Advances in Computer, Signals and Systems (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2023.070108
ISSN 2371-8838 Vol. 7 Num. 1

Research on Automatic Generation of Engineering Geological Survey Report Paragraphs Based on Dynamic Data Template

Aoxue Han[1,a,*], Zhinan Lin[2,b]

[1]Hunan University of Science and Technology, Xiangtan, China
[2]Hunan Vocational Institute of Technology, Xiangtan, China
[a]binbin135197@163.com, [b]409435402@qq.com
**Corresponding author*

Keywords: Paragraph Generation, Dynamic Data, Node Tree, Template Editing

Abstract: As the final result of geological exploration, the engineering geological survey report needs to be prepared in large quantities, and the efficiency of manual preparation of the report is not high enough. To solve this problem, this paper proposes a standardized report paragraph generation model based on dynamic data templates. Through dynamic data template, the structure and content of report paragraphs are directly controlled by data; To solve the problem of template generating documents, a set of node tree rules is designed, and LaTeX files are generated by traversing the template node tree. The experiment shows that paragraphs can be modified by modifying data in terms of structure, content and text format, which improves work efficiency and has high use value in highly standardized engineering geological survey reports.

1. Introduction

In modern geological engineering, a large number of geological survey reports need to be prepared to meet the needs of different projects and different stages. The engineering geological survey report is the final result of the engineering geological survey and an important basis for the design and construction of the engineering foundation and structure foundation[1] .The geological survey report has strong professional relevance, and it needs to meet the specifications of the geological discipline. The geological survey report written and edited by relevant personnel may also have problems such as nonstandard content and incomplete structure, and also need to spend a lot of human resources. Aiming at the above problems, a model that can automatically generate standard document paragraphs is designed.

Document paragraph generation technology is a part of natural language generation, which is a technology from data to text. It can transform huge data into language forms that we can understand, greatly improving the ability of data analysis and understanding. Artificial intelligence is a popular direction of natural language generation, such as the natural language generation technology using LSTM (short and long term memory network) [2]. But if we rely entirely on computers to complete

all this, there will be great unpredictability. For the geological survey report, which is highly professional and normative, there will be non-standard and non-standard problems. The form of template can not only meet the need of saving manpower, but also impose a rule constraint on the document. Most of the domestic research on the generation of data to text is based on templates and generated by manually adding data[3]. However, these methods often have the problem that the structure and content transformation are not flexible enough to achieve the goal of changing with the dynamic data.

Many fields have made systems that meet the requirements of automatic document generation, such as medical examination report, weather detection report [4], etc. Among them, the automatic document generation method using templates has also been used in many fields, such as the design specification for automobile radiator, the oil fingerprint identification report[5], and the flood control document[6]. Research[7] adopts the method of finding and replacing bookmarks, and determines the place of data insertion by establishing a template in advance. This can make the template and data independent of each other and reduce the amount of code. The template is also relatively intuitive, but this method is inefficient. Research[8] through the template engine, the XML template converted from Word document is modified to generate a new template, and then the template is parsed by the template engine to fill in the data to generate the final document. This method requires a lot of time to make templates. Research[9] Formulated the Word generation strategy through OOXML, converted the document.xml file into the Joy template through the design template instruction, and modified the dynamic content in the template to generate the document. This method solves the complex problem of making Joy template, but it also increases the difficulty of development. Research[10] introduces the variable content document generation framework DPLfw based on the principle of software product line engineering, and completes the generation of variable content documents from a domain-oriented perspective. The DPL model uses a process similar to the product line to complete the generation of documents according to the variability of the functional record content. However, the framework must customize the editor to formulate the workflow of document creation, which is very heavy and difficult to develop.

In this paper, the template of node tree structure is used to connect dynamic data and document paragraphs. By setting different types of nodes, the content, structure and text format of the template are controlled, and an automatic generation model of geological survey report paragraphs based on dynamic data template is constructed.

