**Trường Đại Học Công Nghệ Thông Tin - ĐHQG**

**KHOA: KHOA HỌC MÁY TÍNH**



ĐỒ ÁN MÔN HỌC

TRUY VẤN THÔNG TIN ĐA PHƯƠNG TIỆN

Đề tài:

**IMAGES SEARCH BASE ON CAPTION**

Giảng viên hướng dẫn **: Nguyễn Vinh Tiệp**

Sinh viên thực hiện : **Nguyễn Đức Hà (18520689)**

**Võ Gia Bảo (18520502)**

**Trần Lê Duy (18520674)**

**Nguyễn Trung Hiếu (18520751)**

Lớp : CS336.L12.KHCL

Khóa : 2018

***TP.Hồ Chí Minh, ngày 16 tháng 1 năm 2021***

**--------------------------------------**

**MỞ ĐẦU**

Tìm kiếm trong các bộ sưu tập hình ảnh dựa trên nội dung trực quan có khả năng phát triển rất mạnh ẽ về mặt kĩ thuật. Và một trong những thành tựu đó chính là Content-Based Image Retrieval, CBIR (hay còn gọi là truy xuất ảnh dựa trên nội dung).

Đây là 1 trong những đề tài đang được rất thu hút bởi các nhà nghiên cứu từ các lĩnh vực nghiên cứu khác nhau, bao gồm Thị giác máy tính (Computer Vision), Trí tuệ nhân tạo (Artificial Intelligence) và Học máy(Machine Learning).

Mục đích chính của bài báo cáo này là để giới thiệu ngắn gọn về CBIR, tài liệu và các ứng dụng. Báo cáo cung cấp tổng quan về các phương pháp hữu ích và

đưa ra một số thách thức lớn nhất trong CBIR.

**LỜI CẢM ƠN**

Để hoàn thành được đồ án môn học lần này 1 cách thành công tốt đẹp, em xin trân trọng cám ơn tập thể lớp và giảng viên thầy Nguyễn Vinh Tiệp (phụ trách bộ môn Truy vấn thông tin đa phương tiện, mã môn học: CS336.L12.KHCL) đã hỗ trợ nhóm em trong suốt thời gian học tập.

Do kiến thức còn nhiều hạn chế và khả năng tiếp thu thực tế còn nhiều bỡ ngỡ và chưa hoàn hảo nên báo cáo sẽ còn nhiều thiếu sót, kính mong sự góp ý và giúp đỡ từ thầy Nguyễn Vinh Tiệp.

Nhóm em xin chân thành cám ơn!

**NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN**

….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................…...............................................................................................................................................................................................................................................................................…...........................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................….....................................................................................................................................................................................................................................................…..............................................................................................................................................…..............................................................................................................................................…..............................................................................................................................................…..............................................................................................................................................…..............................................................................................................................................…..............................................................................................................................................…..............................................................................................................................................…..............................................................................................................................

**Nội dung**

[Phần 1: Giới thiệu: - 1 -](#_Toc61530854)

[Phần 2: Thiết kế và thực hiện: - 2 -](#_Toc61530855)

[2.1.Tổng quát: - 2 -](#_Toc61530856)

[2.2. Giải quyết bài toán: - 3 -](#_Toc61530857)

[2.2.1. TF – IDF: - 3 -](#_Toc61530858)

[2.2.2: Xử lý dữ liệu với TF-IDF: - 4 -](#_Toc61530859)

[2.2.3: Truy vấn ảnh với TF-IDF: - 5 -](#_Toc61530860)

[Phần 3: Thực nghiệm: - 6 -](#_Toc61530861)

[Phần 4: Hệ thống CBIR hiện có: - 8 -](#_Toc61530862)

[KẾT LUẬN - 9 -](#_Toc61530863)

# **Phần 1: Giới thiệu:**

Công nghệ hiện đại đã dẫn đến sự phát triển nhanh chóng của các bộ sưu tập phương tiện kỹ thuật số, thường chứa cả hình ảnh tĩnh và video. Thiết bị lưu trữ chứa đầy hàng terabyte kỹ thuật số hình ảnh1 (digital images1), khiến việc truy xuất hình ảnh quan tâm từ các bộ sưu tập như vậy ngày càng khó hơn.

Rõ ràng là khả năng tìm kiếm là cần thiết để tìm thấy những gì chúng ta đang tìm kiếm trong bộ sưu tập lớn, nhưng làm thế nào chúng ta có thể làm cho những tìm kiếm như vậy hữu ích?

