Tourism Big Data Construction and Application Strategy Analysis -- Take Qiqihar Tourism Cultural Resources Publicity as an Example

Xiangkun Li¹, Zhiwei Tian²*, Mingli Tian, Yunfeng Han, Ang Li, Xuejuan Zhang, Di Yan and Xinxin Chen

¹Personnel Department, Qiqihar Medical University
²Foreign Language Department, Qiqihar Medical University
³Qiqihar Medical University
⁴Qiqihar Municipal Administration of Culture and Tourism
⁵Qiqihar University
No.333, Bukui Street, Jianhua District, Qiqihar City, Heilongjiang Province, 161006, P.R. China

*Corresponding author e-mail: tzwteller@163.com

Abstract. With the rapid promotion of big data on the Internet, the construction of "big data" in all walks of life has begun to take root. With the continuous rise of smart scenic spots and smart tourism, it is possible to obtain massive tourism big data. The purpose of this paper is to analyze the construction and application strategy of tourism big data based on the promotion of tourism cultural resources in Qiqihar. First of all, it analyzes the advantages and applicability of spark big data platform, and expounds the system module functions from three main aspects: background management module, information service module and big data analysis module. After studying the algorithm of scoring function, it selects the most commonly used centralized data storage engine on MySQL 5.6.17 for research and test. The experimental results show that the InnoDB storage engine takes 10 to 100 times more time than other engines, and the archive storage engine has the advantages of compressing storage and saving space, which is very suitable for storing visitors' access information. The platform collects the cultural resources of Qiqihar, a new tourism city, through fine classification and index of resources, graphic and multimedia.

Keywords: Big Data Construction, Application Strategy, Tourism Culture, Resource Publicity, Qiqihar City

1. Introduction

With the improvement of people's living standard and the change of life concept, tourism seems to be an indispensable part of people's life. The analysis of tourism data is very important, because the rapid development of tourism data, the rich accumulation of tourism data, and the analysis of tourism data promote the development of tourism. Tourism developers predict the types of future tourism resources.
by analyzing the data of tourism resources, and focus on the development. The types of tourist attractions will decide to make the most of commercial interests and help tourism development. Tourists can also improve the quality of tourism by analyzing and comparing the products with high cost performance.

Qiqihar is an important industrial base in our province. It has been in the history for thousands of years and is a new tourist city. The development of tourism in Qiqihar will bring more opportunities through the construction of tourism big data, which will solve some problems for the development of tourism industry. Tourism in Qiqihar is facing some obstacles. The development and application of tourism big data, the promotion of tourism industry and the construction process of regional tourism big data are imperative.

Sanya, as one of the first units to establish a national "one person tourism" demonstration area, is innovating and changing. Lieyun Ding had built the tourism motivation of Sanya's smart tourism city. Combined with the current development of Sanya's tourism industry, it puts forward Sanya's tourism integration and smart city construction, analyzes the advantages and disadvantages of Sanya's smart tourism city construction, and puts forward how to solve these problems Using the Internet and big data will become the strategy of Sanya's smart city construction [1]. Osman Taylan in the big data environment, the key of tourism intelligent information system is to use semantic description and link construction to complete the intensive, intelligent and unified management of tourism. This paper proposes a hierarchical semantic description framework of tourism related data, which consists of four layers: metadata layer, ontology layer, related data layer and data application layer. The main functions of these layers and their interrelations are described. This paper proposes a method of RDF linking between tourism related data, and establishes the semantic association of tourism related data [2].

This paper takes the construction of digital platform of Qiqihar cultural resources as the core, through the summary and analysis of the technical points of the project, first introduces the core technologies involved in the project construction, mainly including Java EE related technology and spark big data processing platform technology. Then, taking the platform construction as the core, it introduces the project demand analysis, design and implementation. The implementation part includes the construction of digital platform of urban cultural resources, the implementation of big data analysis engine, the implementation of system core architecture and components, and the implementation of spark big data engine.

2. Proposed Method

2.1 Advantages and Applicability of Spark Big Data Platform

The big data analysis engine of urban cultural resources is a data analysis and recommendation engine based on the big data platform of urban cultural resources [3]. Considering the above three big data platforms and characteristics, as well as the actual needs of the system, we finally take the spark framework platform as the basic framework of the big data engine. The analysis is as follows:

(1) Data analysis tasks include two categories: access statistics and user information recommendation. Neither of these tasks requires high real-time performance, so spark batch processing can be used to accomplish this task.

