**ĐẠI HỌC CẦN THƠ**

**TRƯỜNG CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**

**KHOA HỆ THỐNG THÔNG TIN**

**🙞 🕮 🙜**



**BÁO CÁO THỰC TẬP THỰC TẾ**

**NGÀNH HỆ THỐNG THÔNG TIN**

**Đề tài**

**ỨNG DỤNG ELASTICSEARCH TỐI ƯU CHỨC NĂNG TÌM KIẾM VĂN BẢN**

**Đơn vị thực tập**

**Trung tâm Công nghệ thông tin – Viễn Thông Tiền Giang**

|  |  |
| --- | --- |
| ***Cán bộ hướng dẫn*:** | ***Giáo viên hướng dẫn:*** |
| **Nguyễn Thái Duy** | **TS Phạm Thị Ngọc Diễm** |
|  |  |
| ***Sinh viên thực hiện*:** | ***Học phần:*** |
| **Hồ Thị Minh Tuyền** | **Thực tập thực tế - HTTT** |
| **MSSV: B1906358** | **Mã học phần: CT472** |
| **Khóa 45** | **Nhóm: 02** |

**Tiền Giang, 07/2023**

**LỜI CẢM ƠN**

Lời đầu tiên, em xin chân thành cảm ơn **Ban giám đốc Trung tâm Công nghệ thông tin - Viễn thông Tiền Giang** đã tạo cho em cơ hội được thực tập tại trung tâm và tạo cho em môi trường tốt nhất để em có thể thuận lợi thực tập. Em cũng bày tỏ lời cảm ơn chân thành đến anh **Nguyễn Thái Duy** là cán bộ hướng dẫn của em, anh đã tận tình hướng dẫn, chỉ bảo em trong suốt 8 tuần qua để em hoàn thành tốt công việc thực tập của mình tại Trung tâm.

Đồng thời, em cũng xin gửi lời cảm ơn đến quý thầy cô **Trường Công nghệ thông tin & Truyền Thông - Đại Học Cần Thơ** đã nhiệt tình giảng dạy và truyền đạt cho em nhiều kiến thức quý báu trong 4,5 năm em học tập và tạo điều kiện thuận lợi để em có cơ hội được trải nghiệm với môi trường mới chuyên nghiệp và năng động.

Đặc biệt, em xin cảm ơn cô **Phạm Thị Ngọc Diễm** là cố vấn học tập của em, cô đã nhiệt tình cung cấp các thông tin về vấn đề thực tập, giải đáp các thắc mắc cho em cũng như các bạn lớp Hệ thống thông tin K45. Và cuối cùng em xin gửi lời cảm ơn đến người thân, gia đình và bạn bè đã hỗ trợ, động viên, giúp đỡ em trong suốt quá trình thực tập vừa qua.

**Xin chân thành cảm ơn!**

Tiền Giang, ngày 05 tháng 07 năm 2023

Sinh viên thực hiện

**Hồ Thị Minh Tuyền**

**PHIẾU ĐÁNH GIÁ BÁO CÁO KẾT QUẢ THỰC TẬP**

**HỌC KỲ 2 – 2022-2023**

**(Dùng cho giáo viên chấm báo cáo thực tập)**

Họ và tên cán bộ chấm báo cáo: TS Phạm Thị Ngọc Diễm

Họ tên sinh viên thực tập: Hồ Thị Minh Tuyền Mã số SV: B1906358

|  |  |  |
| --- | --- | --- |
| **Nội dung đánh giá** | **Điểm**  **tối đa** | **Điểm**  **chấm** |
| **I. Hình thức trình bày** | **0.5** |  |
| I.1 Đúng format của khoa (Trang bìa, trang lời cảm ơn, trang đánh giá thực tập của khoa, trang mục lục và các nội dung báo cáo). Sử dụng đúng mã và font tiếng Việt (Unicode Times New Roman, Size 13) | 0.25 |  |
| I.2 Trình bày mạch lạc, súc tích, không có lỗi chính tả | 0.25 |  |
| **II. Phiếu theo dõi** | **0.5** |  |
| II.1 Có lịch làm việc đầy đủ cho 8 tuần | 0.25 |  |
| II.2 Số buổi thực tập tại cơ quan trong 1 tuần >=6; ít hơn 6 buổi 0.0 điểm | 0.25 |  |
| **III. Đánh giá của công ty** | **8.0** |  |
| Cách tính điểm = (Điểm cộng của cán bộ hướng dẫn/100) x 8.0 | 8.0 |  |
| **IV. Nội dung thực tập (quyển báo cáo)** | **1.0** |  |
| - Có được sự hiểu biết tốt về cơ quan nơi thực tập.  - Phương pháp thực hiện phù hợp với nội dung công việc được giao.  - Kết quả củng cố lý thuyết.  - Kết quả rèn luyện kỹ năng thực hành.  - Kinh nghiệm thực tiễn thu nhận được.  - Kết quả công việc có đóng góp cho cơ quan nơi thực tập | 0.5 |  |
| **TỔNG CỘNG** | **10.0** |  |
| Điểm trừ |  |  |
| **Điểm còn lại** |  |  |

