Research and Design of Dynamic Configurable Data Acquisition System in the Age of Big Data

Demei Gao

Shandong Vocational College Of Light Industry, Zibo, Shandong, China

Keywords: Big Data, Dynamic configuration, data collection, design.

Abstract: With the advent of the era of big data, data collection has become one of the main tasks of the social production and life sector. However, as current data collection still uses traditional collection methods, it is difficult to meet social development needs. Based on this, in order to achieve data dynamic configurability, this paper further studies the dynamic configurable data acquisition system. This data collection system can be personalized and dynamically configured according to the individual needs of users, and then distributed to the information collectors through social tools such as WeChat and Weibo, greatly improving data collection efficiency.

1. Research Background

1.1 Literature review

At present, in order to meet the needs of economic development, dynamic data acquisition infinite transmission devices have been widely used. The main reason is that dynamic data acquisition has the characteristics of remote controllable, gain amplification factor, and can provide users with the frequency and sampling frequency selection function according to the main characteristics of the information collected (Cheng, 2010). At the same time, dynamic configurable data can help people make better use of data resources, understand the true meaning of data, help users correctly analyze business models, and timely understand market development trends and security risk situations in commercial battlefields, and provide reference for enterprises to carry out informationization work. These features of dynamically configurable data make it an indispensable tool in data collection (Wang, 2018). Based on the above literature, Gong Sha et al pointed out that the current human society has entered the era of big data, and the data has shown an exponential growth. Therefore, for data collection, it is necessary to build a configurable automated crawler system to achieve efficient data collection and improve data collection and collection efficiency (Gong et al., 2018). Cao Lie further pointed out that the current data collection system mainly has two methods, one is to make a field custom collection form on the basis of database, the other is dynamic form generation, the form, the field definition, etc. are extracted, and then all stored in In the database table, when the user collects data, the system dynamically draws the collection form on the client (Cao, 2011). Deng Qingxu pointed out that by rebuilding the data acquisition system, the configurability and flexibility of data collection can be achieved. Therefore, Deng et al. proposed a solution that can reconstruct local data collection technology, further improve the configurability of the data collection system, broaden the application range of the system, and achieve high efficiency and real-time data collection (Deng et al. , 2010). Tian Xiaohong and others have realized the flexible configuration of data collection by researching cashback and dynamic configuration design through data data system, which will reduce the difficulty of system debugging to a certain extent (Tian et al., 2011).

1.2 Purpose of research

As the economy continues to develop, massive amounts of data are pouring into people's production and life. As the data continues to emerge, the era of big data is quietly coming. In this era, data collection has become the basis of data analysis. At present, there are many data collection tools, mainly including the Internet and electronic instruments, which can collect a large amount of

  DOI: 10.25236/scmc.2019.059

data. However, using the above tools to collect data will still be the most missing data, and still need to manually fill in relevant information. For example, when using the Internet and electronic instruments to collect data, when filling in personal information, you still need to print out a paper form to fill out, and then the responsible person to recycle, and collect and collect data, so that the human resources consumption in the whole data collection process is better. Big and very error prone. In this context, this paper designs a dynamic configurable data collector to further study the massive data collection in the context of big data, and how to use dynamic configuration of dynamic big data for collection, the main goal is to improve the efficiency of data collection.

2. System Architecture Design

2.1 System architecture design

In order to realize the complete import of data into the Excel template, and export the file in excel format to the collected data, the design system needs to include modules for reading, parsing, and recording metadata. The different templates in the system designed in this paper provide form interfaces for file uploads, allowing users to upload different types of files (Shi et al., 2010). At the same time, different modules in the designed system can create data models and field metadata models that satisfy different types of numbers based on the field names read and parsed by the data, as well as the field types. The mode configuration module in the system can realize user personalization. The user sets the field type and the corresponding form control type according to their own needs (Ma and Yang, 2013). At the same time, in the webpage publishing data collection module, the user can distribute different data collection forms through the webpage. During the release process, new data tools such as QQ, SMS, and WeChat can be sent to the data collector. In this process, the data collected by the collector mainly through the computer, collection and other tools, open the web page to fill in the relevant data information. Once the information is submitted, it is saved directly in the server database system (He, 2017). In addition, in the designed system preview and export module, the collected data is sorted and filtered. In the delete module, the specified data module is deleted, and the metadata and field metadata are deleted in the form of transactions. The main purpose is to maintain data consistency. The export module mainly uses a plug-in called Apache POI to classify the collected data and export it into an excel file.

2.2 System data model design

In order to realize the dynamic configurable mode of the model, on the basis of creating the corresponding data collection model of the Excel document, it is also necessary to specifically design the metadata data model related to the data record table, form, field, etc. These models mainly include the field metadata model and the table. Metadata model. Among them, in the data table metadata model, it is mainly used to store the basic information of each worksheet. The information is mainly stored by Excel template, including the original name, key field, current name, creation time, and publishing URL of the data table. , create a person, fill out the form description and other data. The field metadata model is mainly used to store information related to all fields in the design system, including field specific reference content, field type, field name, key field, foreign key, field comment, field reference mark, field corresponding form control, etc. type of data.

3. System Implementation

3.1 Specific implementation method

This design system mainly uses Apache POI plug-in, JDBC batch processing, and Jquery, Ajax asynchronous processing and other technologies to run. Among them, the mode parsing implementation mainly relies on the system entry form form, and provides an upload interface for the template data file. After the file upload is successful, the mode parsing uses the Apache POI plugin to read and parse the data in the file. In the data parsing process, the pattern parsing module identifies the data file type and field name, as well as the data type. Among them, different types of

data need to be parsed using different POIs. For example, xlsx type data needs to work with multiple data types such as XSSFCell and XSSFRow to complete data file parsing. In addition, when parsing the data content, it is necessary to distinguish the data types required by different template files, including date, number, null value, characters, etc. These parameters will affect the types of data created and collected by subsequent mode modules.

