



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA CÔNG NGHỆ THÔNG TIN



# BÁO CÁO

## BÀI TẬP THỰC HÀNH TUẦN 9

### Khai thác dữ liệu Web



**Giảng viên hướng dẫn :** Khoa Phó Ngọc Đăng

**Sinh viên thực hiện :** Vũ Mạnh Hùng

**Mã số sinh viên :** 1461390

**Lớp :** 15CK3

**Ca :** 1 – C6

TP HCM, tháng 5 năm 2017

I) Từng bước thực hiện từ tập cơ sở dữ liệu thô đến kết quả.

- String docs[] = {"Web mining is useful", "Usage mining applications", "Web structure mining studies the Web hyperlink structure"};
- => Tạo mảng các tài liệu
- String[] NguVung = getUniqueTokens(docs);
- 
- ArrayList<Map<String, Integer>> frequent = getFrequent(docs, NguVung);
- 
- ArrayList< ArrayList<Integer>> reverseIndex = getReverseIndex(docs, NguVung);
- 
- 
- //YOUR CODE HERE: ok
- String query\_string;
- System.out.println("Nhập câu truy vấn: ");
- Scanner scan = new Scanner(System.in);
- query\_string = scan.nextLine();
- printSearch(query\_string, NguVung, frequent, reverseIndex);
- 
- //In kết quả danh sách chỉ mục vào file
- //Ref: <https://ngockhuong.com/java/cach-ghi-them-noi-dung-vao-file-trong-java.html>
- String stringPath =
- "D:\\Couses\_VI\\KTDLWeb\\15CK3\\TH\\W09\\1461390\\output.txt";
- File file = new File(stringPath);
- 
- if (!file.exists()) {
- file.createNewFile();
- }
- 
- FileWriter fw = new FileWriter(file.getAbsolutePath(), true);
- BufferedWriter bw = new BufferedWriter(fw);
- 
- String[] query =
- tokenizeAndStemmingAndRemoveDefaultStopWords(query\_string);
- ArrayList<SearchItem> searchResult = search(query, NguVung, frequent, reverseIndex);
- for (int i = 0; i < searchResult.size(); i++) {
- SearchItem item = searchResult.get(i);
- String data = Integer.toString(item.docIdx) + ": " +
- Integer.toString(item.count) + "\n";
- bw.write(data);
- }
- 
- if (bw != null) {
- bw.close();
- }
- if (fw != null) {
- fw.close();
- }
- 
-

- //Xuất ra màn hình:
- `System.out.println("Output My Teacher");`
- `query_string = "Web";`
- `printSearch(query_string, NguVung, frequent, reverseIndex);`
- `System.out.println("=====");`
- `query_string = "Web Mining";`
- `printSearch(query_string, NguVung, frequent, reverseIndex);`
- `System.out.println("=====");`
- `query_string = "Usage mining";`
- `printSearch(query_string, NguVung, frequent, reverseIndex);`
- `System.out.println("=====");`
- `query_string = "mining";`
- `printSearch(query_string, NguVung, frequent, reverseIndex);`

II) Từng bước được thực hiện bằng những hàm nào?

III) Các dòng code trong các hàm thực hiện công việc gì?