Airline Search Engine

Minjian Li, Lei Chen

Computer Science Department, Washington State University

`minjian.li@wsu.edu`

`lei.chen2@wsu.edu`

Abstract

As the improvement of the global living standards, there is an upward trend in tourism. Simultaneously, multiple kinds of travel pattern, especially self-guided tour, are formed, and become increasingly popular. Therefore, our study group intends to implement an airline search engine which systematically conforms to user's wishes in terms of assistance of quickly searching the information from airlines, timing ravel schedule precisely, and effective arrangement of traveling routes. Besides, the Airline Searching Engine we decide to develop is a searching engine using the MVC (Model View Controller) mode that outstandingly protect database from hackers (compare with C/S structure). During the implementation of our study, the essential developing platform is MyEclipse 10, MySQL and WebStorm 10.

Keywords: **travel, self-guided, airline, searching engine**

I. INTRODUCTION

With the rapid development of the economy and the increase of cross-regional cooperation projects, people tend to travel by air, saving energy and time, and ensuring safety.
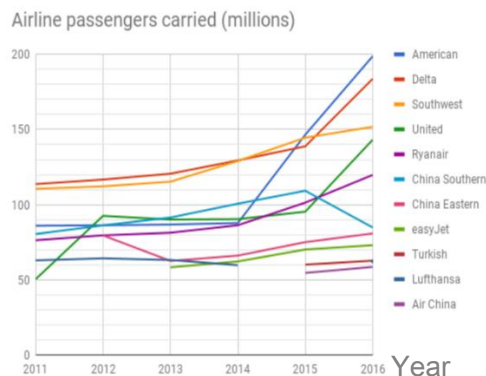

Fig. 1: Passenger capacity of airlines from 2011 to 2016

Sources: All the Data from chart comes from [1]-[6]

From the chart (Fig. 1), we can see that the global passenger capacity of the world's major airlines has grown steadily from 2011 to 2016, specially, the American Airline exponential growth from 2014 to 2016. We can conclude that the demand for airplanes is increasing all over the world.

In order to improve the efficiency of passengers' arrangement of flights and their time, we make this airline search engine. In this project, we use public-available datasets which contain the detail of the flights to make an airline data search engine supported by, Java web SSH structure, MySQL, and MVC pattern. Additionally, our airline search engine is based on "hub-and-spoke" system which is a typical airline system in America [7]. For each airline, hubs are the center of its routes. They are always connected smaller cities are called "spokes" and each one connects at least one hub. Users can use this application to find out facts/trips with requested information/constraints as following shows: with the help of this application, users can get the list of airports operating in the country X (hub city); the list of airlines having X stops (Hub or Spoke); lists of airlines operating with code share; lists of active airlines in the United States.

II. CHALLENGES

Assisting users to make better time planning and to improve work efficiency is our original intention to do this project. However, it is difficult for us to deal with a huge amount dataset of airline, airports and routes; each of these datasets is detailed. For example, the relation of airline, it includes detail such as "Airline ID", "Name", "Alias", "IATA", "ICAO", "Callsign", and "Country".

Another challenge is that there are often a lot of dirty data in our public-available datasets. Some of them are tuples with null value in one or two attributes, and some are incorrect data. Sometimes, there are parts of missing information. If such information is not revised correctly, it can affect the quality of the

data. Therefore, we try to fix it by considering association rules and links mentioned by Professor Yinghui Wu in class.

III. SOLUTION

In order to facilitate the processing of large amounts of data, we classify the data based on different regions. Defining the type of each data and choosing the suitable length of the data which may appear, we could help relational database using as less memory as possible. Another purpose of our classification of data is to make one-to-one comparisons in a small range.

When considering dealing with dirty data, we come up with three ways for data screening. The first one is deleting data with missing items, but the premise is that the amount of data is large enough. The second one is manual completion. This method is inefficient in many situations and brings a lot of labor costs into it. Besides that, we try to use mathematical calculation models to find the average or the predicted value of the same type of data, and replace it with the missing information, this function does not influence the quality of the overall data.

