Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

CJ1 Project Documentation

Atif Saeed

Saleh Gamil

Hajras Algdairy

Chatbot

Fakultät Technik und Informatik
Department Informatik

Faculty of Engineering and Computer Science
Department of Computer Science

Motivation for the idea.

Sometimes we spend our time by chatting on different chatterboxes available on internet, google, facebook and different applications, to make one of them was indeed an interesting idea. Chatbot is a computer program designed to simulate conversation with human users, especially over the internet. Our idea is to Integerate a chatbot for the HAW students (useful for the new students as well for the existing students) in order to help them find important and useful information i.e (scholorships, contacts, study plans, hamburg welcome center e.t.c) which they might need during the course of their studies.

Programmes we used

We used HTML, javaScript to build a chatbot Locally for the HAW website. When you open the chatbot you can see the Welcome message. To perfrom action the user is supposed either to write a keyword or a full sentence which will then try to find these keywords or sentence which are already stored in oure database. We will use MySQL to build our database and use the Java and SQL programming Languages to access this database. In order to search the text or keyword for chatbot search we will use the Elastic search, this tool will help us perform the chatbot search more efficiently.

Different Programms or tools used

Hibernate

Hibernate is a middleware used for object-relational mapping and for performing efficent object persistence. To understand the architecture of Hibernate, let's first take a look at how Hibernates fits in the developement of an application, between an application and the database server.

- Java application code comprises of all the classes that define the business logic of application. These classes communicate with the Hibernate.
- Next, the Hibernate and its core interfaces, using which we could persist(store and retrieve) the objects of our business layer classes by communicating with the database server.
- Hibernate uses the Java core API, Java Database Connectivity(JDBC), Java Transaction API(JTA), Java Naming and Directory Interface(JNDI) to communicate with database in order to persist the state of object by performing create, read, update, delete operations.

The architecture of Hibernate comprises of few important interfaces.

1. Configuration
2. Session
3. Session Factory
4. Transaction
5. Query

6. Callback Interfaces

Spring

Spring framework is an application framework and inversion of control container for the java platform. It's core features can be used by any Java application, but there are extensions for building web applications on top of Java EE(Enterprise Edition) platform in Spring

Modules of Spring framework:

- Core
- AOP (Aspect oriented programming)
- DAO (Data access object)
- ORM (Object Relational Mapping)
- MVC (Model view Controller)
- Context
- Expression language

Advantages of using Spring:

- Open source
- Lightweight and fast
- Moduler structure
- Low coupling
- Resuable software
- Stable and lots of resources
- Projects that make our life easier like Spring security

JPA Repositories

Java Persistence API(JPA) is the standard way of persisting Java objects into relational databases. JPA consist of two parts: a mapping subsystem to map classes onto relational tables as well as Entity Manager API to access the objects, define and execute queries.
Spring Framework offers sophisticated support for JPA to ease repository implementation. It helps to integerate with the Spring transaction abstraction, translate JPA-specific exceptions into Spring's DataAccessException hierarchy.

Lombok

Lombok is used to reduce boilerplate code (biolerplate code refers to sections of code that have to be included in many places with little or no alteration) for data objects e.g it can generate getters and setters for those object automatically by using Lombok annotations. The easiest way is to use the @Data annotation.

Features:

- @Getter, @Setter, @ToString, @EqualsAndHashCode, @Data
- @Cleanup, @Synchronized
- @Log
- @Delegate
- @NoArgsConstructor, @AllArgsconstructor

Details of how @Getter and @Setter can be used

- Automatic Generation of getters and setter
- Method will be public unless you explicitly specify and Access Level
- @NotNull for nullity check
- @Getter or @Setter annotation on class

Gradle

Gradle is a general purpose build system. It comes with a rich build description language (DSL) based on Groovy. It supports build by convention principle, but it is very flexible and extensible. It has built-in-plug-ins for Java, Groovy, Scala & Web, derives all the best and integrates well with Ivy, Ant and Maven.

Gradle is being used by google, hp, NETFLIX, Oracle, paypal, ebay and many other leading international companies.

Features

- Declarative builds and build by conventions.
- Language for dependency based programming and many wazs to manage dependencies.
- Groovy as a base language allows imperative programming
- Deep and rich API for managing projects, tasks, dependency artefacts and much more.
- State of the art support for multiproject builds.
- Free and open source.

Advanced features

- Parallel unit test execution
- Dependency build
- Incremental build support
- Dynamic tasks and task rules

Flyway

Flyway is the open source tool that makes database migrations easy.
It is based around 7 basic commands: Migrate, clean, info, Validate, undo, Baseline and Repair the detail can be found here
It has a command line clientm, a Java API for migrating the database on application startup, a Maven plugin and a Gradle plugin.