Chú thích thủ công của hình ảnh có từ khóa mô tả nội dung hình ảnh có thể giúp bạn dễ dàng tìm thấy hình ảnh của lãi suất, nhưng điều này mất rất nhiều thời gian, làm cho phương pháp này rất tốn kém. Nó cũng chỉ hữu ích ở một mức độ nhất định, vì chúng ta không phải lúc nào cũng biết trước loại tìm kiếm nào điều đó sẽ được thực hiện trong tương lai. Vì tất cả những lý do trên, việc sử dụng truy xuất hình ảnh dựa trên nội dung (CBIR) đã được đề xuất.

Tìm kiếm trong các bộ sưu tập hình ảnh dựa trên nội dung trực quan có khả năng rất mạnh mẽ về mặt kỹ thuật. Vì vậy, một số mục tiêu của báo cáo này là để có được cái nhìn tổng quan về tính năng hiện đại của CBIR là gì, loại hệ thống nào có sẵn trên thị trường thương mại hoặc như các công cụ nguồn mở đã có ngày nay, những gì hiện tại chính những hạn chế của CBIR là gì và loại chức năng nào có thể được mong đợi trong tương lai (được định nghĩa là một vài năm kể từ hiện tại).

# **Phần 2: Thiết kế và thực hiện:**

## **2.1.Tổng quát:**

CBIR đã thu hút các nhà nghiên cứu từ nhiều lĩnh vực nghiên cứu bao gồm cả thị giác máy tính, trí tuệ nhân tạo, nhân tố con người và máy học. Mặc dù vậy, độ tuổi tương đối trẻ của CBIR như một hiện tượng và lĩnh vực nghiên cứu, đã có một sự phát triển rất lớn của các bài báo nghiên cứu về chủ đề này.

Một nỗ lực khá sớm để cấu trúc và đánh giá phần lớn công việc ban đầu về CBIR (từ đầu những năm 90 cho đến năm 2000) được tìm thấy trong Smeulders et al. (2000).

Phần lớn những phát hiện của họ hiện đã lỗi thời, nhưng hai khía cạnh rất quan trọng của CBIR được nêu bật trong bài báo của họ vẫn rất có liên quan ngày nay: Khoảng cách cảm quan và Khoảng cách ngữ nghĩa.

* Khoảng cách cảm quan: là sự khác biệt không thể tránh khỏi giữa đối tượng thực và thông tin người quan sát có thể biên dịch về đối tượng. Khoảng cách này không bao giờ có thể được đóng lại hoàn toàn nhưng nó có thể ở một mức độ ngày càng tăng được thu hẹp bởi các cảm biến được cải tiến, xử lý tín hiệu và Tổng hợp cảm biến (sensor fusion) ...
* Khoảng cách ngữ nghĩa được định nghĩa là “Thiếu sự trùng hợp giữa thông tin người ta có thể trích xuất từ dữ liệu trực quan (bằng các phương tiện tính toán) và giải thích rằng dữ liệu tương tự cho người dùng trong một tình huống nhất định” trong Smeulders et al. (2000).

Mặc dù có sự đơn giản rõ ràng của điều này, nhưng có những trở ngại đáng kể cần phải khắc phục để tạo ra một hệ thống CBIR hữu ích. Để tìm một chữ ký hình ảnh tốt bằng cách sử dụng các tính năng hình ảnh phù hợp còn lâu mới dễ dàng.

## **2.2. Giải quyết bài toán:**

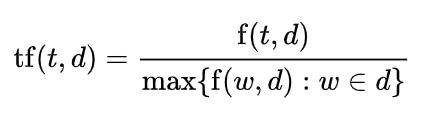
### 2.2.1. TF – IDF:

**TF-IDF** viết tắt là: Term Frequency – Inverse Document Frequency

Term Frequency – Inverse Document Frequency là 1 kĩ thuật sử dụng trong khai phá dữ liệu văn bản. Trọng số này được sử dụng để đánh giá tầm quan trọng của một từ trong một văn bản. Giá trị cao thể hiện độ quan trọng cao và nó phụ thuộc vào số lần từ xuất hiện trong văn bản nhưng bù lại bởi tần suất của từ đó trong tập dữ liệu. Một vài biến thể của tf-idf thường được sử dụng trong các hệ thống tìm kiếm như một công cụ chính để đánh giá và sắp xếp văn bản dựa vào truy vấn của người dùng.  Tf-idf cũng được sử dụng để lọc những từ stopwords trong các bài toán như tóm tắt văn bản và phân loại văn bản.

Sau đây, chúng ta sẽ đi chi tiết vào từng thành phần. Và sau đó, chúng ta đi đến công thức tính TF-IDF.

