xuất, chúng tôi so sánh với kết quả của các công trình liên quan trước đó trên cùng tập dữ liệu (Bảng 3).

Bảng 2. Hiệu suất tìm kiếm của hệ truy vấn SBIR- iR^S T trên các bộ dữ liệu

Bộ dữ liệu thực nghiệm	Avg. precision	Avg. recall	Avg. F-measure	Avg.query time (ms)
COREL	79,45	70,19	74,53	15,23
OxfordFlowers-102	73,16	65,36	69,04	49,27
CUB-200-2011	68,17	61,55	64,69	74,16

Bảng 3. So sánh độ chính xác giữa các phương pháp trên các bộ dữ liệu

Phương pháp	Bộ dữ liệu	Mean Average Precision (MAP), %
Ahmed 2019 [15]	COREL	72,1
Thanh 2021[21]	COREL	76,53
Phương pháp đề xuất (SBIR-iRST)	COREL	79,45
Ahmed 2019 [15]	Oxford Flowers 102	71,4
Phương pháp đề xuất (SBIR-iRST)	Oxford Flowers 102	73,16
Zeng 2019 [16]	CUB-200-2011	70,1
Phương pháp đề xuất (SBIR-iRST)	CUB-200-2011	68,17

Các kết quả thực nghiệm so sánh cho thấy phương pháp đề xuất SBIR-iRST của chúng tôi là hiệu quả trong việc giải quyết bài toán tìm kiếm ảnh theo theo tiếp cận ngữ nghĩa. Với tập ảnh COREL và OxfordFlower-102, phương pháp đề xuất của chúng tôi có độ chính xác vượt trội hơn các phương pháp khác. Trong công trình [15], các tác giả đã trích xuất các đặc trưng màu từ ảnh RGB và sử dụng hình ảnh mức xám cho các đối tượng cục bộ. Tuy nhiên, trong nghiên cứu này, nhóm tác giả chỉ thực hiện việc lập chỉ mục trên cơ sở chủ quan và chưa đề xuất cấu trúc lập chỉ mục dữ liệu để cải thiện hiệu suất truy xuất. Trong công trình này, các tác giả đã lấy top-10 hình ảnh với độ chính xác của giá trị 87,1% (COREL) và 98,7% (Oxford Flowers 102), với top-40 hình ảnh, độ chính xác là 72,1% (COREL) và 71,4% (Oxford Flowers 102). Với tập ảnh CUB-200-2011, độ chính xác của hệ thống đề xuất chưa cao bằng. Zeng và cs. [16] sử dụng phương pháp tìm kiếm ảnh trên mạng CNN. CUB-200-2011 là tập ảnh chi tiết, phù hợp với đặc trưng trích xuất trên mạng CNN hơn, do đó, độ chính xác cũng cao hơn so với đề xuất của chúng tôi. Đây cũng là động lực cho các cải tiến về sau cho các nghiên cứu của chúng tôi. Do đó, mô hình được đề xuất có thể được cải tiến, phát triển để đưa vào các ứng dụng tìm kiếm ảnh trong thực tiễn.

5 Kết luận

Trong bài báo này, chúng tôi đã xây dựng một cấu trúc iR^S -Tree, một cải tiến của cấu trúc R^S -Tree, áp dụng cho bài toán tìm kiếm ảnh. Trong cấu trúc này, thuật toán thêm phần tử được cải tiến để cải thiện thời gian tạo cây đồng thời giúp cân bằng dữ liệu trong cây; thuật toán tách nút được cải tiến dựa trên phương pháp phân cụm k-Means nhằm nâng cao độ chính xác và hạn

chế được chồng lấp không gian dữ liệu. Bên cạnh đó, việc bổ sung dữ liệu cho khung ontology được thực hiện cho truy vấn hình ảnh theo ngữ nghĩa. Trên cơ sở đó, một mô hình tìm kiếm ảnh dựa trên cấu trúc iRS-Tree và ontology được đề xuất để thực nghiệm trên các bộ ảnh khác nhau. Kết quả thực nghiệm trên bộ ảnh COREL, Oxford Flowers-102 và CUB-200-2011 có độ chính xác lần lượt là 79,45, 73,16 và 68,17%. Kết quả thực nghiệm cho thấy tính hiệu quả so với các công trình khác trên cùng một tập dữ liệu ảnh. Điều này cho thấy phương pháp đề xuất của chúng tôi đã làm tăng đáng kể hiệu suất truy vấn ảnh theo ngữ nghĩa. Tuy nhiên, ngữ nghĩa được trích xuất chỉ dừng ở mức khái niệm, định nghĩa và mô tả cho các phân lớp của đơn đối tượng trên ảnh. Chưa trích xuất được mối quan hệ ngữ nghĩa giữa các đối tượng trong cùng một ảnh. Bên cạnh đó, khi so sánh với các kỹ thuật học sâu, phương pháp đề xuất chưa đạt được kết quả mong muốn. Đó đó, trong định hướng phát triển tiếp theo, chúng tôi tiếp tục cải tiến bằng mạng R-CNN để phân lớp các đối tượng và trích xuất mối quan hệ giữa các đối tượng để tạo đồ thị ngữ cảnh cho hình ảnh. Đồng thời, đồ thị tri thức (Knowledge Graph) được xây dựng để lưu trữ các mô tả ngữ nghĩa của các hình ảnh và mối quan hệ giữa các đối tượng trên ảnh.

