DOI: 10.15625/vap.2020.00179

MỘT MÔ HÌNH TÌM KIẾM ẢNH THEO NGỮ NGHĨA DỰA TRÊN THUẬT TOÁN *k*-NN VÀ ĐẶC TRƯNG TÚI TỪ

Nguyễn Hải Yến[1], Nguyễn Thị Định[1], Nguyễn Văn Thịnh[1], Văn Thế Thành[2], Lê Mạnh Thạnh[3,*]

[1]Khoa Công nghệ thông tin, Trường Đại học Công nghiệp thực phẩm TP. HCM
[2]Phòng Quản lý khoa học và Đào tạo sau Đại học, Trường Đại học Công nghiệp thực phẩm TP. HCM
[3]Trường Đại học Khoa học, Đại học Huế
[1,2]{*yennh, dinhnt, thinhnv, thanhvt*}*@hufi.edu.vn*, [3]*lmthanh@hueuni.edu.vn*

TÓM TẮT: *Trong bài báo này, chúng tôi tiếp cận một mô hình truy vấn ảnh theo ngữ nghĩa dựa trên thuật toán láng giềng gần nhất k-NN (k-Nearest Neighbor) và đặc trưng túi từ BoW (Bag of Words). Các đặc trưng thị giác của hình ảnh được trích xuất và gom cụm để làm dữ liệu đầu vào cho quá trình phân lớp ngữ nghĩa hình ảnh và ánh xạ đến túi từ BoW đã xây dựng. Trên cơ sở đó, véctơ từ được trích xuất nhằm tạo câu truy vấn SPARQL để làm cơ sở cho quá trình truy vấn ảnh theo ngữ nghĩa dựa trên Ontology. Kết quả tìm kiếm là một tập các hình ảnh tương tự và ngữ nghĩa của phân lớp hình ảnh đầu vào. Để minh chứng cho cơ sở lý thuyết này, một mô hình tra cứu ảnh theo ngữ nghĩa được xây dựng và thực nghiệm trên bộ ảnh COREL, Wang, ImageCLEF; Kết quả thực nghiệm được đánh giá và so sánh với các công trình khác đã công bố gần đây. Theo kết quả thực nghiệm, phương pháp đề xuất của chúng tôi áp dụng tốt trong các hệ tìm kiếm dữ liệu đa phương tiện.*

Từ khóa: *Tìm kiếm ảnh theo ngữ nghĩa (Semantic Based Image Retrieval), k láng giềng gần nhất (k-Nearest Neighbor), túi từ (Bag of Words), ảnh tương tự (Similar Image), Ontology.*

I. GIỚI THIỆU

Theo số liệu thống kê của tập đoàn dữ liệu quốc tế IDC (International Data Corporation), năm 2018 dữ liệu toàn cầu khoảng 33 zettabyte (1 zettabyte = 1 nghìn tỷ gigabyte), ước tính đến năm 2025 có khoảng 175 zettabyte; trong đó, 90 zettabyte được tạo ra từ các thiết bị IoT, 49 % dữ liệu được lưu trữ trên môi trường đám mây, gần 30 % dữ liệu sẽ được sử dụng để xử lý theo thời gian thực [1, 2]. Thống kê dữ liệu ảnh số năm 2015, tổng số hình ảnh toàn cầu đạt 3,2 nghìn tỷ; năm 2016, có 3,5 triệu hình ảnh được chia sẻ trong mỗi phút và có 2,5 nghìn tỷ hình ảnh được chia sẻ và lưu trữ trực tuyến; năm 2017, tổng số hình ảnh toàn cầu lên đến 4,7 nghìn tỷ [3]. Ảnh số đã trở nên thân thuộc với cuộc sống của con người và được ứng dụng trong nhiều hệ thống tra cứu thông tin đa phương tiện như hệ thống thông tin bệnh viện (Hospital Information System), hệ thống thông tin địa lý (Geographic Information System), hệ thống thư viện số (Digital Library System), ứng dụng y sinh, ứng dụng trong giáo dục đào tạo, giải trí,… [4, 5].

Dữ liệu ảnh số ngày càng gia tăng, điều này đã mang lại nhiều thách thức và cơ hội cho lĩnh vực nghiên cứu về tra cứu ảnh. Các bài toán về tìm kiếm ảnh tương tự, phân lớp hình ảnh,… được ứng dụng trong nhiều hệ thống tra cứu, trong đó phân loại ngữ nghĩa hình ảnh là một trong những bài toán quan trọng của nhiều hệ thống đa phương tiện [6]. Những hệ thống tra cứu hình ảnh được phát triển như: hệ thống tra cứu ảnh dựa trên văn bản TBIR (Text Based Image Retrieval) [9], tra cứu ảnh dựa trên nội dung (CBIR - Content Based Image Retrieval) [10, 11]. Các hệ thống tra cứu ảnh này đa số dựa trên từ khóa, văn bản, nội dung trực quan và chưa phân tích ngữ nghĩa hình ảnh, do đó hiệu suất tìm kiếm chưa cao [7, 8]. Trong cách tiếp cận của chúng tôi, một kỹ thuật phân lớp *k*-NN kết hợp với đặc trưng túi từ được áp dụng cho bài toán tìm kiếm ảnh theo ngữ nghĩa; danh sách phân lớp ngữ nghĩa được trích xuất từ túi từ nhằm tạo câu truy vấn SPARQL và được thực thi trên một Ontology. Chúng tôi đề xuất một cấu trúc tra cứu chỉ mục phân lớp và Ontology mô tả ngữ nghĩa hình ảnh để tra cứu ngữ nghĩa của các phân lớp cho mỗi ảnh đầu vào nhằm tăng độ chính xác, tăng tốc độ tìm kiếm và giảm chi phí tính toán.

Đóng góp của bài báo gồm: (1) Cải tiến thuật toán *k*-NN nhằm tạo ra các phân lớp ngữ nghĩa cho hình ảnh; (2) Xây dựng cấu trúc túi từ thị giác để tìm kiếm hình ảnh tương tự; (3) Thiết kế mô hình tìm kiếm ảnh theo ngữ nghĩa dựa trên việc kết hợp thuật toán *k*-NN, đặc trưng túi từ BoW và tra cứu phân lớp trên một Ontology. (4) Xây dựng thực nghiệm và chứng minh tính đúng đắn của đề xuất trên một số bộ dữ liệu.

Phần còn lại của bài báo gồm: Phần II, chúng tôi khảo sát và phân tích ưu nhược điểm của một số công trình liên quan để chứng minh tính khả thi của bài toán phân lớp và tìm kiếm ảnh tương tự; Phần III, trình bày thuật toán phân lớp CkNN, thuật toán xây dựng túi từ thị giác CBW và phương pháp tìm kiếm ảnh tương tự dựa trên đặc trưng túi từ nhằm tạo câu truy vấn SPARQL và thực hiện trên một Ontology; Mô hình và thực nghiệm được mô tả trong phần IV, kết quả được đánh giá trên bộ dữ liệu ảnh COREL (1.000 ảnh), Wang (10.800 ảnh) ImageCLEF (có 20.000 ảnh); Phần V là kết luận và hướng phát triển tiếp theo.

II. CÁC CÔNG TRÌNH LIÊN QUAN

Cùng với sự phát triển của thiết bị công nghệ, sự gia tăng của dữ liệu ảnh số, đã thúc đẩy nhu cầu sử dụng các ứng dụng thông minh trong nhận diện, phân lớp và tra cứu nguồn gốc hình ảnh. Có nhiều phương pháp khác nhau để phân lớp và tìm kiếm ảnh tương tự, một số công trình sử dụng phương pháp phân lớp ảnh dựa trên thuật toán tìm kiếm láng giềng *k*-NN kết hợp với thuật toán gom cụm K-Means và túi từ thị giác nhằm phân lớp hình ảnh [12-18]. Một

cách tiếp cận khác đã tích hợp phân tích ngữ nghĩa và đặc trưng thị giác hình ảnh nhằm xây dựng hệ thống tra cứu ảnh dựa trên tri thức, kết hợp xây dựng Ontology để tra cứu hình ảnh theo ngữ nghĩa nhằm nâng cao hiệu suất truy vấn ảnh [19-23].