(2) The visit volume of the website mainly comes from the day, and the visit peak will appear in a certain period of the day. However, from 0:00 to 6:00 at night, the server is in a low load state. You can choose to execute the spark batch processing task and calculate the results at this time [4].

(3) The city culture digital platform is developed by Java EE technology, while Scala runs in JVM virtual machine, which can be seamlessly compatible with Java language, and spark also provides Java API.

(4) Compared with Hadoop, spark has higher performance. When the demand is small, we can use Spark's standalone mode; when the demand is not satisfied, we can use spark cluster, and this extension hardly needs to modify the code.
(5) Spark supports more programming models than Hadoop, and has a good extension for real-time analysis (sparkstreaming). Therefore, even if the business requirements change, the big data analysis engine using spark does not need to be redesigned and built.

2.2 System Module Function
The digital platform of urban cultural resources includes several core modules: background management, information service and big data analysis. These modules divide the whole system from the core functions of the system [5]. The functions and contents of each module are briefly introduced as follows:

(1) Background management module: the background management module provides three categories of users, namely, ordinary tourists, cultural resources partners, administrators [6]. Among them, ordinary tourists can participate in cultural tourism information resource review, online purchase and other functions, and cultural resource partners have the right to add various cultural tourism information; administrators have the ability to review resources submitted by cultural resource partners and can modify information such as resource description and priority in cultural resource information database. Such division ensures that the digital platform of urban cultural resources has the ability of open access.

(2) Information service module: the information service module is mainly reflected in the front page of the platform website, which pushes fresh and valuable cultural tourism information to users. It has the ability to enrich the platform content and keep the platform active. In addition, the module provides weather forecast function [7, 8].

(3) Big data analysis module: This module analyzes the user's habits by analyzing the user's browsing history and recommends the tourism cultural information that they may be more interested in, so as to improve the user's experience of using the platform and enhance the platform's stickiness to users [9]. In addition, from the perspective of merchants, the guiding role of the module can bring more traffic and consumption to merchants, and the module has the ability to enhance the commercial value of the whole platform [10].

2.3 Scoring Function
(1) Scoring function based on Bayesian statistics
Both BDe score and BIC score are based on Bayesian statistics. In BN structure learning, structure g and parameter θ are regarded as random variables. The Bayesian network is composed of n variables \( x=\{x_1,x_2,\cdots,x_n\} \). Suppose that variable \( x_1 \) has RI values, its parent node \( Pa(x_1) \) has \( q \) values, parameter \( \theta=\{\theta_{ik} | i=1,\cdots,n;j=1,\cdots,q;k=1,\cdots,r_i\} \), Assuming that the network structure \( g \) obeys a certain distribution with a prior probability of \( P(G) \), for a given data set \( D \), the maximum posterior probability of the network structure can be obtained from the Bayesian formula, that is:

\[
P(G|D) = \frac{P(D|G)P(G)}{P(D)} \tag{1}
\]

(2) Scoring function based on information theory
The minimum description length is a scoring function based on information theory, and BIC scoring is a special case [11]. The evaluation function considers the description length of data and the description length of network structure, and considers that the network with the smallest sum of the two description lengths is the optimal network. Data description length indicates the similarity between data and network, while network structure description length indicates the complexity of network structure [12]. Generally, the function of the number of parameters is regarded as the penalty term of network structure complexity:

\[
C(G) = \frac{1}{2}\log m \sum_{i=1}^{n} (r_i - 1)q_i \tag{2}
\]

3. Experiments
3.1 Analog Location
Qiqihar, located in the Northeast Plain of Heilongjiang Province, is a new tourist city. Qiqihar has a long history, rich history and culture, and beautiful natural scenery. Zhalong Nature Reserve is the largest bird reserve in Qiqihar city. It also provides a breeding ground for the beautiful red crowned crane. Therefore, Qiqihar city is known as "bird flying, flower fragrance and Hexiang" at home and abroad. As a new tourist city, Qiqihar has unique tourist attractions. For example, the world-famous nature reserve, Red Crowned Crane town and Zhalong Nature Reserve are surrounded by water, with magnificent and colorful scenery; Longsha park has beautiful scenery, pavilions, lakes and mountains; Angang West cultural site, Dacheng temple, mosque, Guandi temple and other historical sites in the Neolithic age are especially popular with tourists, and a large number of foreign tourists come to visit each year.