In the system designed in this paper, in order to facilitate data collection, a data model and field metadata are specifically set. These modules are mainly used to collect data and field types of data in metadata. However, the main drawback of this system is that the dynamic configuration of the data needs to depend on the metadata type in its module. Among them, the main storage types in the metadata model include the original name of the data table, the existing name, and information such as creation time and creator. The field metadata model mainly stores the field types, field comments, field reference flags, and low-end corresponding form control types of all data tables.

The mode configuration in the system mainly supports only two template files of structure and structure data. When uploading only the structure template file, there is only one line of sample data in the new model. In this model, all data types are manually entered, and the default form control is a “single line text box”. This data template will also reduce other types of data to be written into the model, reducing the amount of information input by the information, and improving the efficiency of data collection. In the data template of the structure data, the types of the data fields are all reference types. More importantly, in this template, data collectors can also add drop-down menus to select, further reducing data input and improving data collection efficiency. However, the above two data templates have a fixed uniformity and cannot meet the user's personalized data collection requirements. Therefore, on this basis, the data mode dynamic configuration function is also provided in the system.

In the data mode dynamic configuration function in the system, the data fields are mainly composed of the following four types. Self-referencing mainly refers to the information collection, the field value can be selected from the existing data of the template, or the data previously stored in the template, which reduces the information entry time to a certain extent, and can also improve the standardization operation of the data entry system. The limit value is mainly in the field value is a certain specific value. The specific value of this group can be the data uploaded from the template, or it can be set directly from the mode template, which has certain flexibility. Manual input mainly means that a field is completely entered manually by the data issuer. Referencing data in other tables mainly refers to the range of values of a field, which can be a field in an existing table in the system.

3.2 Key technologies for implementing the system

In order to run the design system smoothly, some key technologies need to be applied. The identification of the field data type is an important project, so the identification technology plays an important role in the realization of the system operation process. Specifically, in the actual running process of the system, only one row of fields is provided for the user in the Excel template, and no information of the data is included in the provided field. So to parse the field into a relational data model, you need to apply it to data recognition technology. The system designed in this paper mainly judges the data type through sample data identification. In the specific process, Apache POI technology mainly recognizes five data types such as CELL_TYPE_BLANK and CELL_TYPE_NUMERIC. However, in the actual operation of the system, there are far more than five types of data involved. At the same time, the limited and non-standard data will be involved in the system operation process, which will increase the difficulty of system operation. For example, the data such as the telephone number, QQ, and ID number of the collected information cannot accurately define the data. Therefore, the program needs to perform intelligent recognition and fuzzy judgment, and accurately define different types of data. In this process, when a program has a long string of numbers, the program needs to measure its length in the form of a string. If the length is 15 or 18 bits, the data can be identified as an ID number and the data is defined as a character.

Because of the fixed uniformity in the structural and structural data input methods only, the

system needs to provide dynamic system configuration. Therefore, data related to dynamic system configuration is applied during system operation. Importing with different types of templates also sets the default values. Deciding the default data type is primarily determined based on the sample data in the pattern parsing. When the template data import method is structure only, the default data types are form type and date box. After the data template is imported into the system, the user can dynamically configure the relevance of different fields according to their actual conditions to better meet the personalized experience of the user.

4. Conclusion

The system designed in this paper has already been in actual operation, realizing the basic functions of data acquisition and achieving great results. However, there are still many problems in the actual operation. For example, many user permission settings are not clear, mainly because the user does not have clear access to the page. There is no data filtering and viewing function on the data preview page, which reduces the user experience. Therefore, in order to better promote data collection, and to integrate the system with big data, we should continue to optimize and improve the system.

References

[1] Cheng Y., Zou Y.w., Li J.c., et al. (2010). Design of dynamic data acquisition pre-system based on wireless transmission. Laboratory Research and Exploration, 29(6), 37-39.

[2] Wang P. (2018). Dynamically configurable data visualization display system. Electronic Technology and Software Engineering, 7(6), 178-179.

[3] Gong S., Zhu Y.Q., Liang Y.H. (2018). Design and implementation of Python-based configurable automated crawler system. Computer fans, 109(10), 211.

[4] Cao L. (2011). Dynamic generation and development technology of source data collection form--Taking the source data acquisition system of single well geological database as an example. Computer application of petroleum industry, 18(4), 7-10.

[5] Deng Q.X., Ning B.F., Jin W. (2010). Multi-channel data acquisition system based on local dynamic reconfigurable technology. Small microcomputer system, 31(9), 1779-1783.

[6] Tian X.H., Zhang K., Liu Q. (2011). Programmable High Speed Data Acquisition Module with Multiple Trigger Functions. Single Chip Microcomputer and Embedded System Applications, 11(10), 43-45.

[7] Shi W.R., Yang X.K., Cai Z.L., et al. (2010). XML-based configurable data acquisition gateway. Microcomputer information, 26(24), 91-93.

[8] Ma K., Yang F. (2013). Design of high-speed multi-channel data acquisition system. China Test, 2013, 39(6), 85-88.

[9] He Q. (2017) Design and implementation of reconfigurable data acquisition system. Electronic Measurement Technology, 40(2), 109-112.