Theo Shen Xiaohui và cộng sự (2014), đã xây dựng độ đo tương tự dựa trên ràng buộc không gian giữa các đối tượng đặc trưng để từ đó thực hiện bài toán tìm kiếm ảnh. Trong phương pháp này, nhóm tác giả thực hiện việc kết hợp giữa phương pháp *k*-NN và túi từ thị giác để truy vấn ảnh. Trong túi từ thị giác, các hình ảnh được thống kê và gom nhóm theo kỹ thuật phân lớp *k*-NN để tạo ra nhóm các hình ảnh tương tự [12]. Trong bài báo này, các túi từ thị giác chứa đựng các hình ảnh dựa trên việc phân lớp *k*-NN trong dữ liệu ban đầu nhưng chưa xây dựng được trọng số của mỗi túi từ theo phân lớp các hình ảnh. Hơn nữa, thuật toán *k*-NN được thực hiện trên độ đo của đối tượng đặc trưng và chưa giải quyết việc phân lớp hình ảnh trong trường hợp số lượng phần tử trong các phân lớp bằng nhau.

Dawei Li và cộng sự (2015), xây dựng túi từ thị giác dựa trên lược đồ màu sắc và chọn những hình ảnh đưa vào túi từ dựa trên màu sắc của số lượng điểm ảnh. Với mỗi hình ảnh đầu vào được phân loại dựa trên túi từ thị giác và lấy các hình ảnh lân cận của các ảnh gần nhất trong túi từ để truy xuất tập ảnh tương tự trong dữ liệu ảnh ban đầu. Việc truy xuất tập ảnh tương tự được thực hiện bằng phương pháp phân lớp với thuật toán *k*-NN [13]. Trong phương pháp này, nhóm tác giả thực hiện hai pha của phương pháp *k*-NN kết hợp với túi từ nhưng vẫn chưa xây dựng được mối quan hệ giữa các túi từ và chưa cải tiến thuật toán *k*-NN.

Yanchun Ma và cộng sự (2019), đưa ra mô hình phân lớp theo thuật toán *k*-NN có trọng số (weight *k*-NN) kết hợp phương pháp phân biệt tuyến tính đa nhãn để phân lớp đối tượng dựa trên trọng số nhằm cải thiện độ chính xác trong việc tìm kiếm ảnh theo ngữ nghĩa [14]. Thực nghiệm cho thấy hiệu quả trên các tập dữ liệu lớn; tuy nhiên, trong phương pháp này tốn kém nhiều chi phí thời gian cho pha huấn luyện và gán nhãn cho hình ảnh, vẫn chưa xây dựng một cấu trúc tìm kiếm ảnh tương tự theo nội dung để tăng tính hiệu quả về thời gian.

Safia Jabeen và cộng sự (2018), xây dựng mô hình tìm kiếm ảnh dựa trên túi từ thị giác (Bag of Visual Words) bằng cách gom cụm các đặc trưng thị giác kết hợp với ngữ nghĩa của các bộ phân loại hình ảnh [16]. Tuy nhiên, việc gom cụm các đặc trưng thị giác cấp thấp có thể tạo ra các cụm gồm các hình ảnh có nhiều ngữ nghĩa khác nhau dẫn đến hiệu suất tìm kiếm ảnh theo ngữ nghĩa chưa cao. Do đó, cần kết hợp giữa đặc trưng cấp thấp và ngữ nghĩa cấp cao trong phân lớp hình ảnh. Cùng thời điểm này, Xiao Xie và cộng sự đã đề xuất phương pháp phân lớp các đặc trưng thị giác của hình ảnh dựa trên mạng CNN (convolutional neural network) và kết xuất các từ thị giác (semantic keywords) để tìm kiếm ảnh tương tự. Tuy nhiên, tác giả vẫn chưa xây dựng và thực hiện truy vấn trên Ontology nhằm xác định ngữ nghĩa cho hình ảnh [17].

Theo Shuang Jia và cộng sự (2020), kết hợp thuật toán gom cụm K-Means và túi từ thị giác để ứng dụng cho bài toán tìm kiếm tập ảnh tương tự, trong đó túi từ thị giác được xây dựng dựa trên việc gom nhóm các đặc trưng theo thị giác để hình thành các túi từ lưu trữ các hình ảnh. Ứng với mỗi hình ảnh đưa vào được trích xuất đặc trưng, tính độ tương tự với các túi từ gần nhất để trích xuất ra tập ảnh tương tự [18]. Tuy nhiên, trong phương pháp này các túi từ là độc lập và chưa phân lớp được nội dung của hình ảnh.

M. N. Asim và cộng sự (2019), đã thực hiện khảo sát 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. Nhóm tác giả đã so sánh hiệu suất với các phương pháp tiếp cận trước đó về truy vấn văn bản, dữ liệu đa phương tiện 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]. Tuy nhiên, 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.

Từ các công trình nghiên cứu cho thấy, phương pháp tìm kiếm ảnh tương tự dựa trên kỹ thuật phân lớp *k*-NN kết hợp với đặc trưng túi từ ứng dụng cho bài toán tìm kiếm ảnh là hoàn toàn khả thi. Trong bài báo này, chúng tôi đề xuất một tiếp cận mới dựa trên kỹ thuật *k*-NN cải tiến, đặc trưng túi từ và Ontology để phân lớp đồng thời tìm kiếm một tập ảnh tương tự theo ngữ nghĩa. Trong mô hình túi từ, các đặc trưng hình ảnh được lưu trữ cùng với phân lớp của hình ảnh và liên kết với các túi từ khác dựa trên trọng số tỉ lệ giữa các phân lớp ưu thế. Sau đó, với mỗi hình ảnh đầu vào được phân lớp bằng kỹ thuật *k*-NN dựa trên *k* láng giềng gần nhất và bán kính θ.

III. PHƯƠNG PHÁP TÌM KIẾM ẢNH TƯƠNG TỰ THEO THUẬT TOÁN *k*-NN VÀ ĐẶC TRƯNG TÚI TỪ

A. Thuật toán k-NN cải tiến

Thuật toán k-NN thực hiện phân lớp một ảnh đầu vào *I* dựa trên tập huấn luyện bằng cách so sánh khoảng cách euclide của ảnh *I* với tất cả các ảnh trong tập huấn luyện, sau đó sắp xếp các khoảng cách này theo thứ tự tăng dần, lấy *k* láng giềng gần nhất và thực hiện lấy nhãn của ảnh có tần số suất hiện nhiều nhất trong số *k* láng giềng để gán nhãn cho ảnh *I*. Tuy nhiên, việc phân lớp này gặp khó khăn trong các trường hợp nhiễu khi chọn *k* láng giềng nhỏ và độ phức tạp của thuật toán tăng lên đáng kể khi so sánh khoảng cách euclide của ảnh cần phân lớp với tất cả các ảnh trong tập dữ liệu huấn luyện khá lớn. Để giải quyết vấn đề này, chúng tôi thực hiện cải tiến thuật toán k-NN thành thuật toán CkNN gồm (1) thực hiện gom cụm tập dữ liệu ban đầu bằng thuật toán K-Means và tính khoảng cách euclide từ ảnh I đến các tâm cụm $f_{Ci},(i=1..m)$; (2) sử dụng bán kính θ để bổ trợ cho *k* láng giềng trong trường hợp tần suất các

phân lớp bằng nhau. Gọi *m* là số phân lớp của tập dữ liệu ảnh, thuật toán CkNN thực hiện phân lớp cho một ảnh đầu vào *I* bất kỳ như sau:

Thuật toán CkNN

Đầu vào: Véctơ đặc trưng f_I, tập véctơ F, *k*, bán kính θ

Đầu ra: Phân lớp C_I của ảnh *I*

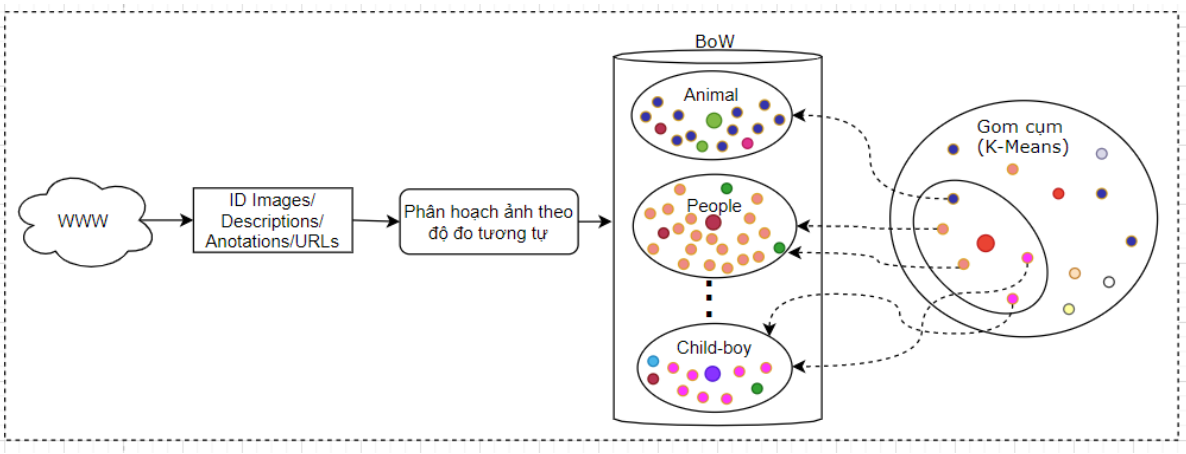
Funtion CkNN (f_I, F, k, θ)
Begin
Foreach ($I \in \{G\}$ **) do**
f_I = *ExtractFeature*(*I*);
Foreach ($f_I \in F$ **) do**
Foreach ($f_{Ci} \in \{f_C\}$ **) do**
{
$d(f_I, f_{Ci}) = euclide(f_I, f_{Ci})$;
$d_{\min} = \min\{d(f_{I,} f_{C1}), ..., d(f_{I,} f_{Cm})\}$;
}
EndForeach
EndForeach
If ($d_{\min} = d(f_{I,} f_{Ci})$ **then** $C_{\min} = C_i$;
kNN:
WordVector = *get* $(f_I, C_{\min}, \theta, k)$;
FreWord = *getFre* $(WordVector, C_{\min}, k)$;
ClassMax = *MaxClass* (*Wordvector, FreWord*);
isTrue = *check* (*FreWord*);
If (isTrue = true) **then**
Return ClassMax;
Else
{
$\theta = \theta + \varepsilon$;
Goto kNN;
}
EndIf
End.

Mệnh đề 1: Độ phức tạp của thuật toán **CkNN** là $O(m*n)$. Với *n* là số phần tử trong tập véctơ đặc trưng F, *m* là số cụm.

Chứng minh: Gọi *n* là số véctơ đặc trưng trong bộ dữ liệu ảnh F. Với mỗi véctơ đặc trưng $f_I \in F$ thuộc bộ dữ liệu, thuật toán tiến hành đo khoảng cách Euclide giữa f_I đến *m* tâm cụm $f_{Ci}, (i = 1..m)$ để tìm ra phân lớp cho ảnh đầu vào. Vì vậy độ phức tạp là O(m*n).

Thuật toán xây dựng túi từ BoW

Mô hình xây dựng túi từ BoW được minh họa như hình 1, với mỗi ảnh đầu vào véctơ đặc trưng được trích xuất và tìm tâm cụm gần nhất để thực hiện quá trình phân lớp dựa trên *k* láng giềng gần nhất. Sau khi tìm được cụm gần nhất, *k* phần tử láng giềng của ảnh đầu vào tại cụm đó được trích xuất nhằm làm đầu vào cho thuật toán CkNN. Các phần tử láng giềng này được liên kết với một *túi từ* và truy xuất ra một danh sách các lớp (gọi là véctơ từ) và tần suất các lớp (gọi là véctơ tần suất), trên cơ sở đó ảnh truy vấn đầu vào được phân loại về các lớp có tần suất nhiều nhất. Việc kết hợp *túi từ* và phương pháp phân cụm, phân lớp CkNN nhằm ánh xạ một ảnh đầu vào trở thành các phân lớp để từ đó tạo câu SPARQL để truy vấn trên Ontology nhằm truy vấn ngữ nghĩa và trích xuất tập ảnh tương tự.



Hình 1. Một minh họa xây dựng túi từ BoW

Thuật toán CBW

Đầu vào: Tập dữ liệu ảnh $f_j \in F$ được gán nhãn, số tâm cụm m

Đầu ra: Tập túi từ Ω.

Begin

Int *numbag* = k;
For *i* = *1* to *numbag* **do**
$\Omega_i = \emptyset$;
EndFor
Foreach ($f_i \in Cluster_{i.center}$) **do**
$\Omega_{i.center} = f_i$;
Foreach ($f_j \in F$) **do**
$D_{ji} = Euclide(f_j, \Omega_{i.center})$;
EndForeach
$D_{\min} = \min\{D_{ji}\}$;
If ($D_{\min} = D_{ji}$) **then**
$\Omega_i = \Omega_i \cup \{f_j\}$;
$\Omega_{i.center} = update(\Omega_{i.center})$;
EndForeach
Foreach ($f_K \in \Omega_i$) **do**
$f_K.Lable = getLable(f_K)$;
$\Omega_{i.VisualWord} = \{f_K.Lable\}$;
EndForeach;
Return Ω ;

End.

Mệnh đề 2: Độ phức tạp của thuật toán **CBW** là $O(n)$. Với n là là số véctơ đặc trưng bộ dữ liệu ảnh.

Chứng minh:

Gọi m là số túi từ cần xây dựng, tiến hành khởi tạo tâm cho từng túi nên độ phức tạp là $O(m)$. Gọi n là số véctơ đặc trưng trong bộ dữ liệu ảnh, với mỗi véctơ ảnh thuật toán tiến hành đo khoảng cách Euclide giữa nó với véctơ tâm của từng túi để tìm ra túi từ chứa ảnh đầu vào nên độ phức tạp là $O(n)$. Vậy độ phức tạp của thuật toán là $O(m*n)$ với m là hằng số, do đó độ phức tạp cần tìm là $O(n)$ ∎

B. Thuật toán trích xuất véctơ từ

Từ tập túi từ Ω đã xây dựng, chúng tôi đề xuất thuật toán trích xuất véctơ từ và tập ảnh tương tự của ảnh truy vấn để làm cơ sở cho truy vấn ảnh theo ngữ nghĩa. Với mỗi ảnh truy vấn I, thuật toán tìm kiếm tập ảnh tương tự dựa vào tập túi từ Ω và phân lớp C_I của ảnh truy vấn. Sau đó trích xuất véctơ từ W của ảnh truy vấn là *Label* của ảnh có tần suất xuất hiện nhiều nhất trong tập ảnh tương tự.