………….., ngày….tháng….năm2023

GIÁO VIÊN CHẤM BÁO CÁO

*(ký tên)*

**MỤC LỤC**

[PHẦN 1: BÁO CÁO KẾT QUẢ TÌM HIỂU VỀ TỔ CHỨC HÀNH CHÍNH, NHÂN SỰ CỦA CƠ QUAN THỰC TẬP 3](#_Toc139551295)

[I. TÌM HIỂU VỀ TỔ CHỨC HÀNH CHÍNH VÀ NHÂN SỰ CỦA CƠ QUAN THỰC TẬP 3](#_Toc139551296)

[1. GIỚI THIỆU VỀ TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM 3](#_Toc139551297)

[1.1 Lĩnh vực kinh doanh 3](#_Toc139551298)

[1.2 Đơn vị thành viên 4](#_Toc139551299)

[2. TÌM HIỂU VỀ TRUNG TÂM CÔNG NGHỆ THÔNG TIN – VNPT TIỀN GIANG 4](#_Toc139551300)

[2.1 Giới thiệu 4](#_Toc139551301)

[2.2 Mô hình tổ chức 4](#_Toc139551302)

[PHẦN 2: TÌM HIỂU VỀ HOẠT ĐỘNG CHUYÊN NGÀNH VÀ MÔI TRƯỜNG LÀM VIỆC CỦA CƠ QUAN THỰC TẬP 5](#_Toc139551303)

[I. HOẠT ĐỘNG CHUYÊN NGÀNH 5](#_Toc139551304)

[II. MÔI TRƯỜNG LÀM VIỆC 5](#_Toc139551305)

[PHẦN 3: NỘI DUNG CÔNG VIỆC ĐƯỢC PHÂN CÔNG 6](#_Toc139551306)

[I. NỘI DUNG CÔNG VIỆC ĐƯỢC PHÂN CÔNG 6](#_Toc139551307)

[1.1 BẢNG PHÂN CÔNG CÔNG VIỆC 6](#_Toc139551308)

[1.2 NHỮNG VIỆC LÀM ĐƯỢC 7](#_Toc139551309)

[1.3 NHỮNG VIỆC CHƯA LÀM ĐƯỢC 7](#_Toc139551310)

[PHẦN 4: PHƯƠNG PHÁP THỰC HIỆN 8](#_Toc139551311)

[1. QUY TRÌNH LÀM VIỆC TRONG QUÁ TRÌNH THỰC TẬP 8](#_Toc139551312)

[2. CƠ SỞ LÝ THUYẾT 8](#_Toc139551313)

[2.1 Giới thiệu Elasticsearch 8](#_Toc139551314)

[2.2 Mô hình tổng thể 9](#_Toc139551315)

[2.3 Cách thức hoạt động 9](#_Toc139551316)

[2.4 Cài đặt bộ phân tích Tiếng Việt vào Elasticsearch 10](#_Toc139551317)

[2.5 Spring Boot 10](#_Toc139551318)

[3. PHƯƠNG THỨC CÀI ĐẶT 10](#_Toc139551319)

[3.1 Cài đặt Elasticsearch 10](#_Toc139551320)

[3.2 Xây dựng ứng dụng Spring boot 10](#_Toc139551321)

[3.3 Cấu hình kết nối Elasticsearch vào Spring boot 11](#_Toc139551322)

[4. TÍNH NĂNG ỨNG DỤNG 11](#_Toc139551323)

[4.1 Tính năng tách từ bằng bộ phân tích vi\_analyzer 11](#_Toc139551324)

[4.2 Tính năng tách từ bằng bộ phân tích my\_vi\_analyzer 11](#_Toc139551325)

[PHẦN 5: KẾT QUẢ ĐẠT ĐƯỢC 11](#_Toc139551326)

[1. NHỮNG KIẾN THỨC ĐƯỢC CỦNG CỐ 11](#_Toc139551327)

[2. NHỮNG KIẾN THỨC ĐƯỢC HỌC HỎI 12](#_Toc139551328)

[3. NHỮNG KINH NGHIỆM THỰC TIỄN ĐƯỢC TÍCH LŨY 12](#_Toc139551329)

[4. CHI TIẾT CÔNG VIỆC ĐÃ ĐÓNG GÓP CHO CƠ QUAN 12](#_Toc139551330)

