

TÌM KIẾM ẢNH THEO NGỮ NGHĨA DỰA TRÊN PHƯƠNG PHÁP GOM CỤM VÀ ONTOLOGY

Lê Thị Vĩnh Thanh^{1,4}, Phan Thị Ngọc Mai², Văn Thế Thành³, Lê Mạnh Thạnh⁴

¹Trường Đại học Bà Rịa - Vũng Tàu

² Khoa Công nghệ thông tin, Trường ĐH Công nghiệp Thực phẩm TP. HCM

³ Phòng Quản lý khoa học và Đào tạo Sau đại học, Trường ĐH Công nghiệp Thực phẩm TP. HCM

⁴Trường Đại học Khoa học- Đại học Huế

thanhlv@bvut.edu.vn, maipn@hufi.edu.vn, thanhvt@hufi.edu.vn, lmthanh@hueuni.edu.vn

TÓM TẮT: Truy vấn ngữ nghĩa hình ảnh là một bài toán có nhiều ứng dụng trong các lĩnh vực khác nhau như: y khoa, giáo dục, giải trí, ... Trong bài báo này, một mô hình tìm kiếm và tra cứu ngữ nghĩa hình ảnh dựa trên phương pháp gom cụm và ontology được đề xuất. Để thực hiện vấn đề này, chúng tôi gom cụm dữ liệu nhằm tìm kiếm tập ảnh tương tự, từ đó trích xuất ngữ nghĩa của các hình ảnh dựa trên cấu trúc từ điển và Ontology đã được xây dựng. Với mỗi ảnh truy vấn đầu vào, đặc trưng thị giác được trích xuất nhằm tìm kiếm các phân lớp ngữ nghĩa trên các cụm bằng phương pháp thống kê tần suất. Tập ảnh tương tự theo ngữ nghĩa được truy vấn trên Ontology đồng thời ngữ nghĩa của phân lớp được truy xuất trên cấu trúc từ điển. Trên cơ sở các phương pháp đã đề xuất, một mô hình tìm kiếm ảnh tương tự và trích xuất ngữ nghĩa cho hình ảnh đầu vào được xây dựng. Kết quả thực nghiệm được chúng tôi đánh giá trên bộ ảnh COREL, ImageCLEF, đồng thời so sánh với các công trình đã được công bố gần đây. Theo như kết quả phương pháp đề xuất, mô hình truy vấn ảnh của chúng tôi có tính khả thi và áp dụng hiệu quả trong các hệ truy vấn ảnh theo ngữ nghĩa trên nhiều lĩnh vực khác nhau.

Từ khóa: Truy vấn theo ngữ nghĩa, gom cụm, từ điển, ảnh tương tự, Ontology.

I. GIỚI THIỆU

Cùng với sự phát triển của internet, dữ liệu ảnh số khổng lồ đã được tạo ra dẫn đến bài toán tìm kiếm ảnh tương tự đang được quan tâm. Các phương pháp tìm kiếm ảnh thông dụng hiện nay bao gồm: tìm theo từ khóa TBIR (*Text-based Image Retrieval*), tìm theo nội dung CBIR (*Content-based Image Retrieval*) hay tìm theo ngữ nghĩa SBIR (*Semantic-based Image Retrieval*) [1-12]. Tìm kiếm hình ảnh dựa trên văn bản (TBIR) là hệ thống truy xuất hình ảnh bằng các câu truy vấn bằng văn bản. Việc trích xuất một hình ảnh tương tự dựa trên quá trình lập chỉ mục, các mô tả, chú thích hình ảnh được cung cấp từ người dùng [1, 2, 3]. Để giải quyết những hạn chế của phương pháp TBIR, phương pháp tìm kiếm ảnh dựa trên nội dung được phát triển. Phương pháp này tập trung vào việc trích xuất và so sánh các đặc trưng cấp thấp từ hình ảnh [4, 5, 6]. Các nhà nghiên cứu trong thập kỷ qua đã chứng minh tính hiệu quả và chính xác của các kỹ thuật dựa trên CBIR. Tuy nhiên, hạn chế của phương pháp này là phụ thuộc vào những kỹ thuật trích xuất đặc trưng hình ảnh cấp thấp đồng thời chưa tạo ra ngữ nghĩa cấp cao cho hình ảnh [7, 8, 9]. Để khắc phục các nhược điểm này, phương pháp tìm kiếm ảnh theo tiếp cận ngữ nghĩa (SBIR) đã được đề xuất. Phương pháp SBIR thực hiện việc tìm kiếm một tập hình ảnh tương tự với một ảnh truy vấn dựa trên ngữ nghĩa thị giác đồng thời tạo ra một ảnh xạ chuyển đổi từ nội dung hình ảnh thành ngữ nghĩa thị giác cũng như ngữ nghĩa mô tả hình ảnh [10, 11, 12].

Trong những thập niên gần đây, tìm kiếm ảnh theo tiếp cận ngữ nghĩa được các nhà khoa học đặc biệt quan tâm và đã sử dụng rất nhiều phương pháp khác nhau như: dựa vào ontology, dựa vào các công cụ học máy, dựa vào phản hồi liên quan (*Relative Feedback*), dựa vào tạo mẫu ngữ nghĩa (*Semantic Template*)... Trong đó, ontology là một trong những công cụ biểu diễn ngữ nghĩa tri thức tiên tiến nhất hiện nay. Hệ truy vấn ảnh SBIR tập trung vào nghiên cứu các kỹ thuật để giảm “khoảng cách ngữ nghĩa” giữa các đặc trưng cấp thấp và ngữ nghĩa cấp cao của hình ảnh [13, 14, 15]. Quá trình ánh xạ ngữ nghĩa được sử dụng để tìm khái niệm phù hợp cho đối tượng hình ảnh thông qua các công cụ học máy có giám sát hoặc không giám sát để liên kết các đặc trưng cấp thấp với khái niệm đối tượng cấp cao [16, 17]. Gần đây, các nhà nghiên cứu đã đề xuất nhiều phương pháp truy xuất hình ảnh đạt được nhiều kết quả khả quan. Tuy nhiên, bài toán tìm kiếm ảnh theo ngữ nghĩa vẫn còn là một thách thức lớn đối với cộng đồng truy vấn ảnh.

Trong bài báo này, chúng tôi đề xuất một mô hình tìm kiếm và tra cứu ngữ nghĩa hình ảnh dựa trên phương pháp gom cụm và ontology. Để thực hiện vấn đề này, chúng tôi gom cụm dữ liệu nhằm tìm kiếm tập ảnh tương tự. Với mỗi ảnh truy vấn đầu vào, đặc trưng thị giác được trích xuất nhằm tìm kiếm các phân lớp ngữ nghĩa trên các cụm bằng phương pháp thống kê tần suất. Tập ảnh tương tự theo ngữ nghĩa được truy vấn trên Ontology đồng thời ngữ nghĩa của phân lớp được truy xuất trên cấu trúc từ điển.

Đóng góp của bài báo gồm: (1) Xây dựng một phương pháp gom cụm cải tiến dựa trên K-Means nhằm tạo ra một mô hình phân loại dữ liệu cũng như giúp quá trình tìm kiếm được hiệu quả về tốc độ và độ chính xác; (2) Trích xuất đặc trưng thị giác nhằm tìm kiếm các phân lớp ngữ nghĩa; (3) Xây dựng một ontology cho bộ dữ liệu hình ảnh dựa trên Protégé nhằm tạo ra một ontology miền; (4) Đề xuất một cấu trúc từ điển nhằm mô tả ngữ nghĩa cho các đối tượng hình ảnh; (5) Xây dựng một hệ truy vấn ảnh theo ngữ nghĩa mà dữ liệu đầu vào là một ảnh truy vấn dựa vào cấu trúc gom cụm và ontology đã xây dựng.

