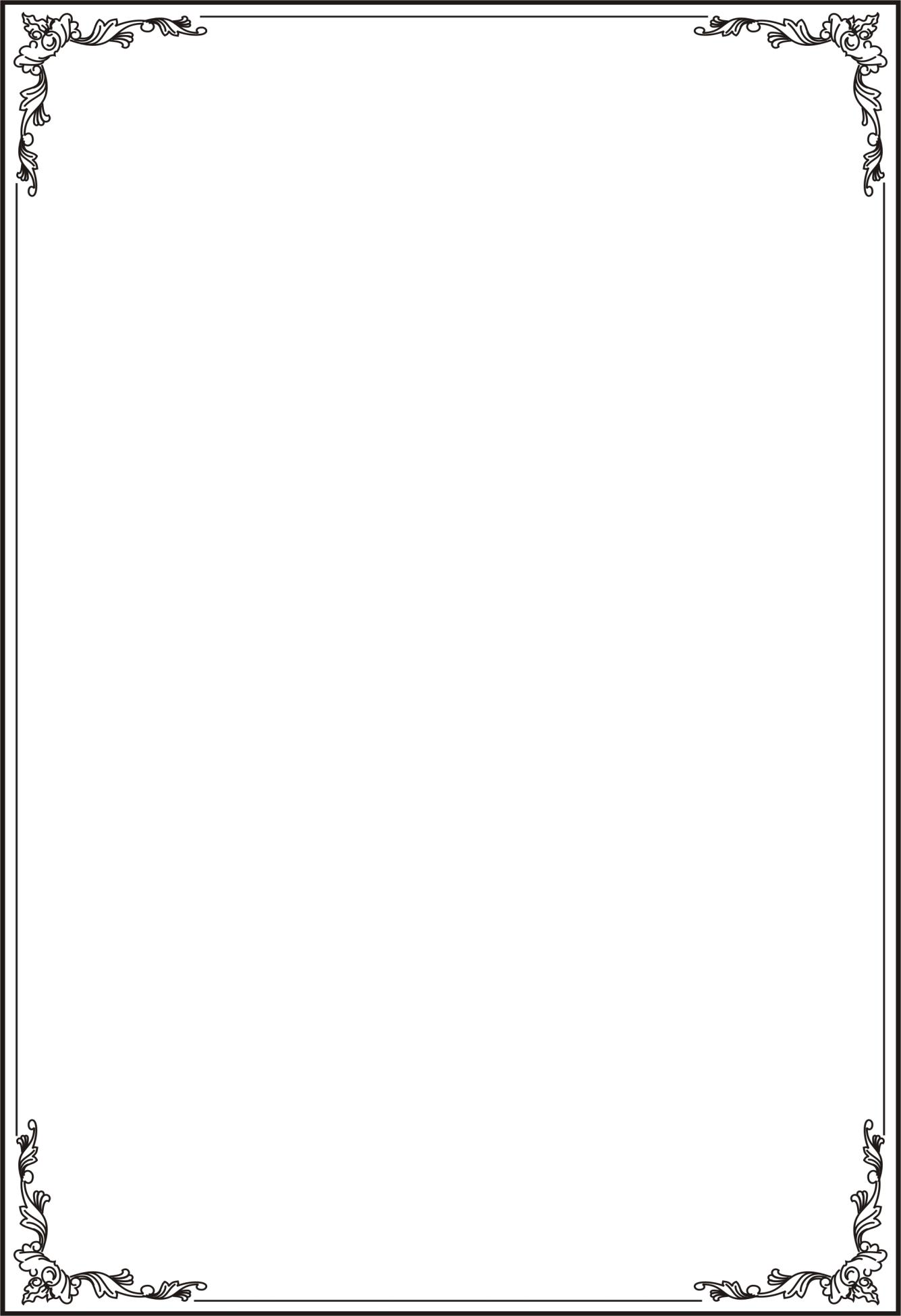
****

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KHOA HỌC MÁY TÍNH**

**BÁO CÁO ĐỒ ÁN MÔN TRUY VẤN THÔNG TIN VÀ ĐA PHƯƠNG TIỆN  
Đề tài: Truy vấn cơ sở dữ liệu ảnh với ảnh mẫu**

**Lớp: CS336.I21.KHTN**

**GVLT: Nguyễn Lưu Huỳnh Ngân**

**GV HDTH: Nguyễn Vinh Tiệp**

**Nhóm thực hiện:**

1. **Nguyễn Minh Dũng - 15520138**
2. **Nguyễn Quốc Danh - 15520092**
3. **Trịnh Hoàng Ngọc - 15520556**