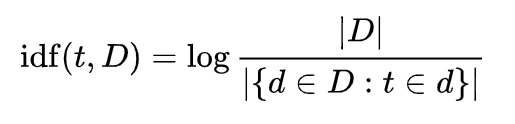
**TF**: Term Frequency (Tần suất xuất hiện của từ) là số lần từ xuất hiện trong văn bản. Vì các văn bản có thể có độ dài ngắn khác nhau nên một số từ có thể xuất hiện nhiều lần trong một văn bản dài hơn là một văn bản ngắn. Như vậy, term frequency thường được chia cho độ dài văn bản (tổng số từ trong một văn bản).



Trong đó:

* Tf (t, d): tần suất xuất hiện của từ t trong văn bản d
* F (t, d): Số lần xuất hiện của từ t trong văn bản d
* Max ({f (w, d): w ∈ d}): Số lần xuất hiện của từ có số lần xuất hiện nhiều nhất trong văn bản d

**IDF**: Inverse Document Frequency (Nghịch đảo tần suất của văn bản), giúp đánh giá tầm quan trọng của một từ. Khi tính toán TF, tất cả các từ được coi như có độ quan trọng bằng nhau. Nhưng một số từ như “is”, “of” và “that” thường xuất hiện rất nhiều lần nhưng độ quan trọng là không cao. Như thế chúng ta cần giảm độ quan trọng của những từ này xuống.



Trong đó:

* IDF (t, D): giá trị idf của từ t trong tập văn bản
* |D|: Tổng số văn bản trong tập D
* | {d ∈ D: t ∈ d} |: thể hiện số văn bản trong tập D có chứa từ t.
* Cơ số logarit trong công thức này không thay đổi giá trị idf của từ mà chỉ thu hẹp khoảng giá trị của từ đó. Vì thay đổi cơ số sẽ dẫn đến việc giá trị của các từ thay đổi bởi một số nhất định và tỷ lệ giữa các trọng lượng với nhau sẽ không thay đổi. (nói cách khác, thay đổi cơ số sẽ không ảnh hưởng đến tỷ lệ giữa các giá trị IDF). Việc sử dụng logarit nhằm giúp giá trị tf-idf của một từ nhỏ hơn, do chúng ta có công thức tính tf-idf của một từ trong 1 văn bản là tích của tf và idf của từ đó.
* Cụ thể, chúng ta có **công thức tính TF-IDF** hoàn chỉnh như sau: **TF-IDF (t, d, D) = TF (t, d) x IDF (t, D)**

Tổng quát lại: Những từ có giá trị TF-IDF cao là những từ xuất hiện nhiều trong văn bản này, và xuất hiện ít trong các văn bản khác. Việc này giúp lọc ra những từ phổ biến và giữ lại những từ có giá trị cao (từ khoá của văn bản đó).

### 2.2.2: Xử lý dữ liệu với TF-IDF:

Nhóm sử dụng dữ liệu có sẵn trên Kaggle với bộ dữ liệu gồm 8000 ảnh và tương ứng với mỗi ảnh là 5 câu mô tả đi kèm theo.

Đầu tiên, móc nối 5 câu mô tả của từng ảnh lại với nhau, lưu chúng vào trong 1 tập Dict với Key là tên của từng ảnh. Bắt đầu xử lý chúng: Tách từ, loại bỏ stopword, tính TF-IDF cho từng văn bản ứng với mỗi ảnh.

Bước cuối cùng là tạo vector trọng số của từng ảnh và lưu vào Database.

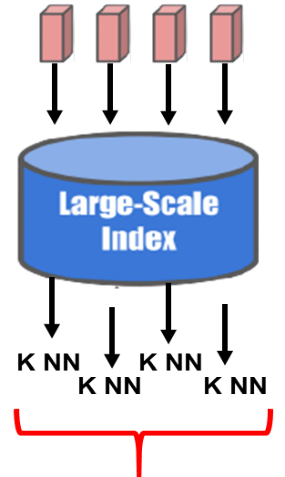
### 2.2.3: Truy vấn ảnh với TF-IDF:

Bằng cách sử dụng TF-IDF, xử lý dữ liệu và đưa vào trong database, tiếp theo sẽ là bước truy vấn hình ảnh trong database dựa trên những đặc trưng này. Các bước thực hiện như sau: Đánh chỉ mục cho từng vecto trọng số của ảnh trong database, sử dụng TF-IDF xử lý câu mô tả đưa vào truy vấn, tạo vecto trọng số cho câu truy vấn, tính độ tương đồng của câu truy vấn và các mô tả đặc trưng của ảnh trong database, xếp hạng, đưa ra các ảnh cần truy vấn.