Phần còn lại của bài báo được tổ chức như sau: Phần II, giới thiệu các công trình liên quan, nhằm phân tích ưu khuyết điểm những công trình đã có và đưa ra cách khắc phục cũng như phân tích tính khả thi của hướng tiếp cận của bài báo. Phần III, xây dựng mô hình cho bài toán truy vấn ảnh theo tiếp cận ngữ nghĩa sử dụng ontology và cấu trúc từ điển. Phần IV, xây dựng ứng dụng thực nghiệm và đánh giá kết quả dựa trên cơ sở lý thuyết đã đề nghị. Phần V, đưa ra kết luận và hướng phát triển.

II. CÁC CÔNG TRÌNH LIÊN QUAN

Các hệ thống truy vấn ảnh trực tuyến ngày càng phát triển. Nhiều kỹ thuật khác nhau được áp dụng cho bài toán tìm kiếm ảnh theo nội dung và theo hướng tiếp cận ngữ nghĩa. Trong đó, ontology là công nghệ biểu diễn ngữ nghĩa được nhiều nhà nghiên cứu quan tâm và đề xuất trong các mô hình tìm kiếm ảnh theo tiếp cận ngữ nghĩa nhằm nâng cao hiệu suất truy vấn. T. Kanimozhi và cộng sự (2013) đã cung cấp tổng quan về phát triển ontology và cách tạo truy vấn SPARQL [18], Kurtz và cộng sự (2014) đã đề xuất một khung ngữ nghĩa mới, cho phép so sánh sự giống nhau giữa hình ảnh dựa trên chủ thích từ một ontology [19], Juli Rejito và cộng sự (2016) đã gom cụm các vectơ đặc trưng dựa trên histogram và đánh chỉ mục ảnh để thực hiện tìm kiếm nhanh tập ảnh tương tự [20], Filali và cộng sự (2016) đã đề xuất một hệ truy vấn ảnh dựa trên từ vựng thị giác (*Visual Vocabulary*) và ontology [21], Mohamed Ouhda và cộng sự (2018) đã phân đoạn ảnh và trích xuất đặc trưng dựa trên vùng sử dụng giá trị kỳ vọng và độ lệch chuẩn của không gian màu RGB nhằm phân cụm và tra cứu hình ảnh tương tự [22], M. N. Asim và cộng sự (2019), đã xem xét các phương pháp truy xuất thông tin dựa trên ontology áp dụng cho truy vấn văn bản, dữ liệu đa phương tiện (hình ảnh, video, audio) và dữ liệu đa ngôn ngữ [23], M. H. Kaur và cộng sự (2019) đã sử dụng thuật toán K-Means kết hợp phân lớp SVM (*Support Vector Machine*) nhằm thực hiện tìm kiếm ảnh dựa trên các đặc trưng cấp thấp [24].

T.Kanimozhi và cộng sự (2013) đã cung cấp tổng quan về phát triển ontology và cách tạo truy vấn SPARQL. Phương pháp này, tích hợp ontology và SPARQL cho web ngữ nghĩa giúp cải thiện hiệu quả tìm kiếm hình ảnh [18]. Trong công trình này, nhóm tác giả chỉ mới thử nghiệm tạo ontology bằng phần mềm Protégé. Kết quả của nhóm tác giả chưa tạo ra cấu trúc dữ liệu để lưu trữ hình ảnh và chưa xây dựng phần mềm để thực hiện truy vấn hình ảnh tương tự.

Kurtz và cộng sự (2014) đã đề xuất một khung ngữ nghĩa mới, cho phép so sánh sự giống nhau giữa hình ảnh dựa trên chủ thích từ một ontology. Khung ontology này bao gồm đặc trưng ngữ nghĩa của hình ảnh và có thể truy vấn hình ảnh tương tự dựa trên mối quan hệ ngữ nghĩa. Hệ thống sử dụng bộ ảnh y khoa về bệnh gan để thực nghiệm và mang lại kết quả khả quan [19]. Trong công trình này, nhóm tác giả chỉ thử nghiệm trên một tập các hình ảnh rất nhỏ (77 hình ảnh), đồng thời chỉ mới xây dựng ontology miền dựa trên chủ thích hình ảnh.

Theo Juli Rejito và cộng sự (2016) đã đề xuất một hệ truy vấn ảnh dựa trên thuật toán K-Means bằng cách phân cụm các đặc trưng histogram của hình ảnh; sau đó, nhóm tác giả thực hiện tạo định danh cho tập hình ảnh nhằm tra cứu nhanh tập ảnh kết quả. Kết quả thực nghiệm của bài báo cho thấy phương pháp áp dụng của nhóm tác giả là hiệu quả [20]. Tuy nhiên, việc phân cụm trên các đặc trưng dựa vào K-Means sẽ tốn kém thời gian xác định tâm cụm, cũng như số cụm ban đầu. Đồng thời, khi dữ liệu tăng trưởng sẽ tốn kém nhiều chi phí định danh các hình ảnh theo từng cụm.

Filali và cộng sự (2016) đã đề xuất một hệ truy vấn ảnh dựa trên từ vựng thị giác và ontology. Với mỗi ảnh truy vấn, xây dựng một từ vựng thị giác và ontology dựa trên chủ thích của hình ảnh. Ontology được làm phong phú bởi các khái niệm và mối quan hệ được trích xuất từ tài nguyên từ vựng BabelNet. Thực nghiệm của phương pháp đề xuất minh chứng hiệu suất truy vấn là khả thi [21]. Tuy nhiên, phương pháp đề xuất thực nghiệm chưa xây dựng cấu trúc để lưu trữ dữ liệu hình ảnh và chưa kết hợp truy vấn ảnh theo nội dung với truy vấn theo ngữ nghĩa.

Mohamed Ouhda và cộng sự (2018) sử dụng kỹ thuật K-Means để phân cụm các đặc trưng dựa trên vùng của hình ảnh. Đặc trưng này được trích xuất trên từng vùng đối tượng theo vị trí và màu sắc trong không gian RGB để từ đó thực hiện bài toán tra cứu ảnh theo nội dung. Theo thực nghiệm, phương pháp đề xuất hiệu quả trên nhiều bộ ảnh khác nhau và có độ chính xác cao [22]. Tuy nhiên phương pháp này tốn kém nhiều chi phí thời gian trong việc phân cụm.

M. N. Asim và cộng sự (2019), đã xem xét các phương pháp truy xuất thông tin dựa trên ontology áp dụng cho truy vấn văn bản, dữ liệu đa phương tiện (hình ảnh, video, audio) và dữ liệu đa ngôn ngữ. Trong công trình này, tác giả sử dụng ngôn ngữ bộ ba RDF để thực hiện lưu trữ và truy vấn trên ontology [23]. Kết quả thực nghiệm cho thấy, nhóm tác giả mới đề xuất mô hình sử dụng ontology để truy vấn đa đối tượng, chưa đề cập đến kết quả thực nghiệm cụ thể để so sánh với các công trình trước đây.