**DANH MỤC HÌNH**

[Hình 1: Mô hình tổng thể 10](#_Toc139467345)

[Hình 2: Mô hình hoạt động Elasticsearch 10](#_Toc139467346)

[Hình 3: Mô hình kiến trúc 12](#_Toc139467347)

**NỘI DUNG BÁO CÁO**

# PHẦN 1: BÁO CÁO KẾT QUẢ TÌM HIỂU VỀ TỔ CHỨC HÀNH CHÍNH, NHÂN SỰ CỦA CƠ QUAN THỰC TẬP

## I. TÌM HIỂU VỀ TỔ CHỨC HÀNH CHÍNH VÀ NHÂN SỰ CỦA CƠ QUAN THỰC TẬP

### 1. GIỚI THIỆU VỀ TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM

**Tập đoàn Bưu chính Viễn thông Việt Nam** (tên giao dịch quốc tế: Vietnam Posts and Telecommunications Group, viết tắt: VNPT) là một doanh nghiệp nhà nước chuyên đầu tư, sản xuất, hoặc kinh doanh trong lĩnh vực bưu chính và viễn thông tại Việt Nam. Hiện tại trụ sở chính đang được đặt tại: Tòa nhà VNPT, số 57 Huỳnh Thúc Kháng, P. Láng Hạ, Q. Đống Đa, TP. Hà Nội. VNPT vừa là nhà cung cấp dịch vụ đầu tiên đặt nền móng cho sự phát triển của ngành Bưu chính, Viễn thông Việt Nam, vừa là tập đoàn có vai trò chủ chốt trong việc đưa Việt Nam trở thành 1 trong 10 quốc gia có tốc độ phát triển Bưu chính Viễn thông nhanh nhất toàn cầu. Theo Quyết định số 888/QĐ-TTg ngày 10/6/2014 của Thủ tướng Chính phủ, VNPT đã thực hiện tái cơ cấu tổ chức. Sau khi tái cấu trúc doanh nghiệp, hiện với gần 40 nghìn cán bộ công nhân viên, hạ tầng công nghệ viễn thông tiên tiến, mạng lưới dịch vụ phủ sóng toàn bộ 63 tỉnh thành trên cả nước, VNPT tự hào là nhà cung cấp dịch vụ bưu chính, viễn thông số 1 tại Việt Nam, phục vụ khoảng 30 triệu thuê bao di động, gần 10 triệu thuê bao điện thoại cố định và khoảng hàng chục triệu người sử dụng Internet.

#### 1.1 Lĩnh vực kinh doanh

Dịch vụ và sản phẩm viễn thông, công nghệ thông tin và truyền thông đa phương tiện; Khảo sát, tư vấn, thiết kế, lắp đặt, khai thác, bảo dưỡng, sửa chữa, cho thuê các công trình viễn thông, công nghệ thông tin; Nghiên cứu, phát triển, chế tạo, sản xuất thiết bị, sản phẩm viễn thông, công nghệ thông tin; Thương mại, phân phối các sản phẩm thiết bị viễn thông, công nghệ thông tin; Dịch vụ quảng cáo, nghiên cứu thị trường, tổ chức hội nghị hội thảo, triển lãm liên quan đến lĩnh vực viễn thông, công nghệ thông tin; Kinh doanh bất động sản, cho thuê văn phòng; Dịch vụ tài chính trong lĩnh vực viễn thông, công nghệ thông tin và truyền thông đa phương tiện.

#### 1.2 Đơn vị thành viên

* Tổng công ty dịch vụ viễn thông (VNPT-VinaPhone);
* Tổng công ty hạ tầng mạng (VNPT-Net);
* Tổng công ty truyền thông (VNPT-Media);
* Công ty Công nghệ thông tin VNPT (VNPT-IT);
* Công ty CP Công nghệ Công nghiệp Bưu chính Viễn thông (VNPT Technology).

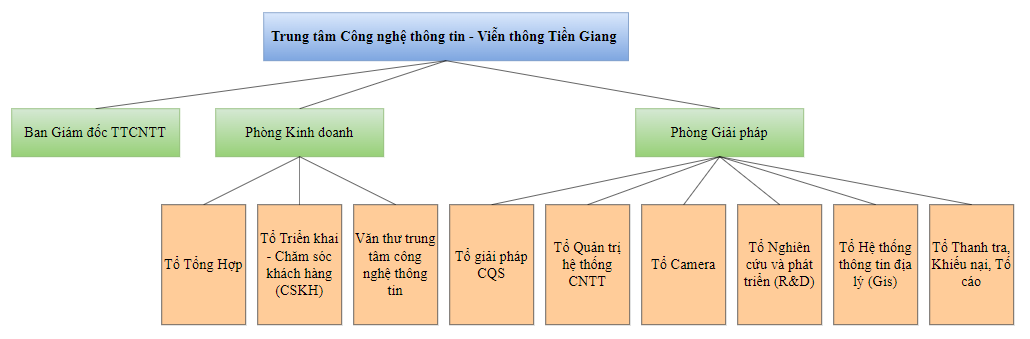
### 2. TÌM HIỂU VỀ TRUNG TÂM CÔNG NGHỆ THÔNG TIN – VNPT TIỀN GIANG