Cấu trúc đánh chỉ mục trong database

|  |  |  |  |
| --- | --- | --- | --- |
| **Database Image Index** | **Description** |  | **Vector** |
|  | Decriptor 0 |  |  |
| 0 | Decriptor 1 |  | Vector 1 |
|  | **…** |  |  |
|  | Decriptor 5 |  |  |
|  | Decriptor 1 |  | Vector 2 |
| 1 | Decriptor 2 |  |
|  | **…** |  |
|  | Decriptor 5 |  |
| **…** | **…** |  |  |
|  | Decriptor 1 |  |  |
| 49 | Decriptor 2 |  | Vector 3 |
|  | **…** |  |
|  | Decriptor 5 |  |  |

Sau khi database đã được xây dựng. Tiếp đến là bước sử dụng TF-IDF để tạo vector trọng số cho câu truy vấn.



***Số lượng Vector tương đồng trong từng ảng thuộc database***

Sau đó, đem vector trọng số của câu truy vấn đem đi so sánh với tất cả các vector trọng số của tất cả văn bảntrong database, mục tiêu là tìm được các vector có độ tương đồng nhất định với đặc trưng của câu truy vấn. Việc tìm kiếm các đặc trưng này dựa trên thuật toán tìm kiếm các neighbor gần nó nhất (Nearest Neighbor search).

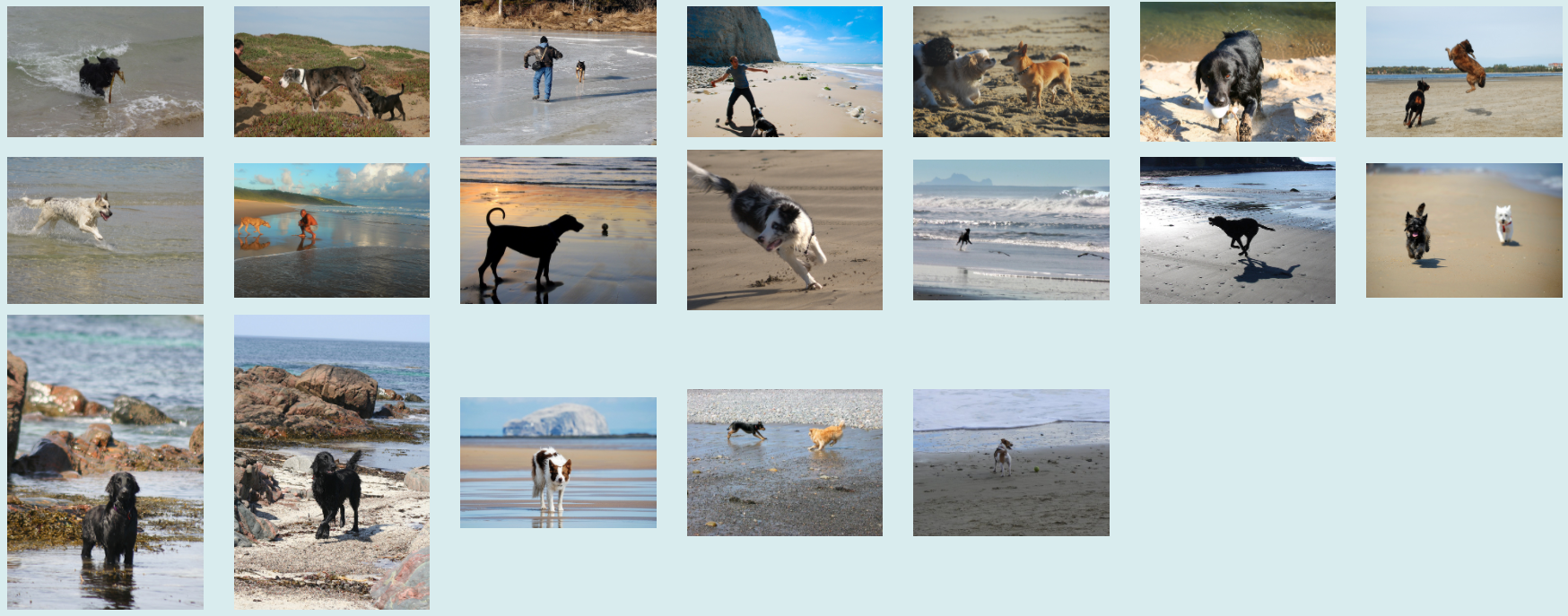
Một khi Vector trọng số câu truy vấn được đưa vào, có thể dễ dàng tìm được K vector gần nó nhất. K là tự chọn. Độ đo được sử dụng để tính khoảng cách hay độ tương đồng giữa 2 đặc trưng là Euclid. Cuối cùng trong bước truy vấn là sẽ tổng hợp các vector trọng số của tập văn bản mô tả để tìm ra K ảnh có độ tương đồng gần nhất.