Theo H. K. Maur và cộng sự (2019) đã vận dụng thuật toán K-Means nhằm thực hiện phân cụm các hình ảnh dựa vào đặc trưng cấp thấp gồm: màu sắc, hình dạng và cấu trúc. Đồng thời, nhóm tác giả áp dụng phương pháp phân lớp SVM để phân loại hình ảnh, từ đó thực hiện tra cứu ảnh. Theo thực nghiệm, phương pháp đề xuất của tác giả đã giải quyết hiệu quả cho bài toán tra cứu ảnh trên nhiều bộ dữ liệu khác nhau [24]. Tuy nhiên, việc phân cụm của nhóm tác giả còn phụ thuộc vào số tâm cụm ban đầu và nếu như dữ liệu tăng trưởng thì việc phân cụm mất nhiều chi phí thời gian.

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 nhằm giảm khoảng cách ngữ nghĩa của hình ảnh. Các kỹ thuật đã khảo sát chỉ mới được giải quyết một phần vì các phương pháp đề xuất phụ thuộc ontology miền hợp lý, các dữ liệu lớn và các ngôn ngữ truy vấn. Vì thế, vấn đề xây dựng ontology cho bộ dữ liệu lớn là một trong những thách thức cho các nhà nghiên cứu. Trên cơ

sở kế thừa và khắc phục những hạn chế của các công trình nghiên cứu liên quan, chúng tôi đề xuất các phương pháp để xây dựng cấu trúc cây phân cụm dữ liệu hình ảnh. Bên cạnh đó, chúng tôi phát triển thuật toán K-Means và độ đo Euclide để phân cụm các vectơ đặc trưng thị giác của hình ảnh; Đặc trưng thị giác được trích xuất nhằm tìm kiếm các phân lớp ngữ nghĩa trên các cụm bằng phương pháp thống kê tần suất; Một Ontology được xây dựng để trích xuất ngữ nghĩa cho bộ dữ liệu hình ảnh. Câu truy vấn SPARQL được tự động tạo dựa trên các vectơ từ thị giác và thực hiện truy vấn hình ảnh ngữ nghĩa trên Ontology, nhằm trích xuất ngữ nghĩa của phân lớp được truy xuất trên cấu trúc từ điển.

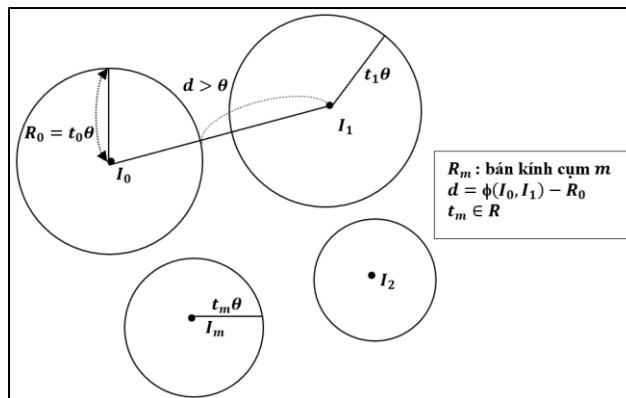
III. MÔ HÌNH HỆ TRUY VẤN ẢNH THEO TIẾP CẬN NGŨ NGHĨA DỰA TRÊN ONTOLOGY

A. Kỹ thuật gom cụm

1) Mô tả kỹ thuật gom cụm

Trong thuật toán K-means, ba tham số cần được khởi tạo ban đầu gồm: số lượng cụm k , tâm cụm và độ đo tương tự; Ngoài ra, nếu bổ sung phần tử mới vào cụm thì phải xác định lại tâm cụm mới. Tuy nhiên, với một bộ dữ liệu bất kỳ cho trước, số lượng các cụm rất khó xác định cũng như việc tăng trưởng dữ liệu có thể làm gia tăng số lượng cụm, điều này gây ra tốn kém nhiều chi phí về thời gian và quá trình thực thi khi tái tạo lại số cụm.

Trong bài báo này, chúng tôi đề xuất một cải tiến thuật toán K-Means nhằm phân cụm dữ liệu và tăng trưởng số cụm theo bộ dữ liệu. Để thực hiện được điều này, chúng tôi sử dụng một giá trị ngưỡng đo độ tương tự giữa các đối tượng dữ liệu, ngưỡng này được kí hiệu là θ . Ngưỡng θ là một giá trị trong khoảng $[0, 1]$ để đánh giá độ tương tự các phần tử thuộc một cụm. Nếu ngưỡng θ càng lớn thì cụm sẽ chứa nhiều phần tử; Trong trường hợp $\theta = 1$ thì tất cả dữ liệu thuộc về một cụm. Do đó, giá trị ngưỡng θ cần phải được lựa chọn để phân cụm bộ dữ liệu ban đầu, ví dụ nếu $\theta = 0,25$ tức là các phần tử có độ tương tự trong cùng một cụm là 75 %. Giá trị θ được chọn tùy vào thực nghiệm và tùy vào bộ dữ liệu đầu vào. Trong thực nghiệm của bài báo này, chúng tôi chọn $\theta = 0,03$ để thực hiện phân chia dữ liệu và dần nới giá trị θ đến 0,06, nếu giá trị $\theta > 0,06$ thì cụm được tách làm đôi bằng cách phân bố dữ liệu cụm hiện tại thành hai cụm khác nhau. Trên cơ sở ngưỡng θ , thuật toán K-Means được cải tiến bằng cách không xác định trước số tâm cụm, vì vậy số cụm dữ liệu tăng trưởng theo sự gia tăng của số lượng hình ảnh.



Hình 1. Mô tả kỹ thuật phân cụm

Trong Hình 1, phương pháp phân cụm dựa trên K-Means được mô tả, với C_i, I_i, R_i, t_i lần lượt là cụm, tâm, bán kính và hệ số dẫn nới bán kính của cụm i , cụm đầu tiên có bán kính và tâm lần lượt là $R_0 = t_0 * \theta, I_0 = v_0$ và d là khoảng cách từ tâm cụm mới đến đường biên của cụm gần nhất. Trên cơ sở giá trị d , qui tắc phân cụm được mô tả như sau:

Gọi $L = \{v_0, v_1, \dots, v_{n-1}\}$ là tập vectơ ảnh ban đầu

- **Bước 1:** Tạo cụm C_0 đầu tiên: $R_0 = \theta$ và $I_0 = v_0$
- **Bước 2:** Xét $v_i \in L$ với $i = 1, \dots, n - 1$

Tìm cụm C_t thỏa $\phi(v_i, I_t) - R_t = \min\{\phi(v_i, I_j) - R_j\}$ với $j = 0, \dots, m - 1$
(trong đó m là số lượng cụm đã tạo)

$$d = \phi(v_i, I_t) - R_t$$

Nếu $(d \leq \theta)$ thì

$$C_t = C_t \cup \{v_i\}; // \text{Thêm } v_i \text{ vào cụm } C_t$$

Nếu $\phi(v_i, I_t) > R_t$ thì

$$R_t = \phi(v_i, I_t); // \text{Cập nhật bán kính cụm } C_t$$

Ngược lại //Tạo cụm mới C_m có tâm là v_i

$$I_m = v_i;$$

$$R_m = \theta;$$

$$\Omega = \Omega \cup C_m;$$

2) Thuật toán gom cụm dữ liệu

Theo phương pháp đề xuất như trên, thuật toán gom cụm ảnh được cải tiến từ K-Means được mô tả như sau:

Thuật toán CTIR

Đầu vào: Ngưỡng tương tự θ và L (tập véctor ảnh ban đầu)