Thuật toán CWV

Đầu vào: Véctơ ảnh truy vấn f_I, tập túi từ Ω

Đầu ra: Tập ảnh tương tự Γ và véctơ từ thị giác W_I

Begin

$\Gamma = \varnothing$;

$W_I = \varnothing$;

$C_I = Classification(f_I)$;

Foreach ($\Omega_i \in \Omega$) **do**

If ($\Omega_i.VisualWord = C_I$) **then**

$\Gamma = \Gamma \cup \Omega_i$;

EndIf

EndForeach

Foreach ($J \in \Gamma$) **do**

Begin

L_J = getLabel(J);

$freq(L_J) = count(L_J)$;

$W_I = W_I \cup max(freq(L_J))$;

End

Return (Γ, W_I);

End.

C. Tạo câu truy vấn SPARQL

Dựa trên véctơ từ đã được trích xuất bằng **Thuật toán CWV**, câu truy vấn SPARQL được tạo ra để làm cơ sở truy vấn trên Ontology đã xây dựng nhằm tìm ra tập ảnh tương tự và ngữ nghĩa hình ảnh. Hình 2 mô tả cách thực hiện truy vấn bằng SPARQL được sinh ra từ véctơ từ thị giác của ảnh 31592.JPG.

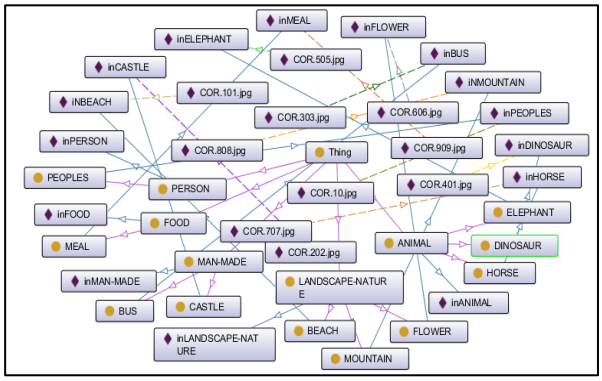
```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX sbir: <http://sbir-hcm.vn/>
SELECT DISTINCT ?Subject
WHERE{{?Subject sbir:opFLOWERBED sbir:inFLOWERBED .}{?Subject sbir:opGRASS sbir:inGRASS .}}
```

```
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:opFLOWERBED sbir:inFLOWERBED . sbir:inFLOWERBED rdf:type owl:NamedIndividual .
 sbir:inFLOWERBED rdf:type sbir:FLOWERBED . } {?IMG sbir:imgName ?ImgName. ?IMG sbir:opGRASS sbir:inGRASS . sbir:inGRASS rdf:type
owl:NamedIndividual . sbir:inGRASS rdf:type sbir:GRASS . }}
```

Hình 2. Một ví dụ thực hiện truy vấn bằng SPARQL

D. Xây dựng Ontology cho tập dữ liệu ảnh

Chúng tôi tạo một Ontology miền mô tả ngữ nghĩa cho bộ ảnh COREL gồm 10 phân lớp. Ontology xây dựng sử dụng ngôn ngữ bộ ba RDF dạng Turtle dựa trên ngữ nghĩa bộ ảnh COREL. Mỗi ảnh được thiết kế là một cá thể thuộc về một lớp đối tượng và được liên kết đến ngữ nghĩa mô tả tương ứng. Hình 3 mô tả mô hình Ontology trực quan được xây dựng trong Protege cho bộ ảnh COREL. Hình 4 minh họa Ontology được thực hiện cho bộ ảnh COREL dạng Turtle.



Hình 3. Một Ontology cho bộ ảnh COREL trên Protege

```
@prefix : <http://sbir-hcm.vn/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix sbir: <http://sbir-hcm.vn/> .
@base <http://sbir-hcm.vn/> .
<http://sbir-hcm.vn/> rdf:type owl:Ontology .
http://sbir-hcm.vn/opANIMAL
:opANIMAL rdf:type owl:ObjectProperty .
http://sbir-hcm.vn/opBEACH
:opBEACH rdf:type owl:ObjectProperty .
http://sbir-hcm.vn/opBUS
:opBUS rdf:type owl:ObjectProperty .
 http://sbir-hcm.vn/opCASTLE
:opCASTLE rdf:type owl:ObjectProperty .
```

Hình 4. Một ví dụ tạo Ontology dạng Turtle

E. Thuật toán tìm kiếm và trích xuất ngữ nghĩa hình ảnh dựa trên véctơ từ và Ontology

Với mỗi ảnh truy vấn I thực hiện trích xuất véctơ từ W_I để tạo câu truy vấn SPARQL. Việc tạo câu truy vấn SPARQL làm cơ sở cho truy vấn ảnh theo ngữ nghĩa để tìm ra tập ảnh tương tự và ngữ nghĩa của hình ảnh từ Ontology đã xây dựng.

Thuật toán KBIR

Đầu vào: Véctơ từ W_I của ảnh I

Đầu ra: Tập ảnh tương tự S_I và phân lớp ngữ nghĩa CS_I

Begin

$S_I = \varnothing$;

$CS_I = \varnothing$;

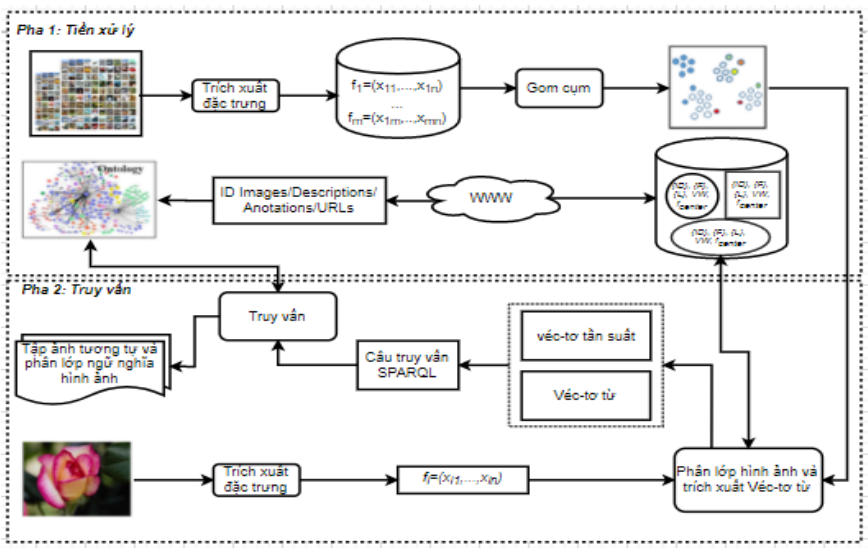
$SP(I) = CreateSPARQL(\mathrm{W}_I)$;

$(S_I, CS_I) = Query(SP(I), \mathrm{Ontology})$;

Return (S_I, CS_I);

End.

IV. MÔ HÌNH TÌM KIẾM ẢNH TƯƠNG TỰ THEO NGỮ NGHĨA



Hình 5. Mô hình tìm kiếm ảnh theo ngữ nghĩa dựa trên thuật toán *k*-NN và túi từ BoW

Mô hình thực nghiệm của hệ thống tìm kiếm ảnh theo nghữ nghĩa dựa trên thuật toán tìm kiếm láng giềng *k*-NN và túi từ được mô tả tại Hình 5. Mô hình tìm kiếm ảnh gồm hai pha: pha tiền xử lý và pha truy vấn.

Pha tiền xử lý: Thực hiện phân lớp tập dữ liệu ảnh, xây dựng túi từ và Ontology gồm các bước như sau:

***(1)*:** Trích xuất véctơ đặc trưng của tập dữ liệu ảnh.

***(2)*:** Gom cụm tập dữ liệu ảnh thành các cụm.

***(3)*:** Xây dựng túi từ thị giác cho tập dữ liệu ảnh.

***(4)*:** Xây dựng Ontology trên protege.

Pha truy vấn: Thực hiện tìm kiếm tập ảnh tương tự dựa trên túi từ đã xây dựng, trích xuất véctơ từ thị giác của ảnh truy vấn và thực hiện truy vấn theo ngữ nghĩa hình ảnh dựa trên Ontology đã xây dựng bằng SPARQL, gồm:

***(5)*:** Trích xuất véctơ đặc trưng ảnh cần truy vấn và thực hiện phân lớp hình ảnh đầu vào đồng thời trích xuất véctơ từ dựa trên túi từ thị giác đã xây dựng.