**Tp.Hồ Chí Minh, ngày 30 tháng 06 năm 2018**

Mục lục

[1 Mô tả bài toán 1](#_Toc514214661)

[2 Các kỹ thuật chính cần có để giải quyết bài toán 1](#_Toc514214662)

[3 Phương pháp cơ bản (baseline) để giải quyết 1](#_Toc514214663)

[4 Cách đánh giá độ chính xác của thuật toán 1](#_Toc514214664)

# Giới thiệu đề tài

- Sự phát triển mạnh mẽ của công nghệ ảnh số làm lượng ảnh lưu trữ trên web tăng  
lên một cách nhanh chóng. Mỗi ngày, có hàng triệu bức ảnh được đăng tải trên các  
trang ảnh trực tuyến như: Flickr, Photobucket, Facebook, ... Vì thế nhu cầu tìm kiếm bằng hỉnh ảnh nhận được nhiều sự quan tâm của người dùng.

- Để giải quyết vấn đề này, các hệ thống tìm kiếm ảnh đã ra đời như: Yahoo, MSN, Google Image Search, Bing,….

- Input: Một tấm ảnh mẫu

- Output: Từ tập ảnh dataset lấy ra các ảnh tương đồng với ảnh mẫu nhất

- Tập dataset trong đồ án này là tập dataset Oxford (845 images) trong source mẫu của thầy Tiệp

Link Dataset: https://github.com/nvtiep/Instance-Search/tree/master/oxford

# Phân tích và thiết kế

Sơ đồ khối:

Mã giả:

B1: Sử dụng RootSift để lấy các điểm đặc trưng của từng image trong dataset lưu vào ListDescs.

B2: Áp dụng thuật toán gom nhóm MiniBatchKmeans để gom nhóm các desc trong listdescs lại với nhau, mỗi nhóm được xem như 1 “word”. Tập hợp các “word” ta được 1 tập Vocabulary.

B3: sử dụng Vector Quantization để tính toán Histogram cho các điểm đặc trưng.