Đầu ra: Tập các cụm Ω

Function Clustering_theta(θ, L)

Begin

 Khởi tạo $\Omega = \emptyset$;

Foreach $\langle v_i \rangle \in L$ **do**

If ($\Omega = \emptyset$) **then** //Tạo cụm đầu tiên C_0

$I_0 = v_0$;

$R_0 = \theta$;

$\Omega = \{C_0\}$;

Else

 Tìm cụm $C_t \in \Omega$ thỏa $\phi(v_i, I_t) - R_t = \min\{\phi(v_i, I_j) - R_j\}$ với $j = 0, \dots, m - 1$

 (trong đó m là số lượng cụm đã được tạo)

$d = \phi(v_i, I_t) - R_t$

If ($d \leq \theta$) **then**

$C_t = C_t \cup \{v_i\}$; //Thêm v_i vào cụm C_t

If $\phi(v_i, I_t) > R_t$ **then**

$R_t = \phi(v_i, I_t)$; // Cập nhật bán kính cụm C_t

EndIf;

Else //Tạo cụm mới C_m có tâm là v_i

$I_m = v_i$;

$R_m = \theta$;

$\Omega = \Omega \cup C_m$;

EndIf;

EndIf;

EndFor;

Return Ω ;

End.

Mệnh đề 1. Thuật toán CTIR có độ phức tạp $O(n * m)$, với n, m lần lượt là số phần tử của bộ dữ liệu và số cụm tạo ra.

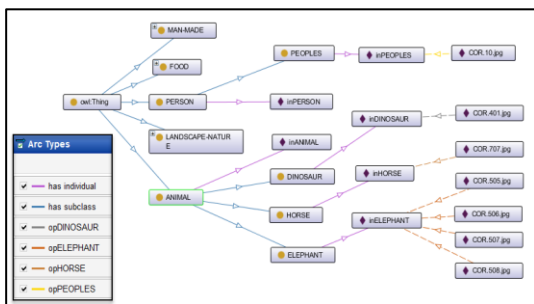
Chứng minh: Giả sử n phần tử trong bộ dữ liệu được phân thành m cụm. Thuật toán cần phải duyệt qua n phần tử dữ liệu, với mỗi phần tử cần phải duyệt qua m cụm để phân bố vào cụm gần nhất. Trong trường hợp xấu nhất, thuật toán CTIR có số lần duyệt là $n * m$ để phân bố các phần tử vào các cụm. Khi số phần tử dữ liệu lớn, giá trị n và m rất lớn thì độ phức tạp của thuật toán CTIR là $O(n * m)$.

B. Xây dựng ontology

Để tạo ra một tập ảnh tương tự và các mô tả ngữ nghĩa của của hình ảnh truy vấn, chúng tôi xây dựng ontology miền cho bộ dữ liệu hình ảnh thực nghiệm và thực hiện truy vấn theo ngôn ngữ SPARQL. Quá trình này đòi hỏi nguồn nhân lực và mất khá nhiều thời gian. Trong phần này, một ontology miền (domain ontology) cho hình ảnh được xây dựng dựa trên phần mềm Protégé. Bộ dữ liệu ảnh COREL (10 phân lớp), ImageCLEF(276 phân lớp) được sử dụng để xây dựng ontology miền.

Quy trình xây dựng ontology được thực hiện trên phần mềm Protégé và kết quả xây dựng được minh họa như trong Hình 2, Hình 3.

Một ontology cho hình ảnh được tạo trên Protégé có thể được lưu trữ dưới dạng tệp OWL. Các hình ảnh được chú thích với các mô tả ngữ nghĩa lưu trữ cùng với tập dữ liệu. Một ngôn ngữ truy vấn SPARQL được sử dụng để thực hiện truy vấn ngữ nghĩa hình ảnh trên ontology được tạo.



Hình 2. Một ví dụ cho ontology trực quan trên Protégé

```

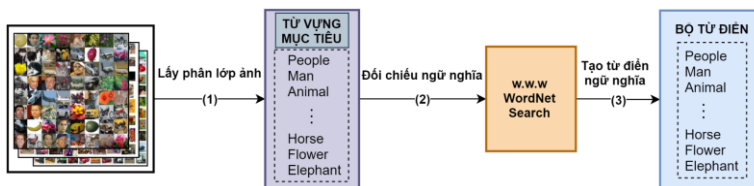
1 @prefix : <http://sbir-hcm.vn/> .
2 @prefix owl: <http://www.w3.org/2002/07/owl#> .
3 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
4 @prefix xml: <http://www.w3.org/XML/1998/namespace> .
5 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
6 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
7 @prefix sbir: <http://sbir-hcm.vn/> .
8 @base <http://sbir-hcm.vn/> .
9
10 <http://sbir-hcm.vn/> rdf:type owl:Ontology .
11
12 #####
13 # Annotation properties
14 #####
15
16 ### http://sbir-hcm.vn/Description
17 sbir:Description rdf:type owl:AnnotationProperty .
18
19 #####
20 # Object Properties
21 #####
22 ### http://sbir-hcm.vn/SPANIMAL
23 sbir:SPANIMAL rdf:type owl:ObjectProperty .

```

Hình 3. Một ví dụ ontology dưới dạng ngôn ngữ RDF

C. 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 mô tả như trong Hình 4.



Hình 4. Quy trình xây dựng từ điển ngữ nghĩa cho hình ảnh

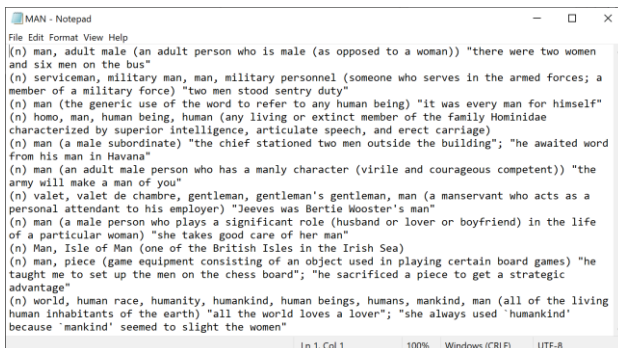
Quy trình xây dựng từ điển gồm ba bước:

Bước 1: Với mỗi ảnh của bộ dữ liệu ảnh, thực hiện lấy phân lớp hình ảnh để tạo tập từ vựng mục tiêu.

Bước 2: Thực hiện kết xuất các khái niệm là danh từ và đồng nghĩa với từ vựng mục tiêu dựa vào dựa vào mạng từ (ontology từ vựng) WORDNET theo cấu trúc: danh từ, từ khóa, định nghĩa cho từ vựng.

Bước 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 như một ví dụ trong Hình 5, ví dụ về từ vựng "MAN" được định nghĩa.



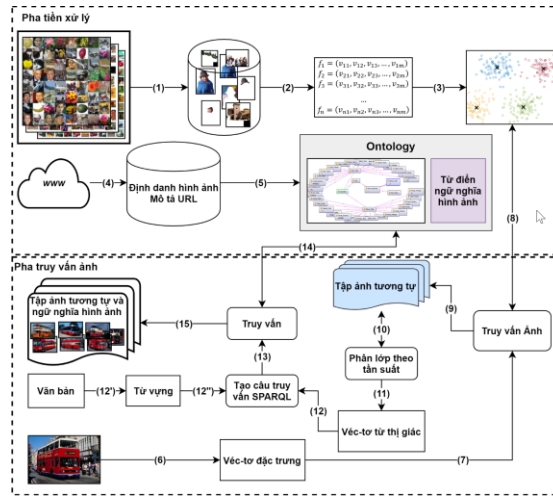
Hình 5. Một ví dụ về khái niệm từ vựng trong từ điển ontology

Theo Hình 5, có thể thấy, với từ vựng mục tiêu "MAN", được thực hiện đổi chiếu trên WORDNET để lấy các định nghĩa và các từ đồng nghĩa: là các danh từ (n); từ khóa: "man", "adult male" ...; khái niệm; các ví dụ để làm rõ ngữ nghĩa theo ngữ cảnh của từ vựng.