3.2 Database Design Steps

Database design can be basically divided into the following four stages:

1. Requirements analysis stage. At this stage, it is necessary to clarify the application requirements of each user. Clarify the information requirements, processing requirements, security and integrity requirements in data management.

2. Conceptual design stage. The main task of this stage is to transform the real world demand into the information world model, and the E-R diagram needs to be completed in this stage. In this stage, we need to pay attention to the definition of entity, relation, code, attribute and integrity constraint.

3. Logic design stage. In this stage, when the transformation of E-R diagram is completed, we need to pay attention to the combination and decomposition of the relations in E-R, so that the database design can meet the requirements of the third paradigm.

4. Physical implementation phase. The transformed E-R diagram is implemented into a data model supported by a specific database, including the design of database tables, physical storage arrangement of database, etc. We need to pay attention to setting appropriate index to improve the efficiency of database query.

4. Discussion

4.1 Database Integration Analysis

The integration of big data engine and digital platform adopts the way of database integration. The database uses platform database, and the selection of data engine is compared with the actual research. This paper selects the most commonly used centralised database storage engine on MySQL 5.6.17 for research and test. This paper tests the time-consuming situation of inserting data 10000 times by different MySQL storage engines in a single machine environment and makes a statistical chart. In order to eliminate the impact of environmental changes on the actual performance of the database, we performed 10 rounds of iterations on each group of experiments, and the final results are shown in Figure 1.

![Figure 1](image_url)  

**Figure 1.** Time consumption of continuous insertion of 10000 pieces of data by MySQL storage engine
It should be noted that the test of InnoDB storage engine shows that it takes 10 to 100 times more time than other engines, so the statistical chart is not shown. The end result is the archive storage engine[13]. In addition, because the archive storage engine has the advantages of compressing storage and saving space, it is very suitable for storing visitor access information.

The platform collects user data in the following format:

**Table 1.** Format description of visitor user access information

| Name | User: userID | Resource: resID |
|------|--------------|-----------------|
| Example | For example: wpxidian | Such as: 1242 |

Corresponding to user a clicking to access resource RESA, there is a message to insert into the database as follows:

(A:userID,resA:resID)

### 4.2 Tourism Cultural Resources Publicity Website

The page components included in the mobile website include but are not limited to: title bar, shortcut menu, primary classification area, secondary classification area, picture rotation, content list, popular recommendation, paging, breadcrumb, detail graphic display, user comment browsing area, landing and other page components.

The title bar has certain navigation ability, and can directly return to the first page and the second level page. At the same time, the right side can directly open the shortcut menu. The design of the module is simple and natural. The design principle of shortcut menu is to show the whole content of mobile website, which is an important entrance to the website.

The first level classification is similar to the second level classification, which is placed under the title bar of the page. It is simple and clear, convenient for users to select and jump to the content types they are interested in. Pictures are rotated on the home page. They are usually used to display advertisements and recommend website contents to users. They support automatic playback and finger sliding playback. The open-source swipe.js library is used, with excellent performance and elegant appearance.

The content list is a list preview display of the website content in the way of mixed arrangement of text and text. Each entry can be divided into three parts: title, content summary and theme picture. Users can click an entry to browse the details of the corresponding entry.

**Table 2.** Contents of the title bar

| Title | Content outline |
|-------|-----------------|
| Item 1 | Zhalong Nature Reserve has a Reserve Management Bureau, which is built in the west of the central part of the reserve. Its main task is to protect waterfowl and cranes in wetland. |
| Item 2 | The island is about 4 kilometers long from east to west, 3 kilometers wide from north to south, and covers an area of about 7.6 square kilometers. Beautiful natural environment. The completion of Longsha animal and plant park will greatly cater to people's return to nature. There are all kinds of rare animals from five continents living in Longsha animal and plant park, as well as more than 120 kinds of precious plants at 47 degrees north latitude. |
| Item 3 | It has opened up a military industry tourism route, and prepared various weapons, equipment and live ammunition, so that the majority of tourists can get to know the military industry at a close distance. |

Hot recommendation recommends hot content under a certain level of classification to users, which conforms to the overall style of the website and focuses on it. Paging is an important control of web
pages. The paging control of Qiqihar mobile website includes two ways: up and down page Jump and page direct, beautiful and intelligent, with good user experience; the detailed display of pictures and texts includes: picture roaming, introduction of scenic spots, list of sub scenic spots, relevant commodities, surrounding accommodation and other contents, with rich display forms, using the means including mixed arrangement of pictures and texts, tables and other means to provide good sensory enjoyment; User comment area includes comment browsing and I want to comment. Comment browsing has paging function and uses Ajax technology to load more content. I want to comment on the part including star rating control, easy to use.