How it works

- Metadata table SCHEMA_VERSION
- Each change is recorded in metadata table
- Scans classpath for available migrations. SQL or Java.
- The mighrations are then sorted based on their version number and applied in order

Flyway is version control for the database. It evolves the database schema easily and reliably acroos all instances. It's been designed from day one to be easy to setup and simple to master. Flyway favors simplicity and convention over configuration. It achieves this by focusing on one problem only and solving it well.

Thymeleaf

Thymeleaf is a Java library. It is an XML/HTML5 template engine able to aply a set of transformations to template files in order to display data and text produced by the applications.

It is better suited for serving web applicatins, but it can also process any XML file .

The main goal of Thymeleaf is to provide an elegant and well formed way of creating templates. In order to acheive this, it is based on XML tags and attributes that define the execution of predefined logic on the DOM(Document Object Model), instead of explicitly writing that logic as code inside the template.

The main features of Thymeleaf are

- Java, DOM based
- Online (web) or offline (email, XML data...)
- Produces XML, XHTML or HTML5
- Full Spring MVC integration.
- Configurable and extensible
- Static prototyping abilities

Vue

Vue.js is a progressive framework that focues on building user interfaces. Unlike other frameworks, Vue is designed from the ground up to be incrementally adoptable. The core library focuses on the view layer only, and is easy to pick up and integrate with other librarier or existing projects. Vue is also capable of powering single-page Applications when used in combination with modern tooling and supporting libraries. Details can be found here

Vue.js offers lot of functionality for the view layer and can be used for building powerful single-page webapp. Follwing are the list of features:

- Reactive Interfaces
- Declarative Rendering
- Data Binding
- Directives
- Template Logic
- Components
- Event Handling
- Computed Properties
- Filters

Vue.js core library is small in size. This ensures the overhead which is added to your project by using Vue.js is minimal and website is loading fast.

How to use Vue.js

There are different ways to include Vue.js in a web project:

- Use CDN by including <script> tag in HTML file
- Install using Node Package Manager(NPM)
- Install using Bower
- Use Vue-cli to setup the project.

Advantages and Disadvantages of Vue

Advantages

- Small size(Vue weigh only 18kb and Vue ecosystem is also small and fast)
- Ease of Understanding and developement
- Simple Integration
- Comprehensive documentation
- Flexibility
- Facilitate two way communication

Disadvantages

- Closed community developement
- Language barrier
- Too felxible can be problematic

Elasticsearch

Elasticsearch is a highly scalable open source full-text search and analytical engine. It allows to store, search, and analyze big volumes of data quickly and in near real time. It is generally used as the underlying engine/technology that powers applications that have complex search features and requirements. The detail explanation of Elasticsearch can be seen .

Use-Cases of Elastic Search

1. An online web store where customers are allowed to search for product that we want to sell. In this case, we can use Elasticsearch to store the entire product catalog and inventory and provide search suggestions for them.
2. Collect log or transaction data and want to analyze this data to look for trends, statistics, summarizations, or anomalies. In this case, we can use part of (Elasticsearch/Logstash/Kibana stack) to collect and feed the data into Elasticsearch.
3. It can be used for price alerting platform which allows price-savy customers to specify a rule „I am Interested in buying a specific electronic gadget and i want to be notified if the price of gadget falls below $X from any vendor within the next month“. In this case it is possible to scrape vendor prices, push them into Elasticsearch and use its reverse-search(Percolator) capability to match price movements against customer queries and eventually push the alerts out to the customer once matches are found.

Advantages and Disadvantages Elasticsearch

Advantages

1. Elasticsearch offers the most powerful full-text search capabilites
2. It Performs linguistic searches against documents and returns the documents that match the search condition.
3. Stores complex entities as structured JSON(Javascript Object Notation) documents and indexes all fields by default, providing a higher performance.

4. It stores a large quntity of semi-structured (JSON) data in distributed fashion. It also attempts to detect the data sturcture, index the data present and makes it search-friendly.
5. Elasticsearch is API-driven, so actions can be performed using a simple Restful API.
6. Elasticsearch records any changes made in transactions logs on multiple nodes in the cluster to minimize the chance of data loss.

Disadvantages

1. Elasticsearch does not have multi-language support in terms of handling request and response data(only possible in JSON) unlike in Apache Solr, where it is possible in CSV, XML and JSON formats.
2. Elasticsearch also has a problem of Split brain situations but in rare cases.
3. It is not as good at being a data store as some others. For small use cases, it will perform fine. If streaming in TB's of data every day, you will find that it either choes or loses data.
4. Elasticsearch is way more powerful and flexible, but it's learning curve is much steeper.