D. Mô hình và thuật toán truy vấn ảnh

1) Mô hình hệ truy vấn ảnh

Dựa trên ontology được xây dựng và dữ liệu được gom cụm dựa trên thuật toán CTIR, mô hình hệ truy vấn theo tiếp cận ngữ nghĩa được đề xuất và được mô tả trong Hình 6. Hệ truy vấn bao gồm 2 pha: (1) pha tiền xử lý nhằm trích xuất 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 Euclide; tạo một ontology cho hình ảnh dựa trên phần mềm Protégé; 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, vectơ từ thị giác và chú thích của từng hình ảnh trong tập ảnh và truy vấn trên ontology dựa vào vectơ từ thị giác, đồng thời ngữ nghĩa hình ảnh được truy xuất từ bộ từ điển được xây dựng.



Hình 6. Mô hình hệ truy vấn ảnh dựa trên phương pháp gom cụm và ontology

Pha tiền xử lý: Kết quả của pha tiền xử lý là xây dựng được mô hình gom cụm dữ liệu dựa trên vectơ đặc trưng của tập dữ liệu ảnh, đồng thời xây dựng được ontology miền gồm các bước như sau:

- (1) Thực hiện kỹ thuật phân đoạn hình ảnh;
- (2) Trích xuất các vectơ đặc trưng cấp thấp và phân lớp của hình ảnh;
- (3) Thực hiện gom cụm dữ liệu theo phương pháp K-Means cải tiến;
- (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) Tạo và bổ sung ontology hình ảnh, đồng thời tạo cấu trúc từ điển ngữ nghĩa cho hình ảnh.

Pha truy vấn ảnh: Pha truy vấn hình ảnh bao gồm hai giai đoạn chính (1) Truy vấn hình ảnh dựa trên cụm dữ liệu đã phân hoạch; (2) Truy vấn ngữ nghĩa dựa vào ontology và cấu trúc từ điển.

Giai đoạn 1: Truy vấn hình ảnh dựa trên cụm dữ liệu đã phân hoạch

- (6) Với một hình ảnh truy vấn đầu vào, hệ thống thực hiện trích xuất các vectơ đặc trưng;
- (7) (8) Thực hiện thuật toán truy vấn trên cụm dữ liệu đã phân hoạch;
- (9) Kết quả của quá trình truy vấn là tập các hình ảnh tương tự và phân lớp của các hình ảnh;
- (10) Thực hiện phân lớp tập ảnh tương tự theo tần suất xuất hiện;
- (11) Kết quả phân lớp là tập vectơ từ thị giác của ảnh truy vấn.

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

- (12) Tạo câu truy vấn SPARQL từ các lớp của vectơ từ thị giác của ảnh truy vấn;
Hoặc
- (12') Từ một văn bản Text đầu vào trích xuất các WORDS tập các lớp Classes của ontology →(12'') tạo câu truy vấn SPARQL từ các WORDS;
- (13) (14) Thực hiện truy vấn ảnh tương tự theo tiếp cận ngữ nghĩa dựa trên ontology đã xây dựng; Trích xuất ngữ nghĩa dựa trên từ điển;
- (15) Kết quả truy vấn tạo ra tập các hình ảnh tương tự và các mô tả ngữ nghĩa của hình ảnh.

2) Thuật toán truy vấn ảnh

a) Truy vấn ảnh theo nội dung dựa trên cụm dữ liệu

Trên cơ sở tập các cụm Ω đã được phân hoạch theo thuật toán CTIR, quá trình tìm kiếm ảnh được thực hiện bằng cách chọn ra cụm C_m có tâm gần nhất với ảnh tra cứu. Tuy nhiên, chúng tôi chọn thêm các cụm láng giềng của C_m dựa trên độ đo giữa các tâm cụm để tăng số lượng kết quả ảnh tra cứu. Khi đó các bước thuật toán tìm kiếm ảnh như sau:

- Bước 1:** Tìm cụm C_m có tâm gần với vectơ ảnh tra cứu nhất.
- Bước 2:** Tìm h cụm láng giềng với cụm C_m . Tập \forall chứa h cụm láng giềng và cụm C_m .
- Bước 3:** Tìm tập \mathcal{E} chứa tất cả các vectơ trong \forall .
- Bước 4:** Sắp xếp \mathcal{E} tăng dần theo độ đo.

Thuật toán SEIR:

Đầu vào: vectơ đặc trưng p (ảnh tìm kiếm), tập cụm Ω và ngưỡng tìm kiếm σ .

Đầu ra: tập Ψ chứa các id (định danh) của các ảnh tương tự với ảnh tìm kiếm.

Function ClusterRetrieval(p, Ω, σ)

Begin

Khởi tạo $\Psi = \emptyset$;

Tìm cụm $C_k \in \Omega : \phi(p, I_k) = \min\{\phi(p, I_i), i = 1..m\}$;
 //Tìm h cụm láng giềng với C_k
 Sắp xếp Ω tăng dần theo $\phi(v_t, v_k) - (R_t + R_k)$
 Khởi tạo $\mathcal{E} = \emptyset$;
If $(\phi(p, I_k) - (R_t + R_k) < \sigma)$ then
 $\mathcal{E} = \mathcal{E} \cup C_i$ với $i = 0..m - 1$;
EndIf
 Sắp xếp tập \mathcal{E} theo $\phi(l, p)$ với $\forall l \in \mathcal{E}$
 Tạo tập định danh hình ảnh Ψ theo thứ tự sắp xếp của tập \mathcal{E} .
Return Ψ ;

End.

Mệnh đề 2. Thuật toán SEIR có độ phức tạp $O(m^2)$, với m là số cụm tạo ra.

Chứng minh: Giả sử tập cụm đầu vào cho thuật toán SEIR có m cụm, khi đó thuật toán thực hiện việc tìm cụm gần nhất đối với ảnh đầu vào. Quá trình tìm kiếm này cần duyệt qua từng cụm, nghĩa là số lần so sánh là m tương ứng với m cụm. Sau khi thực hiện tìm cụm gần nhất với ảnh láng giềng, thuật toán SEIR tìm các cụm lân cận bằng cách sắp xếp lại tập cụm theo độ tương tự của cụm đã tìm được, số phép toán tối đa trong việc sắp xếp này là $O(m^2)$. Sau khi thực hiện tìm kiếm các cụm láng giềng, thuật toán thực hiện việc sắp xếp các hình ảnh theo độ đo tương tự với ảnh đầu vào (*tuy nhiên việc sắp xếp này có thể xử lý bên ngoài thuật toán nên độ phức tạp thuật toán SEIR có thể không bao gồm việc sắp xếp các hình ảnh theo độ đo tương tự*). Vì vậy, độ phức tạp của thuật toán SEIR là $O(m^2)$.

b) Thuật toán phân lớp ảnh truy vấn từ tập ảnh tương tự

Quá trình truy vấn hình ảnh trên các cụm dữ liệu đã phân hoạch tạo ra tập ảnh tương tự và tập các từ thị giác đại diện cho tập ảnh này. Việc phân lớp ảnh truy vấn sẽ được thực hiện dựa vào tập ảnh tương tự. Véc tơ từ thị giác của ảnh truy vấn được xây dựng dựa trên tập các từ thị giác bằng cách lấy các từ có tần suất xuất hiện nhiều nhất trong tập từ thị giác. Số lượng từ của véc tơ từ thị giác bằng với số từ thị giác của hình ảnh truy vấn (topK). Hình 7 là một ví dụ véc tơ từ thị giác của bộ ảnh tương tự được tạo ra từ quá trình truy vấn hình ảnh 25118.jpg của bộ ảnh ImageCLEF. Véc tơ từ thị giác của ảnh 25118.jpg là 3 từ có tần suất xuất hiện nhiều nhất trong tập ảnh tương tự: sky-blue (210), ground (162), man (130).