#### 2.1 Giới thiệu

Trung tâm Công nghệ thông tin - Viễn thông Tiền Giang trực thuộc VNPT Tiền Giang nhiệm vụ chính của trung tâm là nghiên cứu, phát triển, sản xuất, kinh doanh các sản phẩm, dịch vụ công nghệ thông tin để cung cấp cho nội bộ VNPT. Bên cạnh đó trung tâm cũng thiết kế, triển khai các phần mềm theo đơn đặt hàng của khách hàng hoặc thiết kế các phần mềm hướng đến những đối tượng cụ thể và bán các phần mềm đó.

Trung tâm Công nghệ thông tin VNPT Tiền Giang đầu tư, phát triển, quản lý các hệ thống, nền tảng công nghệ thông tin; vận hành khai thác hệ thống điều hành sản xuất kinh doanh, đảm bảo an toàn, bảo mật thông tin cho các sản phẩm và các dịch vụ công nghệ thông tin của VNPT cung cấp cho khách hàng.

#### 2.2 Mô hình tổ chức



# PHẦN 2: TÌM HIỂU VỀ HOẠT ĐỘNG CHUYÊN NGÀNH VÀ MÔI TRƯỜNG LÀM VIỆC CỦA CƠ QUAN THỰC TẬP

## I. HOẠT ĐỘNG CHUYÊN NGÀNH

Trung tâm công nghệ thông tin – Viễn Thông Tiền Giang với nhiệm vụ chuyên viết các sản phẩm phần mềm theo đơn đặt hàng của khách hàng và cung cấp sản phẩm phần mềm của tập đoàn ra bên ngoài. Trung tâm có đội ngũ cán bộ công nhân viên có trình độ cao, tác phong chuyên nghiệp và trách nhiệm, luôn đảm bảo đưa ra các sản phẩm phần mềm hoàn thiện nhất.

Đồng thời, với sự phát triển không ngừng của công nghệ thông tin hiện nay, Trung tâm luôn tạo những điều kiện tốt nhất để cán bộ, công nhân viên phát huy hết khả năng và không ngừng học tập, cập nhật xu hướng mới để theo kịp với sự phát triển của xã hội.

## II. MÔI TRƯỜNG LÀM VIỆC

* Môi trường làm việc hiện đại, năng động và cởi mở giúp nhân viên có thể phát huy tối đa khả năng của mình.
* Cơ sở vật chất, trang thiết bị đầy đủ và hiện đại và bố trí thuận tiện để nhân viên có thể sử dụng, không gian làm việc rộng rãi, thoáng mát tạo cho nhân viên tinh thần thoải mái góp phần tăng hiệu suất công việc.
* Có cơ hội tiếp cận và học hỏi thêm nhiều lĩnh vực mới. Được đào tạo về kỹ năng mềm và các kỹ thuật chuyên sâu. Trải nghiệm các công nghệ mới, cùng những nhân viên trẻ đầy nhiệt huyết, đam mê, trong môi trường toàn cầu, chuyên nghiệp, với những cơ hội phát triển bản thân toàn diện.
* Trải qua gần 2 tháng thực tập tại Trung tâm CNTT- Viễn Thông Tiền Giang, em nhận thấy đây là môi trường làm việc rất tốt cho những sinh viên công nghệ thông tin có đam mê, tâm huyết và muốn gắn bó lâu dài với nghề.