IV. DEVELOPMENT TOOLS

In this part, we introduce the tools used in our project. We use Java to write programs because it has powerful features, and it is easy to use. Java is a computer programming language with cross-platform, object-oriented, and generic programming features that are widely used in enterprise-class Web application development and mobile application development. Today's Java ecosystem includes multiple Java web servers and application servers. Although Apache Tomcat, Simple, Jo! Rimfaxe Web Server (RWS) and Project Jigsaw occupy Web server space, WebLogic, WebSphere and JBoss EAP play an important role in the field of business application servers [8]. It is suitable for Java SSH framework which supports the MVC pattern that we used in our project very well.

For demo design section, MyEclipse remarkably support a wide range of applications and different kinds of programming language such as C/C++, PHP, Java, Android and so forth. It is a powerful enterprise-level integrated development environment developed on the basis of eclipse and its own plug-in. MyEclipse is very powerful, and it supports a wide range of applications, especially for various open source products; it is mainly used for the development of Java, Java EE and mobile applications.

For database design, MySQL has become the most popular open-source database in the past due to its high performance, low cost, and good reliability. Airline searching engine is required to conduct both relational database by utilizing MySQL.

For user interface, WebStorm, a JavaScript development tool owned by JetBrains, is honored as the most powerful HTML5 editor by a majority of developers. It has intelligent code completion, code formatting, code debugging and other advanced features.

V. DATABASE DESIGN

MySQL is the platform used in this project to give a support for the airline search engine. In order to resolve the issue—a huge amount datasets with more than 20,000 tuples per each—previously mentioned in this report, we decided to choose suitable domain with reasonable length to store the data in database. Also, classification would be considered as a useful approach for solving this problem.

A. Domain of Attributes

Domain is the element type of a relation, also known as data type [10]. Each attribute in a relation has its own domain, such as "int", "char", "varchar", "text", and so forth. Generally, different type of domain is used for different type of data. For example, "Int" is used for numeral data, "char", "varchar", and "text" are used for letters, words and strings. It is worth to notice that the value range of each domain is different from others. In other words, although two domains can be used for representing the same data type, they have variety of storage size in database. Unsuitable storage size could negatively affect the read/write speed of disk and the cost of memory. Therefore, accurately picking up the domain for each attribute is of vital importance. One of the relation designs from our project is shown in Table 1 below.

TABLE I

THE DOMAIN FOR EACH ATTRIBUTE IN THE RELATION AIRLINE

Name	Type
AirlineID	smallint
Name	varchar
Alias	varchar
IATA	varchar
ICAO	varchar
Callsign	varchar
Country	varchar
Active	char

In Table 1, the first column is the name of each attribute, and the second column is domain. As can be seen, "AirlineID" is defined as "smallint". The reason we use "smallint" rather than "int" is that according to [9], the value range of "int" is from -2,147,483,648 to 2,147,483,648 which is extremely large, while the value range of "smallint" is only from -32,768 to 32,767; in relation airline, the value of "AirlineID" is less than 30,000. In addition, "varchar" is used for defining most of the attributes in this relation because most values of attributes are words and strings. Another reason we use "varchar" is that it is flexible for data storage because its value range can be dynamically adjusted according to the length of the data. Based on the above rules, we can design the rest of relations.

B. Classification

This project, the airline searching engine, uses tuple-by-tuple function to achieve the goal required from users. However, the original datasets used in this project have more than 20,000 tuples in each relation. As a result, time complexity of this search engine is high. Fortunately, our group discover by analyzing our datasets that each relation can be separated into several sub-relations using classification mentioned by our professor, and we separate each relation into sub-relations based on countries because we notice that all relations have "countries" as its attributes. After the classification of each relation, whenever the engine is searching on data, it only scans a small scale of relation according to the name of the country inputted by users, which successfully increase the efficiency of the search engine.