5. Conclusions
This paper is a summary and analysis of the construction of digital platform of urban cultural resources and big data analysis engine. From the above summary, we can see that there are still improvements in the architecture of the platform; and we need to continue to study and improve the open resource API. In the future, the platform will do further work on the separation of view layer logic and business logic layer logic by adjusting the architecture. In the aspect of open interface, a reasonable and controllable restful API is developed to the third-party platform by studying a set of reasonable scheme and following jax-rs specification, so as to further promote resource sharing and multi win-win situation, and realize the economic value and social value of cultural resources. In terms of big data engine, we will provide more abundant data analysis means with big data engine as the core, and further study and improve the algorithm, so as to provide more accurate services for users and more powerful data analysis capabilities for administrators and businesses.

Acknowledgment
This work was supported by philosophy and social sciences research program of Qiqihar City, the project number QSX2019—18JL.

References
[1]. Lieyun Ding, Shengyu Guo. Study on Big Data-based Behavior Modification in Metro Construction [J]. Frontiers of Engineering Management, 2015, 2(2):131.
[2]. Osman Taylan. Contractor Selection for Construction Projects Using Consensus Tools and Big Data [J]. International Journal of Fuzzy Systems, 2017, 20(21):1-15.
[3]. Wang, Fan. Analysis on the Construction of Computer Data Processing Mode Based on Era of Big Data [J]. Journal of Computational & Theoretical Nanoscience, 2016, 13(12):10347-10351.
[4]. Guanlin Chen, Erpeng Wang, Xinxin Sun. An Intelligent Approval System for City Construction based on Cloud Computing and Big Data [J]. International Journal of Grid & High Performance Computing, 2016, 8(3):57-69.
[5]. Zhisheng Zhao, Jing Li, JunHua Liang. An Empirical Study of Medical Big Data and Hospital Information System Construction Based on Cloud Computing [J]. Journal of Computational and Theoretical Nanoscience, 2016, 13(12):10358-10363.
[6]. Zheng Xiang, Qianzhou Du, Yufeng Ma. Assessing reliability of social media data: lessons from mining TripAdvisor hotel reviews [J]. Information Technology & Tourism, 2017, 18(6):1-17.
[7]. Lizzie Hare. The Anthropocene Trading Zone: The New Conservation, Big Data Ecology, and the Valuation of Nature [J]. Behaviour & Information Technology, 2015, 6(1):109-127.
[8]. LI Jun-yi, TANG Jia, FENG Na. Tourists’ Spatio-temporal Behavior Based on Socially Aware Computing [J]. Scientia Geographica Sinica, 2015, 18(2):239-249.
[9]. MOU Naixia, ZHANG Hengcai, CHEN Jie. A Review on the Application Research of Trajectory Data Mining in Urban Cities [J]. Journal of Geo-Information Science, 2015, 17(10):1136-1142.
[10]. Wenhai H U, Jianping S , Feifei Y U . Temporal-spatial evolution patterns of regional tourism economic development in Anhui province [J]. Analytical & Quantitative Cytology, 2015, 5(1):28.

[11]. B M Sukojo, H Hidayat, D Ratnasari. Important Value of Economic Potency Mangrove Using NDVI Satellite High Resolution Image To Support Eco Tourism Of Pamurbaya Area (Case Study: East Cost of Surabaya)[J]. IOP Conference Series Earth and Environmental Science, 2017, 98(1):012035.

[12]. Huiwen H E, Hongbo C, University N . An Empirical Study of the Exhibitor's Ego-network and Its Influence on the Exhibitor's Performance [J]. Tourism Tribune, 2015, 30(5):97-106.

[13]. Yan Li. A Comparative Study of Tourism Business Models in the Internet Era. International Journal of Frontiers in Sociology (2019), Vol. 1, Issue 1: 99-108.