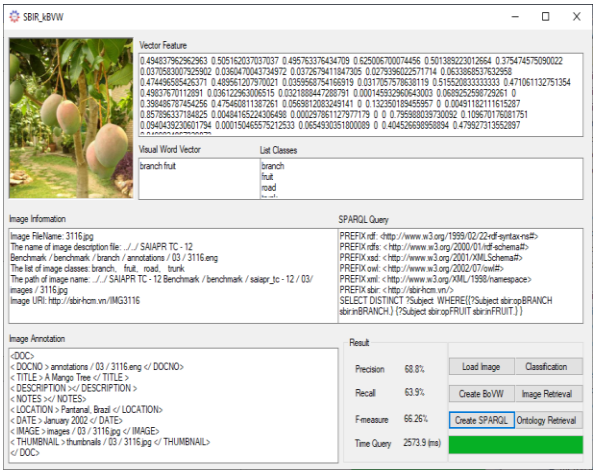
***(6)*:** Từ véctơ từ, tạo câu truy vấn SPARQL.

***(7)*:** Thực hiện truy vấn trên Ontology tìm kiếm tập ảnh tương tự theo ngữ nghĩa dựa vào câu SPARQL đã tạo.

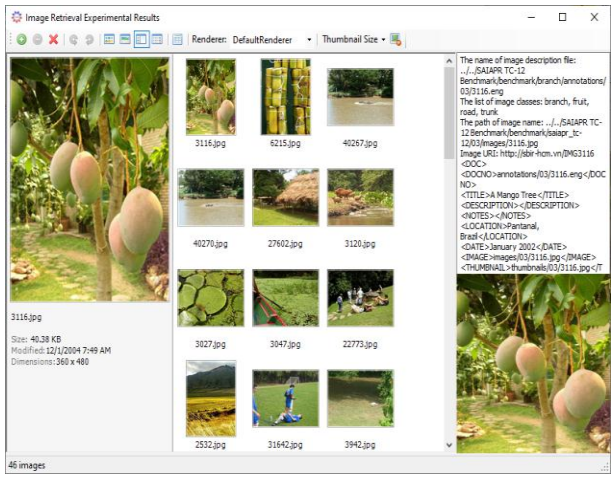
V. THỰC NGHIỆM

A. Môi trường xây dựng thực nghiệm

Thực nghiệm trích xuất đặc trưng và tìm kiếm ảnh tương tự **SBIR-kBVW** được xây dựng trên nền tảng dotNET Framework 4.5, ngôn ngữ lập trình C#. Các đồ thị được xây dựng trên Mathlab 2015. Cấu hình máy tính thực nghiệm: Intel(R) Core™ i5-5200U, CPU 2.2GHz, RAM 8GB và hệ điều hành Windows 10 Professional. Trong bài báo này, chúng tôi tiến hành thực nghiệm trên 3 bộ dữ liệu. Bộ ảnh COREL có 1000 ảnh được chia thành 10 chủ đề: *beach, bus, castle, dinosaur, elephant, flower, horse, meal, mountain, peoples*. Bộ ảnh Wang gồm 10.800 ảnh được chia thành 80 chủ đề và được chia thành 4 nhóm thực nghiệm: Nhóm 1 gồm các bộ ảnh từ 1 - 20 (*art_1 .. obj_cards*); nhóm 2 gồm các bộ từ 21 - 40 (*obj_decoys .. sc_autumn*); nhóm 3 gồm các bộ ảnh 41 - 60 (*sc_cloud .. wl_buttrfly*) và nhóm 4 gồm các bộ ảnh 61 - 80 (*wl_cat .. woman*). Bộ ảnh ImageCLEF gồm 20.000 ảnh được chia thành 41 chủ đề khác nhau (00 - 40).



Hình 6. Hệ truy vấn ảnh **SBIR-kBVW**



Hình 7. Một kết quả truy vấn trên **SBIR-kBVW**

Kết quả thực nghiệm của phương pháp đề xuất được trình bày trong Bảng 1, 2, 3.

Bảng 1. Hiệu suất tìm kiếm ảnh của phương pháp đề xuất trên bộ dữ liệu COREL

Tập ảnh	Độ chính xác trung bình	Độ phủ trung bình	Độ dung hòa trung bình
01-10	**0,756233**	**0,667691**	**0,703921**

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

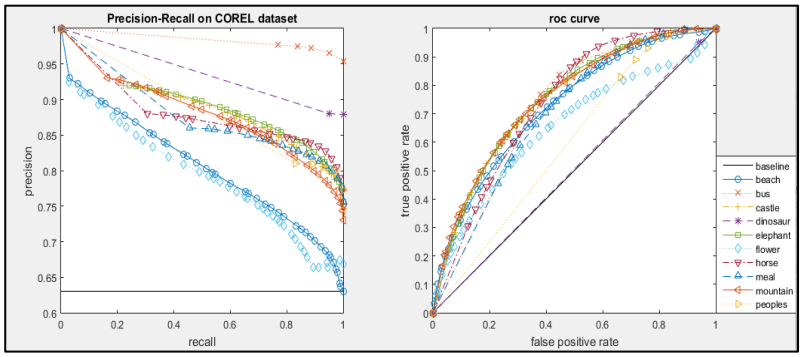
Tập ảnh	Độ chính xác trung bình	Độ phủ trung bình	Độ dung hòa trung bình
01-20	0,7382664	0,6624342	0,6976893
21-40	0,7712558	0,6925048	0,7292711
41-60	0,7331795	0,6588582	0,6938968
61-80	0,6920792	0,6211132	0,6541083
Trung bình	**0,733695**	**0,658728**	**0,693741**

Bảng 3. Hiệu suất tìm kiếm ảnh của phương pháp đề xuất trên bộ dữ liệu ImageCLEF

Tập ảnh	Độ chính xác trung bình	Độ phủ trung bình	Độ dung hòa trung bình
01-20	0,656621	0,443969	0,514976
21-40	0,703109	0,473882	0,549430
Trung bình	**0,678076**	**0,457775**	**0,530877**

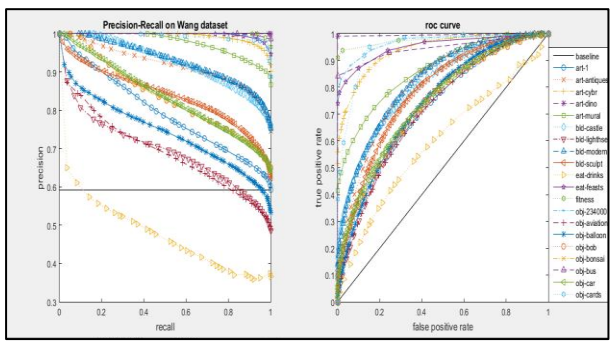
B. Đánh giá kết quả thực nghiệm

Mỗi đường cong trên đồ thị mô tả kết quả truy vấn độ chính xác (precision) và độ phủ (recall) từ một chủ đề ảnh trong bộ dữ liệu COREL, Wang, ImageCLEF. 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 được tính đúng đắn của các kết quả truy vấn. Hình 8, 9, 10, 11, 12, 13 mô tả hiệu suất và tính đúng đắn của kết quả truy vấn trên các bộ ảnh COREL, Wang và ImageCLEF.