# PHẦN 3: NỘI DUNG CÔNG VIỆC ĐƯỢC PHÂN CÔNG

## I. NỘI DUNG CÔNG VIỆC ĐƯỢC PHÂN CÔNG

### 1.1 BẢNG PHÂN CÔNG CÔNG VIỆC

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Tuần** | **Nội dung công việc được giao (phù hợp với đề cương)** | **Dự kiến tổng số buổi sinh viên làm việc tại cơ quan trong 1 tuần** |
| 1 | Từ ngày 15/05/2023  đến ngày 19/05/2023 | Tìm hiểu về ElasticSearch: Khái niệm, cơ chế hoạt động, dựng server và kết nối trao đổi thông tin | 10 buổi |
| 2 | Từ ngày 22/05/2023  đến ngày 26/05/2023 | Tìm hiểu về ElasticSearch: Khái niệm, cơ chế hoạt động, dựng server và kết nối trao đổi thông tin | 10 buổi |
| 3 | Từ ngày 29/05/2023  đến ngày 02/06/2023 | Tìm hiểu về Spring Boot và cách xây dựng một service để giao tiếp với ES | 10 buổi |
| 4 | Từ ngày 05/06/2023  đến ngày 09/06/2023 | Tìm hiểu về Spring Boot và cách xây dựng một service để giao tiếp với ES | 10 buổi |
| 5 | Từ ngày 12/06/2023  đến ngày 16/06/2023 | Xây dựng Chức năng tìm kiếm Văn bản cho hệ thống Văn phòng điện tử sử dụng ES và Spring Boot microservice. | 10 buổi |
| 6 | Từ ngày 19/06/2023  đến ngày 23/06/2023 | Xây dựng Chức năng tìm kiếm Văn bản cho hệ thống Văn phòng điện tử sử dụng ES và Spring Boot microservice. | 10 buổi |
| 7 | Từ ngày 26/06/2023  đến ngày 30/06/2023 | Xây dựng Chức năng tìm kiếm Văn bản cho hệ thống Văn phòng điện tử sử dụng ES và Spring Boot microservice. | 10 buổi |
| 8 | Từ ngày 03/07/2023  đến ngày 07/07/2023 | Demo và viết báo cáo. | 10 buổi |

### 1.2 NHỮNG VIỆC LÀM ĐƯỢC

* Tìm hiểu cơ sở lý thuyết về khái niệm, cấu trúc và hoạt động của Elasticsearch.
* Cài đặt Elasticsearch trên Docker cấu hình bộ phân tích Tiếng Việt.
* Tìm hiểu và tạo ứng dụng Spring boot, lập trình tạo ra các phương thức API cơ bản.
* Tìm hiểu các phương thức RESTful API trong Elasticsearch: API Index, API document, API Search và API Bulk,...
* Kiểm thử các phương thức bằng công cụ Postman.

**Về công cụ:**

* Docker Desktop: công cụ giúp cho việc tạo ra và triển khai các container để phát triển, chạy ứng dụng được dễ dàng.
* Gitlab: công cụ quản lý code và báo cáo.
* IntelliJ IDEA 2020.1.3: trình IDE dùng để lập trình Java (nó cũng được sử dụng để lập trình một số ngôn ngữ khác như Node.js, Python…)
* Postman: Cho phép làm việc với các API, giúp ích rất nhiều cho việc kiểm thử.

### 1.3 NHỮNG VIỆC CHƯA LÀM ĐƯỢC

* Một số kỹ thuật trong cách thiết kế và kết nối chưa thật sự nắm vững dễ xảy ra lỗi và phương thức xây dựng chưa tối ưu.
* Tài liệu tham khảo khá nhiều, chưa thể tiếp thu hết và chọn lọc cho phù hợp với dự án.

# PHẦN 4: PHƯƠNG PHÁP THỰC HIỆN

## 1. QUY TRÌNH LÀM VIỆC TRONG QUÁ TRÌNH THỰC TẬP

**-** Tìm hiểu về đề tài bằng những tài liệu từ cán bộ hướng dẫn, có thể đặt câu hỏi cho cán bộ hướng dẫn. Ngoài ra, tìm các tài liệu liên quan trên Internet.

- Lập bảng phân bố công việc theo tuần để có thể hoàn thành đúng hạn.

- Bổ sung, chỉnh sửa các chức năng trong dự án theo yêu cầu của cán bộ hướng dẫn.

- Trong quá trình phân tích, thiết kế và cài đặt nếu gặp vấn đề không thể khắc phục được có thể nhờ sợ hỗ trợ của cán bộ hướng dẫn.

- Kiểm thử, sửa lỗi và hoàn thiện các chức năng của ứng dụng.