Segmented image			Objects detected			Retrieved label Occurrences		
Label	Occurrences	Name	Label	Occurrences	Name	Label	Occurrences	Name
224	1	sky-blue	224	210	sky-blue	224	210	sky-blue
160	1	man	118	162	ground	118	162	ground
118	1	ground	160	130	man	160	130	man
						Intersected ratio: 100%		

Hình 7. Ví dụ véc tơ từ thị giác cho ảnh truy vấn 25118.jpg

Thuật toán CIMG:

Input: Tập ảnh tương tự IMG .

Output: Véc tơ từ thị giác W

Function getVisualWord(IMG)

Begin

$L = \emptyset; W = \emptyset; C = \emptyset;$

For ($I \in IMG$) **do**

$L = L \cup \{I.label\}$

End For

For ($L_i \in L$) **do**

$W = topK(freq(L_i, L));$

End For

Return W ;

End.

c) Truy vấn ảnh theo ngữ nghĩa trên ontology

Quá trình truy vấn ảnh theo tiếp cận ngữ nghĩa dựa trên ontology được thực hiện bằng hai cách:

Cách 1: Từ một ảnh đầu vào, được trích xuất tập véc tơ từ thị giác, dựa vào tập véc tơ từ hệ thống sẽ tự động tạo ra câu truy vấn 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.

Cách 2: Với một văn bản Text theo người dùng được đưa ra, hệ thống sẽ phân tích và chọn lọc để lấy ra các lớp có trong danh sách lớp của ontology. Từ các lớp được trích xuất, hệ thống sẽ tự động tạo ra câu truy vấn 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 OSBIR:

Input: Tập véctor từ Y_I của ảnh I , hoặc tập các lớp Y_T của một văn bản Text
Output: Tập ảnh theo ngữ nghĩa \mathbb{C}_{SI} và tập các mô tả ngữ nghĩa ảnh truy vấn \mathcal{M}_{SI}
Function getSBIR_ontology(string [] Y_I, Y_T)
Begin
 $\mathbb{C}_{SI} = \emptyset, \mathcal{M}_{SI} = \emptyset;$
 $(\mathbb{C}_{SI}, \mathcal{M}_{SI}) = CreateSPAQRL(Y_I, Y_T);$
 Return (($\mathbb{C}_{SI}, \mathcal{M}_{SI}$);
End.

d) Tra cứu ngữ nghĩa dựa trên từ điển

Mô tả: Dựa vào phân lớp ảnh đầu vào, hệ thống thực hiện truy vấn SPAQRL trên ontology kết quả trả về tập ảnh theo ngữ nghĩa \mathbb{C}_{SI} và tập các mô tả ngữ nghĩa ảnh truy vấn \mathcal{M}_{SI} . Từ tập kết quả mô tả ngữ nghĩa của ảnh truy vấn, hệ thống truy cập vào từ điển của ontology để truy xuất ngữ nghĩa cấp cao của lớp trong hình ảnh truy vấn.

Thuật toán DSBIR:

Input: Tập véctor từ Y_I của ảnh I , hoặc tập các lớp Y_T của một văn bản Text, file từ điển $Y_I[i].txt$
Output: Tập các mô tả ngữ nghĩa cấp cao \mathcal{H}_I của ảnh truy vấn I
Function getSBIR_dictionary(string [] Y_I, Y_T)
Begin
 $\mathbb{C}_{SI} = \emptyset, \mathcal{M}_{SI} = \emptyset;$
 $(\mathbb{C}_{SI}, \mathcal{M}_{SI}) = CreateSPAQRL(Y_I, Y_T);$
 $\mathcal{H}_I = extractSI(\mathcal{M}_{SI}, Y_I[i].txt);$
 Return $\mathcal{H}_I;$
End.

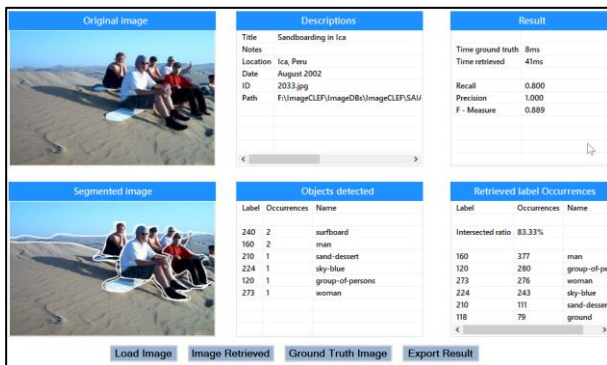
IV. THỰC NGHIỆM

A. Môi trường thực nghiệm

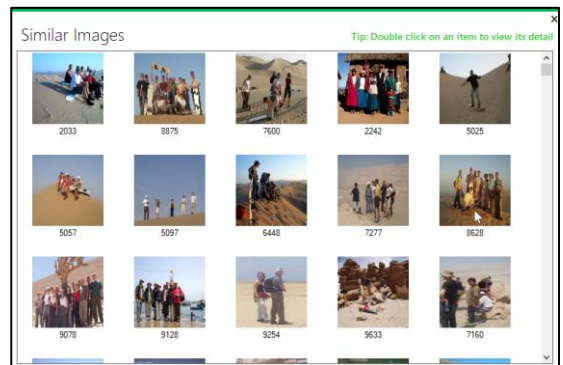
Pha tiền xử lý được thực hiện trên máy PC CPU 2.3GHz 8-core 9th-generation Intel Core i9, 16GB 2666MHz memory, 1TB flash storage. Pha tìm kiếm được thực nghiệm trên máy PC CPU Intel Core i7-6500U CPU @ 2.50GHz, 8.0GB RAM, hệ điều hành Windows 10 Pro 64 bit.

B. Ứng dụng thực nghiệm

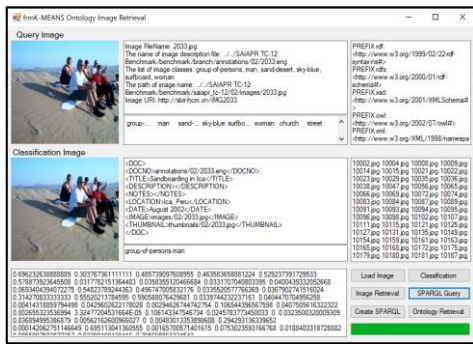
Trong bài báo này, chúng tôi tiến hành thực nghiệm trên hai bộ ảnh COREL, ImageCLEF với các chủ đề khác nhau. Bộ ảnh COREL có 1000 ảnh được chia thành 10 phân lớp: *beach, bus, castle, dinosaur, elephant, flower, horse, meal, mountain, peoples*. Bộ ảnh ImageCLEF chứa 20,000 ảnh, được chia thành 276 phân lớp. Ứng dụng thực nghiệm được minh họa trong Hình 8-11.