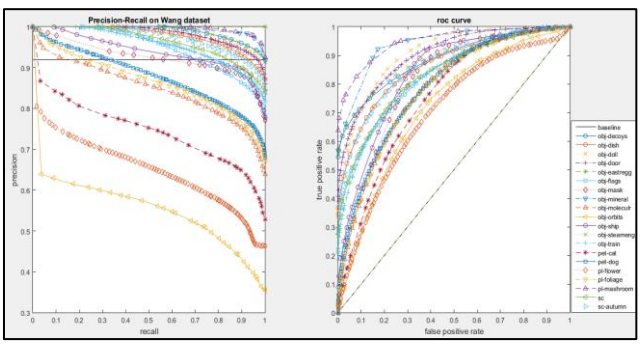


Hình 8. Precision-Recall và đường cong ROC của bộ dữ liệu ảnh COREL

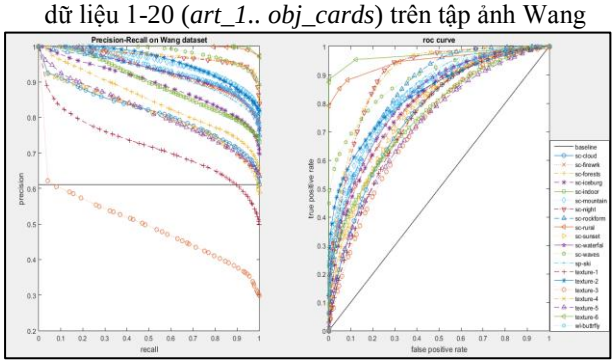
Hình 8 là đồ thị của giá trị Precision-Recall và đường cong ROC cho bộ dữ liệu COREL. Đồ thị cho thấy tính chính xác của hệ truy vấn nằm tập trung ở vùng [0,63; 1,0]. Đồ thị đường cong ROC biểu diễn các giá trị true positive và false positive theo độ phủ Recall, các giá trị nằm tập trung trên đường cơ sở (baseline), nhiều giá trị nằm trong vùng true positive.



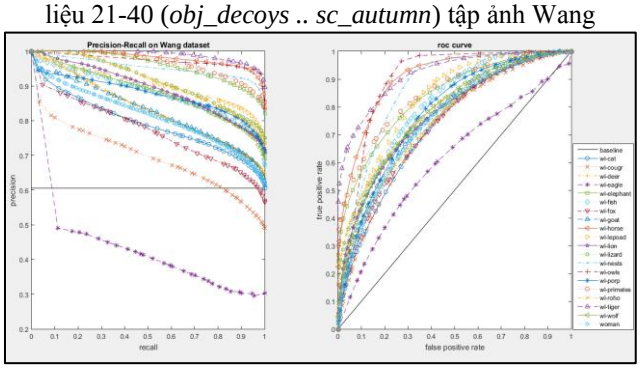
Hình 9. Precision-Recall và đường cong ROC của bộ dữ liệu 1-20 (*art_1.. obj_cards*) trên tập ảnh Wang



Hình 10. Precision-Recall và đường cong ROC của bộ dữ liệu 21-40 (*obj_decoys .. sc_autumn*) tập ảnh Wang

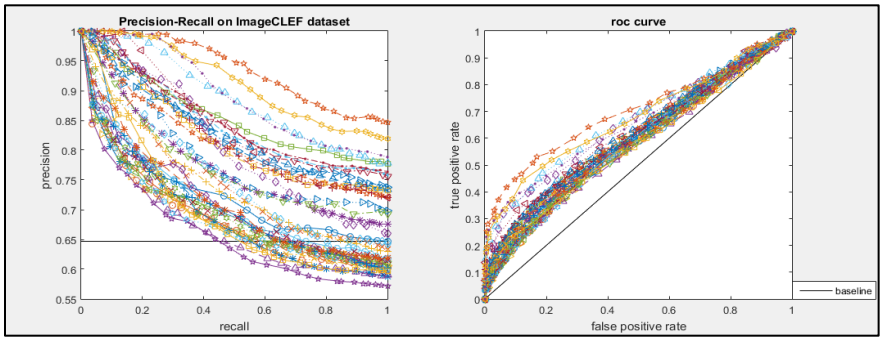


Hình 11. Precision-Recall, đường cong ROC bộ dữ liệu 41-60 (*sc_clound..wl_buttrfly*) trên bộ ảnh Wang



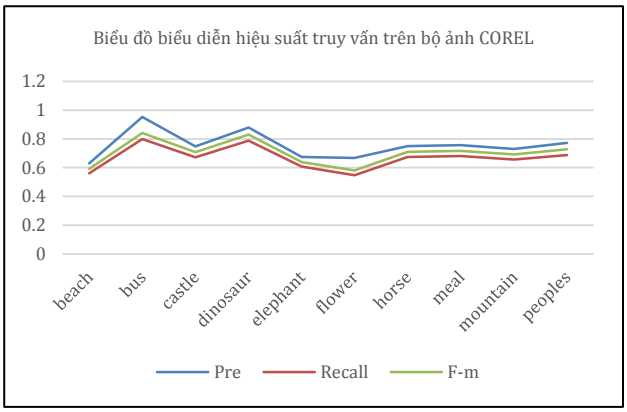
Hình 12. Precision-Recall và đường cong ROC của bộ dữ liệu 61-80 (*wl_cat .. woman*) trên tập ảnh Wang

Hình 9-12 là đồ thị của giá trị Precision-Recall và đường cong ROC cho bộ dữ liệu Wang. Đồ thị cho thấy tính chính xác của hệ truy vấn nằm tập trung ở vùng [0,5; 1,0], chỉ có vài bộ dữ liệu có độ chính xác trong vùng [0,35; 0,8]. Đồ thị đường cong ROC biểu diễn các giá trị true positive và false positive theo độ phủ Recall, các giá trị nằm tập trung trên đường cơ sở, nhiều giá trị nằm trong vùng true positive hơn vùng false positive.

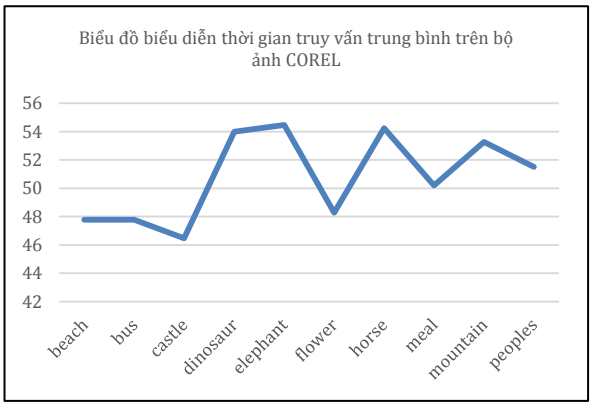


Hình 13. Precision-Recall và đường cong ROC của bộ dữ liệu ImageCLEF

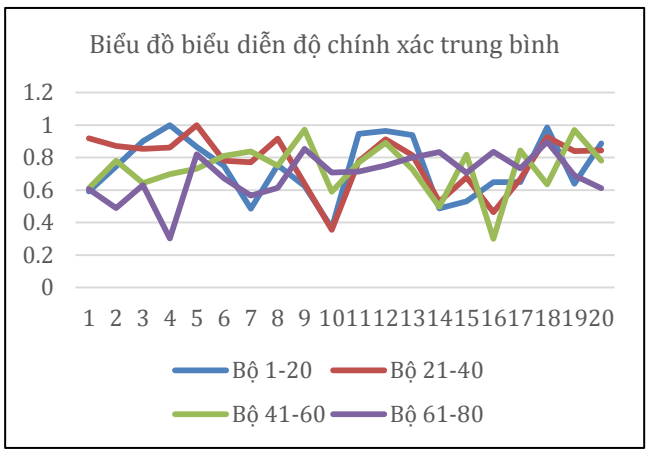
Hình 13 là đồ thị của giá trị Precision-Recall và đường cong ROC cho bộ dữ liệu ImageCLEF. Đồ thị cho thấy tính chính xác của hệ truy vấn nằm tập trung ở vùng [0,56; 1,0]. Đồ thị đường cong ROC biểu diễn các giá trị true positive và false positive theo độ phủ Recall, các giá trị nằm tập trung trên đường cơ sở.