## 2. CƠ SỞ LÝ THUYẾT

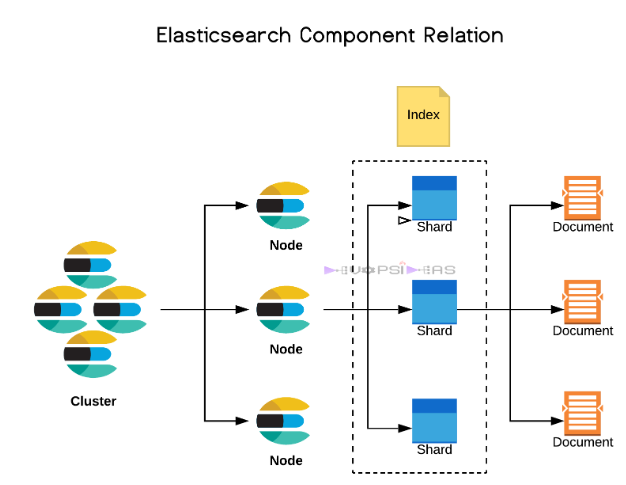
### 2.1 Giới thiệu Elasticsearch

Elasticsearch là một công cụ tìm kiếm và phân tích dữ liệu mã nguồn mở. Được xây dựng trên nền tảng Apache Lucene bằng ngôn ngữ JAVA với giao diện web HTTP có hỗ trợ JSON. Cho phép ta chuyển dữ liệu vào và tìm kiếm trực tiếp, thay vì tìm kiếm trong một hệ quản lý cơ sở dữ liệu thông thường như Oracle, MySQL,… Là một thư viện tìm kiếm văn bản mạnh mẽ, và được thiết kế để tìm kiếm và phân tích dữ liệu phân tán với hiệu suất cao.

Elasticsearch cung cấp khả năng tìm kiếm và phân tích dữ liệu nhanh chóng và dễ dàng, với các tính năng như tìm kiếm văn bản đầy đủ, tìm kiếm tương tự, tìm kiếm theo trường, và tìm kiếm đa trường. Elasticsearch cũng hỗ trợ phân tích dữ liệu với các tính năng như tạo báo cáo, tìm kiếm theo dạng đồ thị, và phân tích thống kê.

Elasticsearch được thiết kế để hoạt động trong môi trường phân tán, cho phép nó xử lý dữ liệu lớn và đảm bảo khả năng mở rộng. Nó cũng có thể tích hợp với các công cụ khác trong hệ sinh thái Elastic, như Logstash và Kibana, để cung cấp một giải pháp toàn diện cho việc quản lý dữ liệu và phân tích trong môi trường phân tán.

### 2.2 Mô hình tổng thể

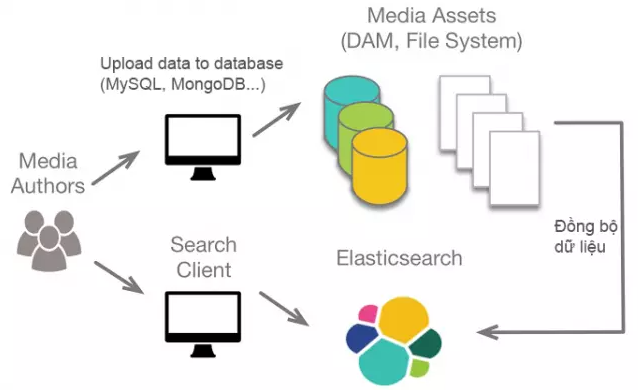


Hình 1: Mô hình tổng thể thành phần Elasticsearch

### 2.3 Cách thức hoạt động

Elasticsearch hoạt động như một Cloud Server có khả năng tìm kiếm thông qua cơ chế RESTful API. Trong đó, người dùng tạo ra các HTTP Request và dữ liệu dạng JSON, sau đó nhập vào Elasticsearch. Các dữ liệu này đều được đánh Index, giúp đem lại hiệu quả tìm kiếm.

1. Phân tích xử lý dữ liệu từ quá trình nhập liệu của người dùng.
2. Dữ liệu được phân loại và chỉ mục, được đẩy lên Server ElasticSearch.
3. Người dùng có thể tạo các truy vấn phức tạp và lấy dữ liệu được trả về từ server Elasticsearch.



Hình 2: Mô hình hoạt động Elasticsearch

### 2.4 Cài đặt bộ phân tích Tiếng Việt vào Elasticsearch

Bộ phân tích Tiếng Việt tích hợp phân tích ngôn ngữ Tiếng Việt vào Elasticsearch. Sử dụng C++ tokenizer cho thư viện Tiếng Việt được phát triển bởi một nhóm chuyên gia công nghệ và nhà phát triển tại công ty CocCoc. Bộ phân tích này được cung cấp từ (*https://github.com/duydo/elasticsearch-analysis-vietnamese*).

Bộ phân tích Tiếng Việt cho phép Elasticsearch thực hiện các nhiệm vụ như tách từ, chuẩn hóa từ, xử lý ký tự đặc biệt, áp dụng các quy tắc phân tích văn bản Tiếng Việt. Điều này giúp Elasticsearch hiểu và xử lý đúng cú pháp và ngữ nghĩa của văn bản Tiếng Việt, giúp tăng cường khả năng tìm kiếm và phân tích trong các ứng dụng sử dụng Elasticsearch.