Sau khi tìm ra số bức ảnh có Vector TF-IDF gần với câu truy vấn nhất, tuy nhiên để tăng độ chính xác nên cần phải thực hiện bước cuối cùng là Ranking.

# **Phần 3: Thực nghiệm:**

Sau khi thực hiện cài đặt hệ thống truy vấn thì thu được kết quả như sau:

Dog on beach



Child on grass



Nhận xét: với hệ thống truy vấn mà nhóm đã cài đặt thì nhiều câu truy vấn vẫn còn có độ sai sót khá cao, cụ thể như những câu truy vấn có ít thành phần mô tả, tuy nhiên còn phụ thuộc vào Dataset, chúng em sẽ cố gắng cải thiện chúng để cho hệ thống hoạt động tốt hơn thế nữa. Ngoài ra, hệ thống truy vấn của nhóm cho kết quả truy vấn vẫn còn khá chậm vì phải xử lý lại tất cả các bước mỗi khi reset lại web, khoảng 20s cho lần đầu truy vấn, những lần sau 10 – 13s.

# **Phần 4: Hệ thống CBIR hiện có:**

Trong những năm qua, nhiều hệ thống CBIR đã được phát triển. Hầu hết trong số này là nguyên mẫu và do đó không dành cho việc sử dụng công cộng. Tuy nhiên, một số được phát hành dưới dạng mở giấy phép nguồn để những người khác có thể sử dụng và phát triển thêm. Ngoài ra còn có một

số lượng hệ thống thương mại trên thị trường, trong đó khả năng của CBIR đã được kết hợp.

Và đây là 2 trong số hệ thống CBIR mà nhóm nghiên cứu đó là:

1. **IBM’s QBIC (Query By Image Content):**

* Là một trong những hệ thống CBIR nổi tiếng nhất và có lịch sử lâu đời. Công cụ CBIR được sử dụng trong QBIC, đã được sử dụng để tìm kiếm trong các kho lưu trữ nghệ thuật nổi tiếng thế giới tại trang website Hermitage.
* Ngày nay chức năng QBIC đã được tích hợp vào phần mềm IBM dành cho Quản lý Nội dung Doanh nghiệp (Enterprise Content Management), nhưng thật khó để tìm thấy bất kỳ thông tin hữu ích nào về điều này. Do đó vẫn chưa biết chức năng của QBIC ngày nay trông như thế nào.

1. **Virage**:

* Là một phần của Autonomy (được công ty Hewlett Packard – HP mua lại).
* Có nhiều sản phẩm Autonomy Virage khác nhau để xử lý hình ảnh như nhận dạng biển số xe, nhận dạng khuôn mặt (face recognition) và phân tích bối cảnh thông minh (intelligent scene analysis).
* Hệ thống CBIR có có thể được tích hợp vào một hoặc nhiều sản phẩm này, nhưng nó không hiển thị rõ tràng trên trang web.

# **KẾT LUẬN**

Trong bài báo cáo kì này, chúng em xin phép giới thiệu hệ thống Truy vấn dựa trên câu mô tả để truy xuất thông tin.

Nhờ đó, có lẽ đã giúp chúng em nắm thêm những kiến thức về việc truy vấn thông tin sao cho có độ chính xác cao.

Vì đây là lần đầu chúng em làm 1 bài báo cáo đầy đủ nên có điều gì sai sót và chưa hoàn chỉnh mong thầy góp ý để chúng em rút kinh nghiệm cho những bài báo cáo sau.

**DANH MỤC TÀI LIỆU THAM KHẢO**

1. Smeulders A.W.M. et al., "Content-Based Image Retrieval at the End of the Early Years" IEEE Trans. PAMI, vol 22, no 12, pp. 1349-1380, Dec. 2000.
2. <https://en.wikipedia.org/wiki/List_of_CBIR_engines>
3. Flores-Pulido, L.; Rodríguez-Gómez, G.; Starostenko, O.; Santacruz-Olmos, C., "Radial Basis Function for Visual Image Retrieval", Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2010.
4. Kwitt R, Uhl A, "Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retreival", 15th IEEE International Conference on Image Processing (ICIP 2008), 2008.