Lời cảm ơn

Chúng tôi xin trân trọng cảm ơn Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế và nhóm nghiên cứu SBIR-HCM đã góp ý chuyên môn cho nghiên cứu này. Chúng tôi xin trân trọng cảm ơn Trường Đại học Bà Rịa-Vũng Tàu và Trường Đại học Sư phạm Tp. HCM đã tạo điều kiện về cơ sở vật chất giúp chúng tôi hoàn thành nghiên cứu này.

Tài liệu tham khảo

1. Huang, Z., Wang, R., Shan, S. & Chen, X. (2015). Projection metric learning on Grassmann manifold with application to video based face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 140-149).
2. Zhang, Y., Pan, P., Zheng, Y., Zhao, K., Zhang, Y., Ren, X. & Jin, R. (2018, July). Visual search at alibaba. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 993-1001).
3. Yasmin, M., Sharif, M. & Mohsin, S. (2013). Neural networks in medical imaging applications: A survey. World Applied Sciences Journal, 22(1), 85-96.
4. Li, X., Yang, J. & Ma, J. (2021). Recent developments of content-based image retrieval (CBIR). Neurocomputing, 452, 675-689.
5. Alzu'bi, A., Amira, A. & Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. Journal of Visual Communication and Image Representation, 32, 20-54.

6. Shaaban, A. M., Salem, N. M. & Al-atabany, W. I. (2020). A semantic-based scene segmentation using convolutional neural networks. *AEU-International Journal of Electronics and Communications*, 125, 153364.
7. Jia, S., Ma, L., Yang, S. & Qin, D. (2021). Semantic and Context Based Image Retrieval Method Using a Single Image Sensor for Visual Indoor Positioning. *IEEE Sensors Journal*, 21(16), 18020-18032.
8. Gonçalves, F. M. F., Guilherme, I. R. & Pedronette, D. C. G. (2018). Semantic guided interactive image retrieval for plant identification. *Expert Systems with Applications*, 91, 12-26.
9. Fadzli, S. A. & Setchi, R. (2010, September). Semantic approach to image retrieval using statistical models based on a lexical ontology. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 240-250). Springer, Berlin, Heidelberg.
10. Ruan, N., Huang, N. & Hong, W. (2006, July). Semantic-based image retrieval in remote sensing archive: An ontology approach. In *2006 IEEE International Symposium on Geoscience and Remote Sensing* (pp. 2903-2906). IEEE.
11. Manzoor, U., Balubaid, M. A., Zafar, B., Umar, H. & Khan, M. S. (2015). Semantic image retrieval: An ontology based approach. *International Journal of Advanced Research in Artificial Intelligence*, 4(4), 1-8.
12. Allampalli-Nagaraj, G. & Bichindaritz, I. (2009). Automatic semantic indexing of medical images using a web ontology language for case-based image retrieval. *Engineering Applications of Artificial Intelligence*, 22(1), 18-25.
13. Liu, Y., Huang, Y., Zhang, S., Zhang, D. & Ling, N. (2017, June). Integrating object ontology and region semantic template for crime scene investigation image retrieval. In *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)* (pp. 149-153). IEEE.
14. Sulaiman, M. S., Nordin, S. & Jamil, N. (2017). An object properties filter for multi-modality ontology semantic image retrieval. *Journal of Information and Communication Technology*, 16(1), 1-19.
15. Ahmed, K. T., Ummesafi, S. & Iqbal, A. (2019). Content based image retrieval using image features information fusion. *Information Fusion*, 51, 76-99.
16. Zeng, H., Lai, H. & Yin, J. (2019). Simultaneous region localization and hash coding for fine-grained image retrieval. *arXiv preprint arXiv:1911.08028*.
17. Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., ... & Khalil, T. (2019). Content-based image retrieval and feature extraction: a comprehensive review. *Mathematical Problems in Engineering*, 2019.
18. Belongie, S., Carson, C., Greenspan, H. & Malik, J. (1998, January). Color-and texture-based image segmentation using EM and its application to content-based image retrieval. In *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)* (pp. 675-682). IEEE.
19. Nguyen Thi Uyen Nhi, Thanh Manh Le, Thanh The Van (2022). A Model of Semantic-based Image Retrieval using C-Tree and Neighbor Graph, *International Journal on Semantic Web and Information Systems (IJSWIS)*, IGI Global, 18(1), 1-23.
20. Lê Thị Vĩnh Thanh và cs. (2020). Tìm kiếm ảnh theo ngữ nghĩa dựa trên phương pháp gom cụm và ontology. *Kỷ yếu Hội thảo Quốc gia về Nghiên cứu cơ bản và ứng dụng CNTT (FAIR), ĐH Nha Trang, Nhà xuất bản Khoa học Tự nhiên và Công nghệ*, ISBN: 978-604-9985-77-5, tr.612-622, 2020.

21. Lê Thị Vĩnh Thanh, Văn Thế Thành, Lê Mạnh Thạnh, “Một phương pháp tìm kiếm ảnh hiệu quả dựa trên cấu trúc R-Tree”, Kỷ yếu Hội thảo Quốc gia về Công nghệ thông tin và ứng dụng trong các lĩnh vực (CITA), Đại học Đà Nẵng, Nhà xuất bản Đà Nẵng, ISBN: 978-604-84-5998-7, trang 259-271, 2021.
22. White, D. A., and Jain, R.n, “Similarity indexing with the SS-tree”, In Proceedings of the Twelfth International Conference on Data Engineering, IEEE. (1996):516-523.