Việc cài đặt bộ phân tích tiếng Việt vào Elasticsearch, giúp tận dụng các tính năng và chức năng mạnh mẽ của Elasticsearch để xây dựng các ứng dụng tìm kiếm, phân tích và khám phá dữ liệu Tiếng Việt một cách chính xác và hiệu quả.

### 2.5 Spring Boot

Spring Boot là một framework được xây dựng trên nền tảng của Spring Framework, nhằm giúp phát triển ứng dụng Java dễ dàng và nhanh chóng. Spring Boot cung cấp một cách tiếp cận cấu hình tập trung và tự động hóa nhiều công việc phức tạp trong quá trình phát triển ứng dụng.

## 3. PHƯƠNG THỨC CÀI ĐẶT

### 3.1 Cài đặt Elasticsearch

* Cài đặt Elasticsearch trên Docker, sử dụng Docker CLI để tải và chạy Elasticseach từ Docker Hub.
* Sau khi chạy Elasticsearch trên Docker các cấu hình mặc định sẽ được áp dụng như: network.host, cluster.name, node.name,..
* Elasticsearch được chạy trên cổng 9200 và giao tiếp bằng Postman thông qua các API cơ bản như: API Index, API Document, API Search, ..

### 3.2 Xây dựng ứng dụng Spring boot

* Xây dựng trên dự án Maven, Spring boot 3.3.1 và Java 17.
* Cấu hình các dependencies: Spring Data Elasticsearch 4.0.0.RELEASE, Spring Web và Lombok.
* Ứng dụng được xây dựng trên IntelliJ IDEA 2020.1.3.

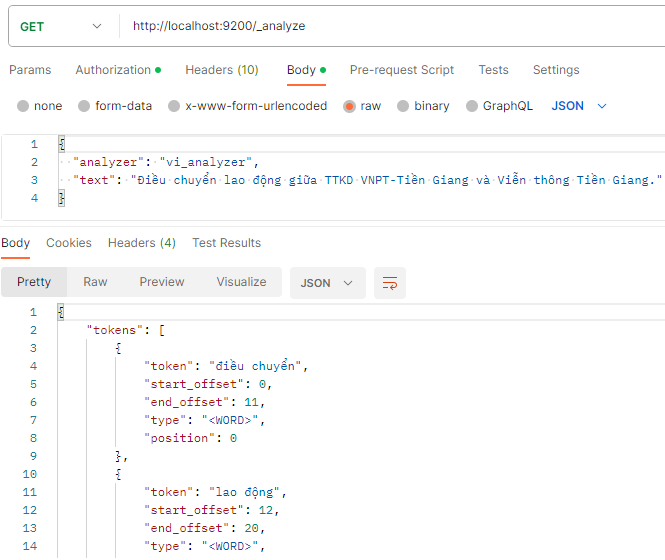
### 3.3 Cấu hình kết nối Elasticsearch vào Spring boot

* Thông qua *ClientConfiguration*, bao gồm địa chỉ Elasticsearch và thông tin xác thực, và sau đó tạo ra một phiên bản *RestHighLevelClient* đã được cấu hình để kết nối và tương tác với Elasticsearch.
* Tiến hành xây dựng các phương thức GET, POST, PUT và DELTE vào ứng dụng và được giao tiếp bằng công cụ Postman.

## 4. TÍNH NĂNG ỨNG DỤNG

### 4.1 Tính năng tách từ bằng bộ phân tích vi\_analyzer

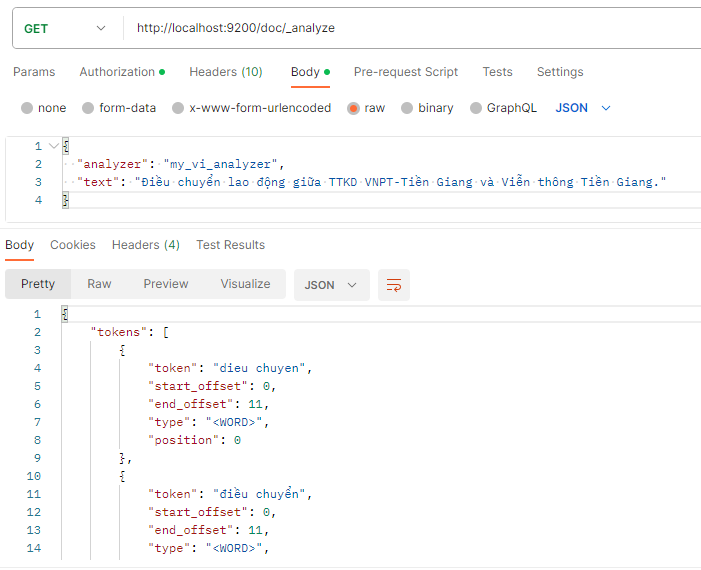
Bộ phân tích vi\_analyzer thực hiện tách từ có nghĩa trong một đoạn văn bản