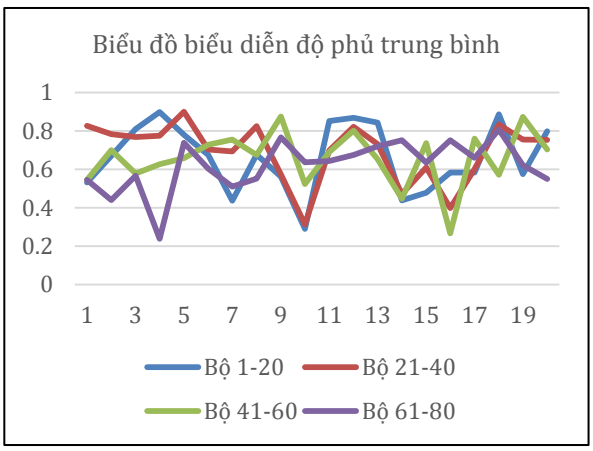
Hình 14. Hiệu suất truy vấn trung bình bộ ảnh COREL



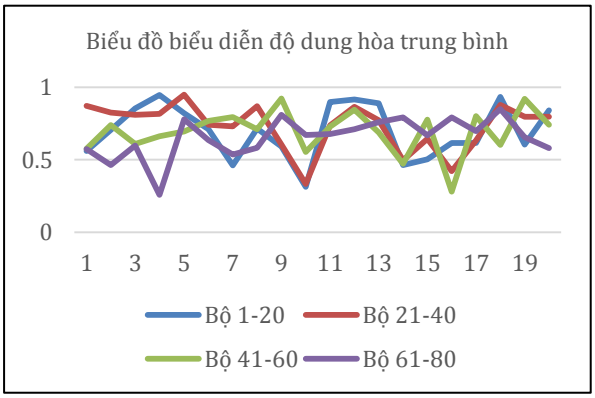
Hình 15. Thời gian truy vấn trung bình bộ ảnh COREL



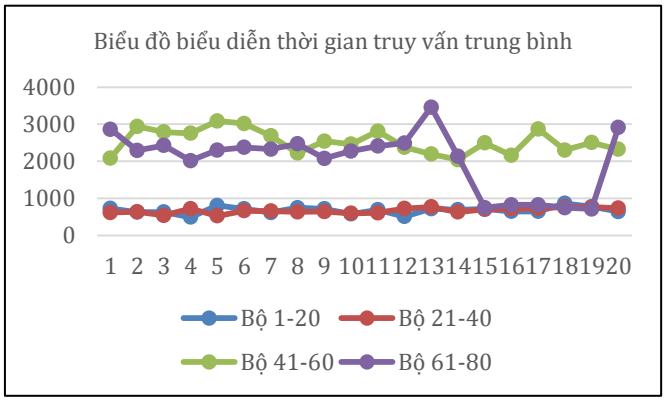
Hình 16. Hiệu suất trung bình trên bộ ảnh Wang



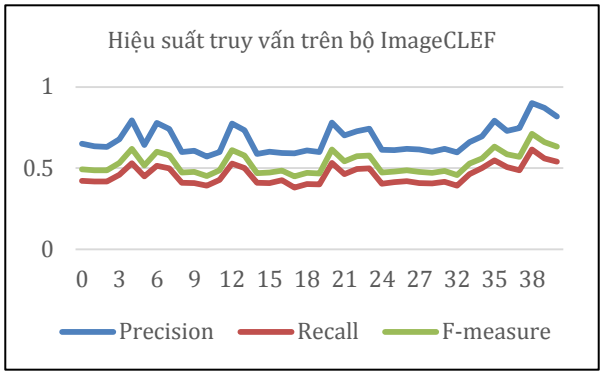
Hình 17. Độ phủ trung bình trên bộ ảnh Wang



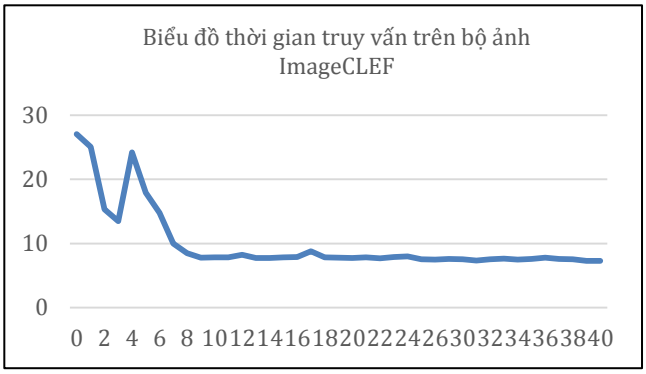
Hình 18. Hiệu suất trung bình trên bộ ảnh Wang



Hình 19. Thời gian truy vấn trung bình trên bộ ảnh Wang



Hình 20. Hiệu suất truy vấn trên bộ ảnh ImageCLEF



Hình 21. Thời gian truy vấn trên bộ dữ liệu ImageCLEF

Để minh chứng cho mô hình truy vấn ảnh theo ngữ nghĩa đề xuất là hiệu quả, chúng tôi so sánh kết quả thực nghiệm với một số công trình gần đây trên cùng bộ dữ liệu trong Bảng 4, 5, 6.

Bảng 4. So sánh hiệu suất truy vấn giữa các phương pháp trên bộ dữ liệu COREL

Phương pháp	Bộ dữ liệu	Độ chính xác trung bình
Relevance Feedback, 2004 [24]	COREL	68,0 %
B_SHIFT, 2016 [25]	COREL	72,0 %
Phương pháp đề xuất	**COREL**	**75,6 %**

Bảng 5. So sánh hiệu suất truy vấn giữa các phương pháp trên bộ dữ liệu Wang

Phương pháp	Bộ dữ liệu	Độ chính xác trung bình
CBIR, 2013 [26]	WANG	61,0 %
MLP, 2018 [27]	WANG	51,0 %
Phương pháp đề xuất	**WANG**	**73,4 %**

Bảng 6. So sánh hiệu suất truy vấn giữa các phương pháp trên bộ dữ liệu ImageCLEF

Phương pháp	Bộ dữ liệu	Độ chính xác trung bình
CBIR, 2013 [28]	ImageCLEF	46,78 %
MLP, 2018 [29]	ImageCLEF	46,18 %
Phương pháp đề xuất	**ImageCLEF**	**67,8 %**

VI. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong bài báo này, chúng tôi đã xây dựng một mô hình truy vấn ảnh tương tự và trích xuất ngữ nghĩa theo nội dung của hình ảnh. Trong mô hình này, chúng tôi đã kết hợp phương pháp gom cụm hình ảnh theo đặc trưng thị giác để xây dựng túi từ thị giác nhằm làm cơ sở cho việc tìm kiếm phân lớp của hình ảnh theo nội dung. Ngữ nghĩa hình ảnh được trích xuất dựa trên câu truy vấn SPARQL trên Ontology đã được chúng tôi xây dựng từ tập ảnh độc lập. Thực nghiệm được xây dựng trên các bộ ảnh COREL, Wang, ImageCLEF để minh chứng tính khả thi của mô hình mà chúng tôi đề xuất. Kết quả thực nghiệm được đánh giá dựa trên độ chính xác, độ phủ và độ chính xác dung hòa nhằm so sánh với các công trình đã công bố gần đây. Theo đó, độ chính xác trung bình tương ứng từng bộ ảnh lần lượt là: 75,6 %, 73,4 %, 67,8 % cho thấy mô hình đã đề xuất là hiệu quả và có thể áp dụng được cho các hệ thống tìm kiếm ảnh trên các lĩnh vực khác nhau. Hướng phát triển tiếp theo, chúng tôi sẽ xây dựng cấu trúc túi từ thị giác từ dữ liệu WWW dựa trên một cấu trúc dữ liệu phân lớp dưới dạng tiếp cận Kd-Tree.