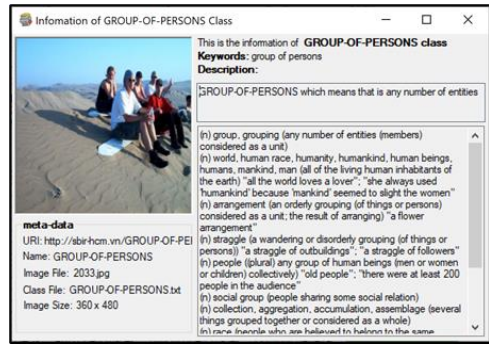
Hình 8. Giao diện truy vấn và phân lớp hình ảnh dựa trên phương pháp gom cụm cải tiến K-Means



Hình 9. Một kết quả truy vấn dựa trên phương pháp gom cụm cải tiến K-Means



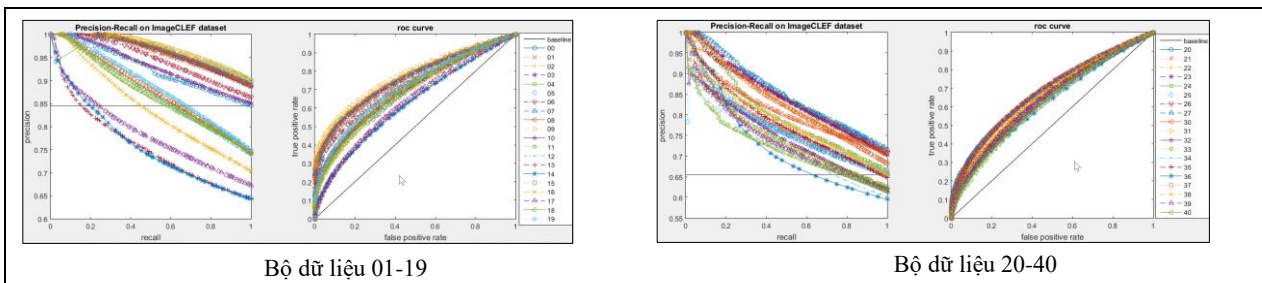
Hình 10. Giao diện truy vấn ngữ nghĩa hình ảnh dựa trên Ontolog



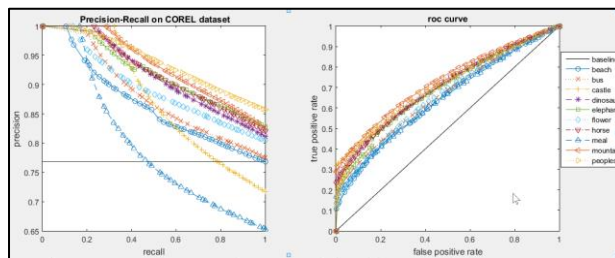
Hình 11. Giao diện trích xuất ngữ nghĩa dựa vào từ điển

C. Kết quả thực nghiệm

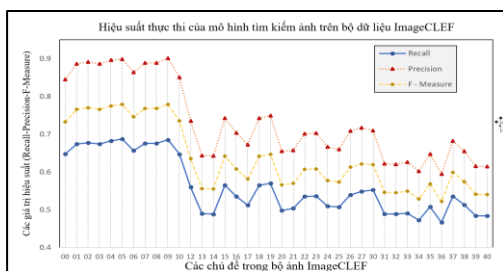
Để đánh giá hiệu quả của phương pháp tìm kiếm ảnh, phần thực nghiệm được đánh giá các giá trị gồm: độ chính xác (*precision*), độ phủ (*recall*) và độ đo dung hòa F-measure. Kết quả thực nghiệm được thể hiện như trong Hình 12-15. Mỗi đường cong trên đồ thị mô tả kết quả truy vấn từ một chủ đề ảnh trong bộ dữ liệu ImageCLEF, COREL.



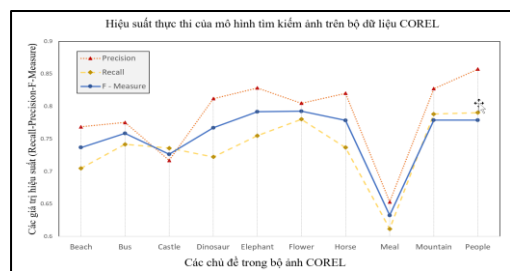
Hình 12. Precision-Recall và đường cong ROC của bộ dữ liệu 01-40 trên tập ảnh ImageCLEF



Hình 13. Precision-Recall và đường cong ROC của bộ dữ liệu ảnh COREL



Hình 14. Giá trị trung bình của Precision, Recall, F-measure của tập dữ liệu ImageCLEF



Hình 15. Giá trị trung bình của Precision, Recall, F-measure của tập dữ liệu COREL

Hiệu suất truy vấn của thuật toán gom cụm cải tiến K-Means được biểu diễn trong Bảng 1, Bảng 2. Để đánh giá hiệu suất của hệ tìm kiếm ảnh tương tự đề xuất, chúng tôi so sánh với kết quả của các công trình liên quan trước đó trên cùng tập dữ liệu được mô tả trong Bảng 3, Bảng 4.

Bảng 1. Hiệu suất truy vấn ảnh của phương pháp đề xuất trên tập ảnh ImageCLEF

Tập ảnh	Precision	Recall	F-Measure
00-09	0.884605851	0.673427364	0.764717822
10-19	0.720307622	0.547789515	0.622307938
20-30	0.686428196	0.525406381	0.594894109
31-40	0.628040842	0.493009485	0.55133095
Kết quả	0.729845628	0.559908186	0.633312705

Bảng 2. Hiệu suất tìm kiếm ảnh của phương pháp đề xuất trên bộ dữ liệu COREL

Tập ảnh	Precision	Recall	F-Measure
01-10	0.786146243	0.736479184	0.753905261

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

Phương pháp	Mean Average Precision (MAP)
Douik, 2016 [8]	0.7200
Bella M. I. T., 2019 [7]	0.6090
Abeer Al-Mohamade, 2020 [6]	0.7058
Phương pháp của chúng tôi	0.7861

Bảng 4. So sánh độ chính xác giữa các phương pháp trên bộ dữ liệu ImageCLEF

Phương pháp	Mean Average Precision (MAP)
Y. Cao, 2016 [13]	0.7236
M. Jiu, 2017 [12]	0.5970
N.M.Hai, 2019 [14]	0.6753
Phương pháp của chúng tôi	0.7298

Theo như kết quả phương pháp đề xuất, mô hình truy vấn ảnh của chúng tôi có tính khả thi và áp dụng hiệu quả trong các hệ truy vấn ảnh theo ngữ nghĩa trên nhiều lĩnh vực khác nhau.