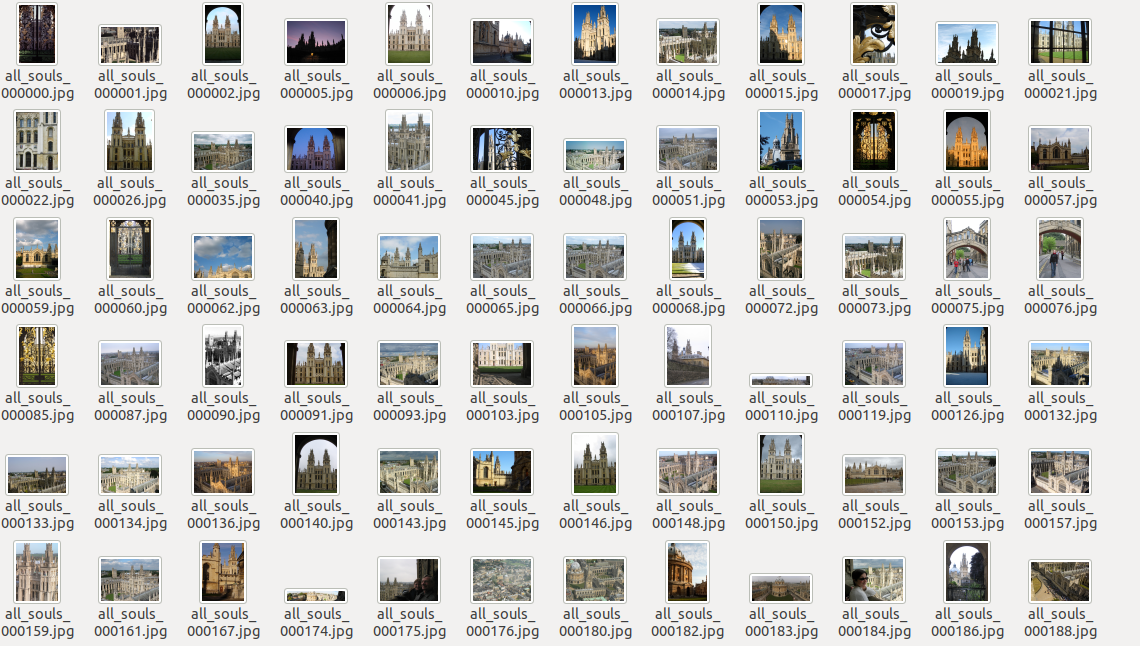
B4: tính Tf-Idf và L2 normalization

B5: tính toán descs, Tf-Idf và L2 normalization của Image query, Sắp xếp các images theo thứ tự tương đồng với image query nhất.

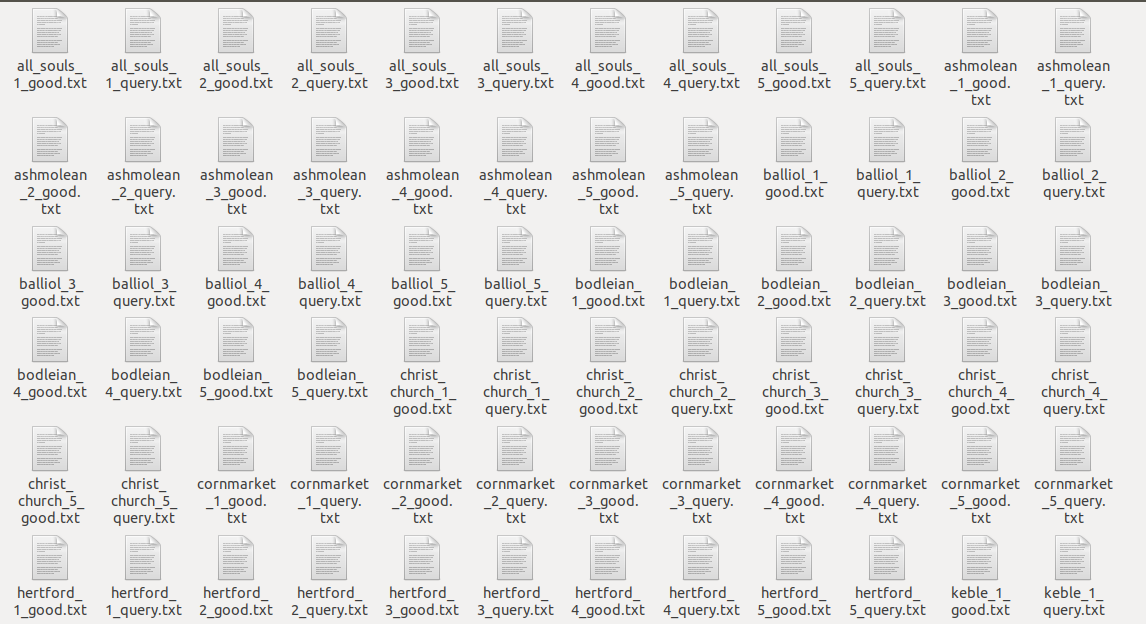
# Đánh giá độ chính xác:

## Thông tin dataset

- Số lượng images: 845 images



**-** Số lượng query: 55 images. Mỗi query có một tập \*\_good.txt để kiểm tra độ chính xác của kết quả tìm được.



## Cách đánh giá

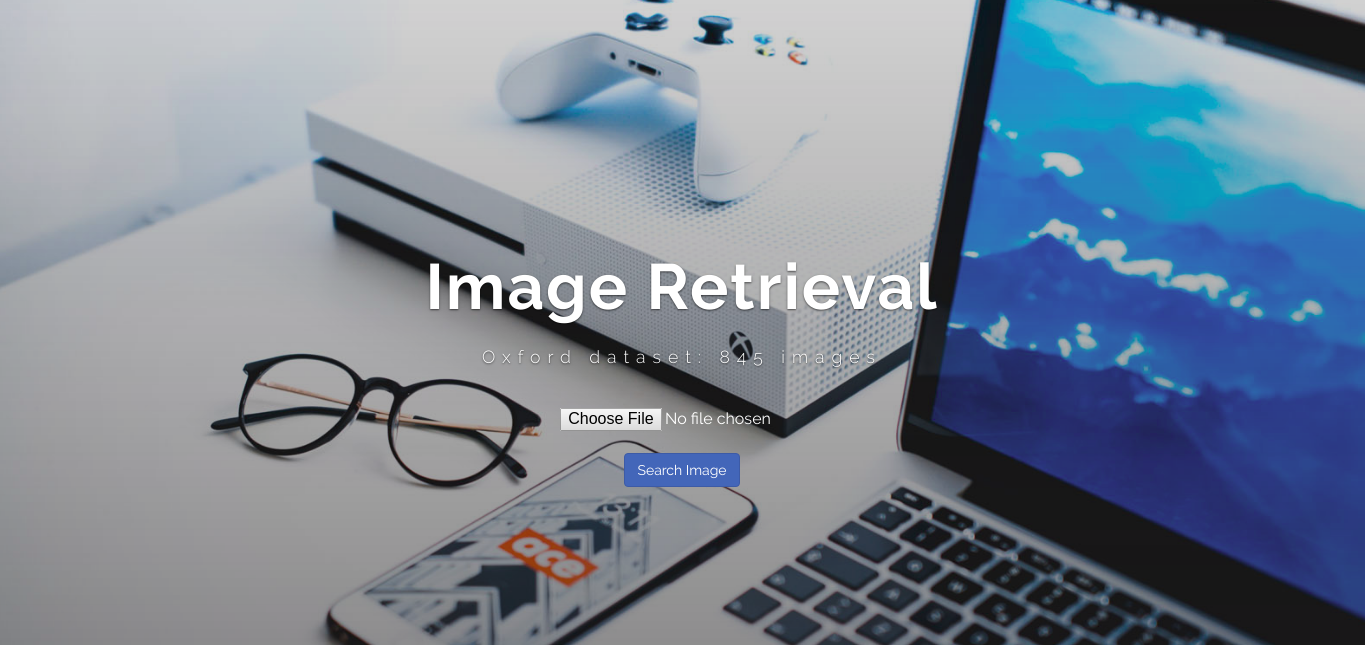
- Sử dụng Average Precision (AP) để tính toán cho từng query

**-** Tính trung bình các AP vừa tìm được ta có kết quả đánh giá MAP

# Kết quả thực nghiệm

- Link images query:

- Web search demo: <http://nguyendu393.pythonanywhere.com/search>



Link demo: <https://youtu.be/QlRbKrQSomk>

- Thời gian rút trích đặc trưng: ~250s

**-** Thời gian training:  **~**300s

- MAP: ~74%

# Đóng góp của từng thành viên

|  |  |  |
| --- | --- | --- |
| **Họ và Tên** | **MSSV** | **Công việc** |
| Nguyễn Minh Dũng | 15520138 | - Tìm hiểu đề tài, các phương pháp để giải quyết đề tài.  - Code chính |
| Nguyễn Quốc Danh | 15520092 | - Tìm hiểu đề tài, các phương pháp để giải quyết đề tài.  - Hỗ trợ code chính trong quá trình code, reviewcode  - Viết giao diện Web |
| Trịnh Hoàng Ngọc | 15520556 | - Tìm hiểu đề tài, các phương pháp để giải quyết đề tài.  - Viết báo cáo môn học |
|  |  |  |

# Tài liệu tham khảo

<https://github.com/nvtiep/Instance-Search>

<http://yael.gforge.inria.fr/tutorial/tuto_imgindexing.html>

<https://github.com/willard-yuan/image-retrieval/tree/master/bag-of-words-python-dev-version>

<https://github.com/willard-yuan/SoTu>

<https://github.com/Pranshu258/svtor/blob/master/svtor.py>

Các nguồn khác trên internet