****

Hình 3 Bộ phân tích vi\_analyzer

### 4.2 Tính năng tách từ bằng bộ phân tích my\_vi\_analyzer

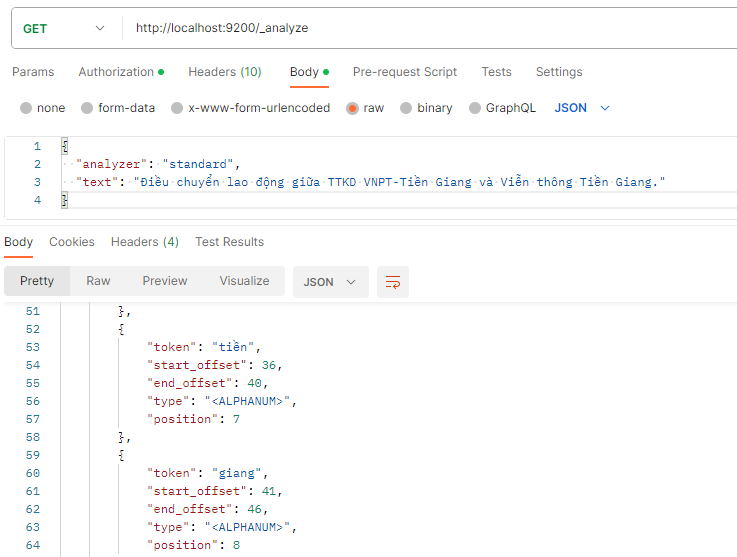
Bộ phân tích my\_vi\_analyzer thực hiện xử lý ký tự không dấu và tách từ có nghĩa trong một đoạn văn bản



Bộ phân tích my\_vi\_analyzer

### 4.3 Tính năng tách từ bằng bộ phân tích standard

Bộ phân tích standard cắt từ bằng khoảng trắng và loại bỏ các ký tự không phải chữ



Bộ phân tích standard

# PHẦN 5: KẾT QUẢ ĐẠT ĐƯỢC

## 1. NHỮNG KIẾN THỨC ĐƯỢC CỦNG CỐ

* Củng cố kiến thức về ngôn ngữ Java và Spring Boot.
* Củng cố kiến thức về hoạt động các phương thức API để lập trình, thiết kế và truy vấn dữ liệu.

## 2. NHỮNG KIẾN THỨC ĐƯỢC HỌC HỎI

* Kỹ năng phân tích, thiết kế, kiểm thử, sửa lỗi.
* Học được khả năng tự tìm hiểu, học hỏi và giải quyết vấn đề trong môi trường làm việc thực tế.
* Phương pháp học tập và quản lý thời gian hiệu quả.

## 3. NHỮNG KINH NGHIỆM THỰC TIỄN ĐƯỢC TÍCH LŨY

* Sử dụng Spring Boot để xây dụng ứng dụng.
* Sử dụng Docker Desktop để quản lý các container.
* Sử dụng Postman để kiểm thử các phương thức đã tạo.
* Quản lý code và báo cáo bằng Markdown trên Gitlab .
* Nắm được những công nghệ mới và những lợi ích, hiệu quả mà nó mang lại.
* Kỹ năng giao tiếp và giải quyết các vấn đề phát sinh.
* Môi trường làm việc thực tế tại công ty.
* Đóng góp ý kiến, đặt câu hỏi trong cài đặt và xây dựng ứng dụng.
* Áp lực công việc đòi hỏi phải tập trung, nghiêm túc làm việc, cố gắng thực hiện và hoàn thiện đúng tiến độ.
* Trách nhiệm trong công việc và trong sinh hoạt tại cơ quan.

## 4. CHI TIẾT CÔNG VIỆC ĐÃ ĐÓNG GÓP CHO CƠ QUAN

* Xây dựng ứng dụng Elasticsearch có cài đặt bộ phân tích Tiếng Việt, dữ liệu được quản lý trong Elasticsearch. Tạo phương thức thêm sửa xóa, tìm kiếm dữ liệu, tìm kiếm dữ liệu theo từ khóa được lập trình bằng ngôn ngữ Java trên dự án Spring boot.
* Lập trình các source code từ đơn giản đến nâng cao, tăng khả năng logic, phân tích, giải quyết vấn đề, sẽ là một khía cạnh để cơ quan đánh gia kiến thức sinh viên đã được học và áp dụng vào thực tế như thế nào.
* Hoàn thành các yêu cầu nghiệp vụ, sẽ là một đóng góp cho công ty cải thiện các chức năng sẵn có và sẽ bổ sung các chức năng chưa có cho phù hợp yêu cầu.