VI. PROJECT MODEL

Considering the search engine is used for publics, the security of database should be protected well. Therefore, MVC design pattern standing for model-view-control model is accepted in this project [11]. One of the main factors for MVC used as the model of this project is that it divides the system into three parts named model section, view section, and controller, shown in Fig.2, which means that the users cannot access to the database, and each of request from users should be identified as legal or illegal by controller. If the request is illegal, the controller will cease the process immediately and return a warning message to users; otherwise, the process continuously runs [11]. Thus, using MVC pattern avoids situations that the records in database accidentally
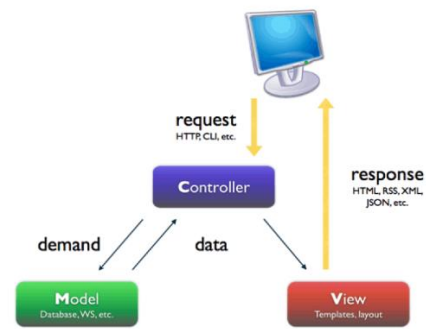
modified, added and removed by users.

Fig. 2 the MVC pattern

Source: http://stackoverflow.com/questions/5966905/which-mvc-diagram-is-correct-web-app

A. Model

The model is part of our system that manages all tasks in terms of data updated including data query, data modification, data insert and data removing. Also, this part can successfully decrease the complexity of the code that the programmer has to write [11]. Moreover, it is responsible for returning the data from database to the controller section. Typically, the model is the only section that is allowed to access to the database of our project.

B. View

The view is the visual of our search engine. It is responsible for showing the searching result to users by the HTML elements. This section basically receives the responses sent from the controller. Generally, what are the searching results shown to users depends on what the controller responses. Furthermore, the techniques that our group mainly uses to develop the view section are HTML, CSS and JavaScript.

C. Control

The controller is responsible for events handling [12]. In this project, control section handles and manages the interaction between users and our system. It responds to users' requests, interactions with the model section and the decision of what view section should be generated and displayed [11]. In this part, we decide to package the searching results as JSON data responded by the model section; then, the view sections will resolve the JSON data using AJAX functions in JavaScript.

VII. PROJECT FRAMEWORK

In the demo coding process in this course project, Java Web SSH Framework would be used for the implementation of the MVC pattern. According to an article "Integration of Struts, Spring and Hibernate for an E-Commerce System", SSH structure outstandingly support the MVC development pattern; it provides a lot of services matching to the MVC pattern, such as UI tags and data validation [12]. Therefore, we believe that Java Web SSH Framework can be successfully support the airline search engine.

A. Struts

Strut is an infrastructure of the project; it belongs to the controller part of the MVC pattern. In this project, strut acts as a bridge among model section and view section in controller of MVC pattern, providing services including receiving requests from model section, a logical approach to deal with the affairs in model section, and returning the results to the view section for display.

B. Spring

Spring manages both struts and hibernate in SSH framework [12]. Like struts, spring works at controller section in MVC pattern as well. Basically, spring possesses wonderful capability of simplicity, testability and interconnection which is convenience for developers [12].

C. Hibernate

Hibernate is a supporter in our project which works at model sections in MVC pattern and works as a connector between database and the search engine. There are five core interfaces in hibernate: Session, Session Factory, Transaction, Query and Configuration [12]. Through these interfaces, not only data can be read and write into the database, but also the transactions between model section and database can be controlled. Here is the effects of the five interface in hibernate used in our project:

Session: This interface can communicate with databases [12].

Session Factory: It used for initializing hibernate [12].

Transaction: used for transaction processing [12].

Query: It is responsible for executing several of database queries [12].

Configuration: This interface is responsible for configuring and launching the hibernate [12].

VIII. CONCLUSION

In this project, we believe that this airline search engine will improve the efficiency of passengers' itinerary in the future. It can quickly search the information of airlines because we classify the database into several smaller databases to increase the speed of the tuple-by-tuple search function. Also, considering the protection of database in this project, we discuss the positive influences by using MVC pattern to develop the search engine. Moreover, this paper demonstrate how MVC can be supported by Java Web SSH framework, and how Struts, Spring and Hibernate coordinate with each other to marvelously work on MVC pattern.