V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Chúng tôi đã xây dựng một phương pháp truy vấn ngữ nghĩa hình ảnh dựa trên một cấu trúc từ điển được xây dựng bằng cấu trúc ontology và tra cứu theo chỉ mục. Dựa trên phương pháp này, chúng tôi đã đề xuất một mô hình tra cứu ảnh theo ngữ nghĩa và các thuật toán truy vấn ngữ nghĩa hình ảnh trên ontology và từ điển. Nhằm minh chứng tính đúng đắn của mô hình đã đề xuất, thực nghiệm được đánh giá trên các bộ ảnh COREL, ImageCLEF dựa trên độ chính xác và độ phủ của từng hình ảnh để từ đó xây dựng các đường cong precision - recall và đường ROC; độ chính xác trên các bộ ảnh lần lượt là 78,61 %, 72,98 %. Kết quả thực nghiệm được chúng tôi so sánh với các kết quả của các công trình đã công bố gần đây, từ đó minh chứng phương pháp của chúng tôi là hiệu quả. Hướng phát triển tiếp theo của bài báo này đó là chúng tôi kết hợp cấu trúc từ điển và ontology, đồng thời phân cụm hình ảnh đầu vào dựa trên cấu trúc R-Tree để cải tiến tốc độ và đảm bảo được độ chính xác cho bài toán tìm kiếm ảnh theo ngữ nghĩa.

VI. LỜI CẢM ƠN

Nhóm tác giả chân thành cảm ơn Trường Đại học Sư phạm TP. HCM, Trường Đại học Công nghiệp Thực phẩm TP. HCM, Trường Đại học Bà Rịa - Vũng Tàu là những nơi bảo trợ cho nghiên cứu này. Chúng tôi trân trọng cảm ơn nhóm nghiên cứu SBIR-HCM đã hỗ trợ về chuyên môn giúp chúng tôi hoàn thành bài nghiên cứu này.

TÀI LIỆU THAM KHẢO

- [1] Li, Wen, et al., "Text-based image retrieval using progressive multi-instance learning". In: 2011 International Conference on Computer Vision. IEEE, pp. 2049-2055, 2011.
- [2] Barrios, Juan Manuel; Díaz-espinoza, Diego; Bustos, Benjamin. "Text-based and content-based image retrieval on Flickr". In: 2009 Second International Workshop on Similarity Search and Applications. IEEE, pp. 156-157, 2009.
- [3] Tollari, Sabrina, et al., "Exploiting visual concepts to improve text-based image retrieval". In: European Conference on Information Retrieval. Springer, Berlin, Heidelberg, pp. 701-705, 2009.
- [4] Liu, Ying, et al., "A survey of content-based image retrieval with high-level semantics", Pattern recognition 40.1 pp. 262-282, 2007.
- [5] Karthikeyan, T.; Manikandrabhu, P.; Nithya, S., "A survey on text and content based image retrieval system for image mining". International Journal of Engineering, pp. 3, 2014.
- [6] Al-mohamade, Abeer; Bchir, Ouiem; Ben Ismail, Mohamed Maher. "Multiple Query Content-Based Image Retrieval Using Relevance Feature Weight Learning". Journal of Imaging, 6.1: 2, 2020.
- [7] Bella, Mary I. Thusnavis; Vasuki, A. "An efficient image retrieval framework using fused information feature". Computers & Electrical Engineering, 75: pp. 46-60, 2019.
- [8] Douik, Ali; Abdellaoui, Mehrez; Kabbai, Leila. "Content based image retrieval using local and global features descriptor". In: 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE, pp. 151-154, 2016.
- [9] Bouchakwa, Mariam; Ayadi, Yassine; Amous, Ikram. "A review on visual content-based and users' tags-based image annotation: methods and techniques". Multimedia Tools And Applications, 2020.

- [10] Wang, Xinsheng, et al. "Unsupervised semantic-based convolutional features aggregation for image retrieval". *Multimedia Tools and Applications*, 79.21: pp.14465-14489, 2020.
- [11] Manzoor, Umar, et al., "Semantic image retrieval: An ontology based approach". *International Journal of Advanced Research in Artificial Intelligence (IJARAI)*, 1.4: pp. 1-8, 2015.
- [12] Jiu, Mingyuan; Sahbi, Hichem. "Nonlinear deep kernel learning for image annotation". *IEEE Transactions on Image Processing*, 26.4, pp. 1820-1832, 2017.
- [13] Cao, Yue, et al., "Deep visual-semantic hashing for cross-modal retrieval". In: *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 1445-1454, 2016.
- [14] Nguyễn Minh Hải, Lê Thị Vĩnh Thanh, Văn Thế Thành, Trần Văn Lăng, "Tra cứu ảnh theo ngữ nghĩa dựa trên cây phân cụm phân cấp", *Kỷ yếu Hội thảo Quốc gia về Nghiên cứu cơ bản và ứng dụng CNTT (FAIR), ĐH Huế, Nhà xuất bản Khoa học tự nhiên và Công nghệ*, ISBN: 978-604-913-915-4, pp. 502-511, 2019.
- [15] Alzubaidi, Mohammad A., "A new strategy for bridging the semantic gap in image retrieval". *International Journal of Computational Science and Engineering*, 14.1: pp. 27-43, 2017,.
- [16] Liu, Ying, et al., "A survey of content-based image retrieval with high-level semantics". *Pattern recognition*, 40.1: pp. 262-282, 2007.
- [17] WANG, Josiah, et al., "Automatic image annotation at ImageCLEF". In: *Information Retrieval Evaluation in a Changing World*. Springer, Cham, pp. 251-273. 2019.
- [18] Kanimozhi, T.; Christy, "A. Incorporating ontology and SPARQL for semantic image annotation". In: *2013 IEEE Conference on Information & Communication Technologies*. IEEE, pp. 26-31, 2013.
- [19] Kurtz, Camille, et al., "A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations". *Journal of biomedical informatics*, 49, pp. 227-244 2014.
- [20] Rejito, Juli, et al., "Image indexing using color histogram and k-means clustering for optimization CBIR in image database". *JPhCS*, 893.1: 012055, 2017.
- [21] Filali, Jalila; Zghal, Hajer; Martinet, Jean. "Towards visual vocabulary and ontology-based image retrieval system". 2016.
- [22] Ouhda, Mohamed, et al., "A content based image retrieval method based on K-means clustering technique". *Journal of Electronic Commerce in Organizations (JECO)*, 16.1: pp. 82-96, 2018.
- [23] Asim, Muhammad Nabeel, et al., "The use of ontology in retrieval: a study on textual, multilingual, and multimedia retrieval". *IEEE Access*, 7: pp.21662-21686, 2019.
- [24] Maur, Harleen Kaur; Faridkot, Punjab; Jain, Puneet. "Content based Image Retrieval System using K-Means Clustering Algorithm and SVM Classifier Technique", pp. 39-43, 2019.

IMAGE RETRIEVAL BASE ON CLUSTERING METHOD AND ONTOLOGY

Le Thi Vinh Thanh, Phan Thi Ngoc Mai, Van The Thanh, Le Manh Thanh

ABSTRACT: *Image semantic query is a problem that has many applications in different fields such as medicine, education, entertainment... In this paper, a model of image semantic search and lookup is based on proposed clustering and ontology methods. To do this, we gathered data clusters to search for similar sets of images, thereby extracting the semantics of these images on a dictionary structure and an already built ontology. For each input query image, visual characteristics are extracted to search semantic classes on clusters by frequency statistical method. The semantic similar image file is queried on Ontology and the semantics of the classification are accessed on the dictionary structure. Based on the proposed methods, a similar image search model and semantic extraction for the input image are built. The experimental results we evaluate on COREL, ImageCLEF series, and compare with the recently published works. According to the results of the proposed method, our image query model is feasible and effectively applied in semantic image query systems in many different fields.*

Keywords: *Semantics based image retrieval (SBIR), clustering, dictionary, similar image, ontology.*