VII. LỜI CẢM ƠN

Nhóm tác giả chân thành cảm ơn Trường Đại học Công nghiệp thực phẩm TP. HCM là nơi bảo trợ cho nghiên cứu này. Trân trọng cảm ơn nhóm nghiên cứu SBIR-HCM và Trường Đại học Sư phạm TP. HCM đã hỗ trợ về chuyên môn và cơ sở vật chất để nhóm tác giả hoàn thành nghiên cứu này.

TÀI LIỆU THAM KHẢO

[1] A Patrizio, "Data center explorer", Network World, 03/12/2018.
https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html.

[2] David Reinsel, John Gantz, John Rydning, "The Digitization of the World: From Edge to Core" sponsored by Seagate, IDC Technical Report, 2018. https://www.seagate.com/as/en/our-story/data-age-2025/.

[3] Deloitte, "Photo sharing: trillions and rising", Deloitte Touche Tohmatsu Limited, Deloitte Global, 2016.

[4] P. Muneesawang, N. Zhang, L. Guan, "Multimedia Database Retrieval: Technology and Applications", Springer, New York Dordrecht London, 2014.

[5] X. Xie, X. Cai, J. Zhou, N. Cao, Y. Wu, "A Semantic-based Method for Visualizing Large Image Collections", IEEE Transactions on Visualization and Computer Graphics, IEEE Computer Society, Vol. 25, No. 7, pp. 2362-2377, 2019.

[6] L. Deligiannidis, H. R. Arabnia, "Emerging Trends in Image Processing, Computer Vision, and Pattern Recognition", ed S. Elliot, Elsevier, USA: Morgan Kaufmann, Waltham, MA 02451, 2015.

[7] Ying Liu, Dengsheng Zhang, Guojun Lu, Wei-Ying Ma (2007), "A survey of content-based image retrieval with high-level semantics", Pattern Recognition Journal 40, pp. 262-283.

[8] Alzu'bi A, Amira A, Ramzan N, "Semantic content-based image retrieval: A comprehensive study", J Vis Commun Image Represent 32, pp. 20-54, 2015.

[9] I. K. Sethi, I. L. Coman, "Mining association rules between low-level image features and high-level concepts", Proceedings Volume 4384, Data Mining and Knowledge Discovery: Theory, Tools, and Technology III, 2001. https://doi.org/10.1117/12.421083.

[10] J. Eakins, M. Graham, "Content-based image retrieval", Technical Report, University of Northumbria at Newcastle, 1999.

[11] T. Gevers, A. Smeulders, "Content-based image retrieval by viewpoint-invariant color indexing", Image Vision Computing, Vol. 17, No. 7, pp. 475-488, 1999.

[12] Shen, Xiaohui, et al. "Spatially-constrained similarity measure for large-scale object retrieval", IEEE transactions on pattern analysis and machine intelligence 36.6 (2013): pp. 1229-1241.

[13] Li, Dawei, and Mooi Choo Chuah. "A Novel Unsupervised 2-Stage k-NN Re-Ranking Algorithm for Image Retrieval", 2015 IEEE International Symposium on Multimedia (ISM). IEEE, 2015.

[14] Ma, Yanchun, et al., "A weighted KNN-based automatic image annotation method", Neural Computing and Applications (2019): pp. 1-12.

[15] Kan, Shichao, et al. "A supervised learning to index model for approximate nearest neighbor image retrieval", Signal Processing: Image Communication 78 (2019): 494-502

[16] S. Jabeen, Z. Mehmood, T. Mahmood, T. Saba, A. Rehman, M. T. Mahmood, "An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-wordsmodel", PLoS ONE, Vol. 13, No. 4, pp. 1-24, 2018.

[17] X. Xie, X. Cai, J. Zhou, N. Cao, Y. Wu, "A Semantic-based Method for Visualizing Large Image Collections", IEEE Transactions on Visualization and computer graphics, IEEE Computer Society, Vol. xx, No. xx, pp.xx-xx, 2018.

[18] Jia, Shuang, et al., "Bag-of-Visual Words based Improved Image Retrieval Algorithm for Vision Indoor Positioning", 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring). IEEE, 2020.

[19] M.K., Y.I.A. and Mohd Noah, S.A. (2017), "Semantic text-based image retrieval with multi-modality Ontology and DBpedia", The Electronic Library, Vol. 35, No. 6, pp. 1191-1214. https://doi.org/10.1108/EL-06-2016-0127.

[20] Manzoor, Umar, et al. "Semantic image retrieval: An Ontology based approach." International Journal of Advanced Research in Artificial Intelligence (IJARAI) 1.4 (2015): pp. 1-8.

[21] V. Vijayarajan, M. Dinakaran, P. Tejaswin, M. Lohani, "A generic framework for Ontology-based information retrieval and image retrieval in web data", Human-centric Computing and Information Sciences, Vol. 6, No. 18, pp. 1-30, 2016.

[22] Spanier, Assaf B., D. Cohen, and Leo Joskowicz. "A new method for the automatic retrieval of medical cases based on the RadLex Ontology" International journal of computer assisted radiology and surgery 12.3 (2017): pp. 471-484.

[23] Asim, Muhammad Nabeel, et al. "The Use of Ontology in Retrieval: A Study on Textual, Multilingual, and Multimedia Retrieval" IEEE Access 7 (2019): 21662-21686.

[24] Gia, Giorgio, and Fabio Roli. "Instance-based relevance feedback for image retrieval." Advances in neural information processing systems. 2005.

[25] Douik, Ali, Mehrez Abdellaoui, and Leila Kabbai. "Content based image retrieval using local and global features descriptor", 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE, 2016.

[26] Lande, Milind V., Praveen Bhanodiya, and Pritesh Jain. "An effective content-based image retrieval using color, texture and shape feature". Intelligent Computing, Networking, and Informatics. Springer, New Delhi, 2014. pp. 1163-1170.

[27] Chhabra, Payal, Naresh Kumar Garg, and Munish Kumar. "Content-based image retrieval system using ORB and SIFT features". Neural Computing and Applications 32.7 (2020): pp. 2725-2733.

[28] M.E. Hakan Cevikalp, Savas Ozkan, "Large-scale image retrieval using transductive support vector machines", Computer Vision and Image Understanding, Vol. 173, No., pp. 2-12, 2018.

[29] M.D. V. Vijayarajan, P. Tejaswin, M. Lohani, "A generic framework for ontology-based information retrieval and image retrieval in web data", Human-centric Computing and Information Sciences, Vol. 6, No. 18, pp. 1-30, 2016.

A SEMANTIC-BASED IMAGE RETRIEVAL MODEL BASE ON k-NN ALGORITHM AND BAG OF WORD FEATURES

Nguyen Hai Yen, Nguyen Thi Dinh, Nguyen Van Thinh, Van The Thanh, Le Manh Thanh

ABSTRACT: *In this paper, we approach a semantic-based image retrieval model base on k-Nearest Neighbor algorithm (k-NN) and bag of visual word (BoVW). The visual features of image are extracted and clustered for input data of semantic classification process and mapping to BoW built. On that basis, the word vectors are extracted for creating SPARQL query as the basis for semantic image retrieval which based on the Ontology. The retrieval results are set of similar images and image classification semantics performed. To verifing for this theoretical, a semantic image retrival model is built and experimented on COREL, Wang, ImageCLEF dataset. Experimental results are evaluated compared to other recently published methods on the same dataset. According to the experimental result show that our proposed method is effective and can be applied in many multimedia data systems.*

Keywords: *SBIR, k-NN, Bag of Words, similar image, Ontology.*