This course project is required for all students by Professor Yinghui Wu. In this project, Minjian Li and Lei Chen collaborate to accomplish the related tasks. Here is the task assignment table of this project:

TABLE II

THE TASK ASSIGNMENT FOR EACH GROUP MEMBER

Team member	Tasks assigned	Rate of completion
Lei Chen	Research(Problem)	100%
	Data Analysis	100%
	Database Design	100%
	Document(Presentation)	100%
	Document(Final Paper)	100%
Minjian Li	Research(Solution)	100%
	Demo Coding	95%
	UI Design	100%
	Document(Presentation)	100%
	Document(Final Paper)	100%

According to the table 2 above, Lei Chen takes charge of finding the outside sources related to the background information of our topic, the analysis of the raw data giving by our professor, database design and documents (part of presentation and final paper). Lei Chen is a hard-working partner, and he did a great job on each part, especially database design. At the very beginning of this part, is struggling with how the data that there are more than 20,000 tuples on each relation can be efficiently recorded, read, and/or written in a relational database. In order to solve this problem, Lei Chen stayed up late finding research every night

before he addressed this problem by using accurate data type, for example using varchar instead of Sting, to design the database of our search engine. Moreover, he would like to help his teammate regarding certain technical terminology if necessary. To sum up, Lei Chen is dedicated to this project.

The table above also shows that Minjian Li is responsible for five sections in terms of research for solution, implementation of demo coding, UI design, and document for both presentation and final paper. He is always strict in his own during this project, and achieves an excellent result in his sections, while there are certain flaws on demo coding section. Although there are several bugs needed to be fixed in his demo coding section, he still tries hard to deal with such bugs. Also, he always discusses with his teammate, sharing his own ideas, whenever the group is facing a challenge. Overall, his patience is appreciated.

REFERENCES

[1] (2017) The Flight Global website. [Online]. Available: https://www.flightglobal.com/

[2] Global Air Traffic & Top Airline 2016 (IATA). (2016) News homepage on IMM International. [Online]. Available: https://www.imm-international.com/global-air-traffic-top-airlines-2016-iata/

[3] Scheduled Passengers Carried. (2015) World Air Transport Statistics 59th Edition homepage on IATA. [Online]. Available: https://web.archive.org/web/20150618070729/http://www.iata.org/publications/pages/wats-passenger-carried.aspx

[4] Scheduled Passengers Carried. (2015) World Air Transport Statistics 58th Edition homepage on IATA. [Online]. Available: https://web.archive.org/web/20150102034843/http://www.iata.org/publications/pages/wats-passenger-carried.aspx

[5] Scheduled Passengers Carried. (2013) World Air Transport Statistics 57th Edition homepage on IATA. [Online]. Available: https://web.archive.org/web/20131113175503/http://www.iata.org/publications/pages/wats-passenger-carried.aspx

[6] Scheduled Passengers Carried. (2012) World Air Transport Statistics 56th Edition homepage on IATA. [Online]. Available: https://web.archive.org/web/20121217084752/http://www.iata.org/publications/Pages/wats-passenger-carried.aspx

[7] Z. Lawrence J., *The Hub-and-Spoke System*, Ed. United States of America, 2010, pp. 118-120.

[8] F. Tetiana. (2017) To Java or Not to Java: What Statistics Show? homepage on Jelastic. [Online]. Available: https://jelastic.com/blog/java-stacks-usage-statistics-2017/

[9] (2017) The Microsoft website. [Online]. Available: https://docs.microsoft.com/en-us/sql/t-sql/data-types/int-bigint-smallint-and-tinyint-transact-sql?view=sql-server-2017

[10] H. Garcia-Molina, J. D. Ullman, and J. Widom, *Database Systems: The Complete Book*, 2nd ed., Upper Saddle River, 2009, pp. 23.

[11] D. Pop* and A. Altar, "Designing an MVC Model for Rapid Web Application Development," *Procedia Engineering*, vol. 69, pp.1172-1179, 2014.

[12] N. Munsi, N. Sehrawat, and M. Jain, "Intergration of Struts, Spring and Hibernate for an E-Commerce System," *International Journal of Computer Science and Mobile Computing*, vol. 3, pp.853-859, Oct. 2014.