2. Correlation Theory

2.1. Theoretical Knowledge

2.1.1. Data Transmission

The minimum storage structure of the tree is node, and each node contains dynamic data and node relationship information. Each type of node is stored in different tables and contains node type codes for query. Dynamic data represents the value of the structure or content represented by a node. A node can contain multiple data. The root node stores the sub-node information; Branch nodes store parent node, child node and sibling node information; Leaf node stores parent node and sibling node information.

The interaction between template and database is realized through SQL statements. The parameters contained in the node are strings composed of database table names, column names and query conditions, which can accurately locate the data in the database. When editing a template, the goal of storing the template tree is achieved through the parent-child relationship of the node, without generating additional complex template tables. After the Latex file is generated, the

document tree will delete itself. Only the template is stored in the database, not the document, which reduces the storage.

2.1.2. LaTeX

LaTeX code can meet the requirements of paragraphs in most address survey report documents. LaTeX code is often used in the editing of professional documents, such as professional journal papers[11]. The text in the document tree node contains the text format with LaTeX code. The format and text are written into the LaTeX file together, and the PDF report can be generated for visual viewing.

2.2. Technical Route

A scheme of automatic generation of paragraphs in engineering geological survey report based on dynamic data template is proposed. This scheme summarizes and summarizes the situations encountered in the paragraphs, and represents different sentence structures with different nodes, which together constitute the template of node tree structure. Get dynamic data from the database by traversing the template tree, generate a document of node tree structure, and finally convert it into a string type and pass it into the LaTeX file.

2.2.1. Template Editing

Different paragraph requirements are different. If no template can meet the paragraph requirements, you need to create a new template and edit it. You should add, delete and modify nodes in the template tree on the interface, and express all the structure and content in the paragraph in the template. We also need to provide query parameters for each node, so that the information required by the node can be quickly found in the database, and achieve the goal of dynamic generation. The edited template is saved in the database as a node.

2.2.2. Document Generation

Document generation is the process from template to document and then to LaTeX file. It is divided into two parts. The first part is the change from template tree to document tree, and the second part is the document generation LaTeX file.

The first part is the process of template to document from one tree to another, which is completed by traversing the template tree. First, the template tree is cloned, and the cloned new tree is traversed in order, and the document tree is generated by executing different node attributes and methods in different node types.

There are three types of nodes: root node, branch node and leaf node. The branch node has a set of leaf nodes. The root node is only responsible for carrying paragraph information, such as writing time and author, and does not participate in document generation. The three structures represented by branch nodes correspond to three functional modules, and the two leaf nodes represent one functional module. All sub-nodes are converted into sub-nodes of the root node of the document tree through different functional modules, and have format information.

(1) Leaf node module

Leaf nodes are divided into fixed value leaf nodes and variable value leaf nodes. Variable value leaf nodes need to obtain dynamic data from the database according to query parameters. After the variable value leaf node obtains the data, the node is converted into a fixed value leaf node, and the obtained data is used as the text of the node. The fixed value leaf node itself does not carry text format. It needs to be encoded according to the default text format of the template.

(2) Statement induction branch node module

The statement induction branch node represents a sentence structure, and its sub-node set is a complete sentence. The statement induction branch node can be regarded as a sentence template. The process from acquiring data to traversing all sub-nodes of the statement induction branch node can be seen as the process of generating sentence objects from the sentence template, which is the instantiation of the statement induction branch node. There are some continuous sentences with the same structure, but only the dynamic data are different. Such sentences can be generated continuously by the same sentence induction node. Statement induction branch node is also the basis for selecting branch node and special format branch node. Both structures operate on statement induction node. The specific process of the module is as follows

(3) Select branch node module

Select a branch node to make a selection of a sentence or phrase in a sentence. Select the sub-node of the branch node, including the statement induction branch node and two leaf nodes, corresponding to the sentence and phrase. In actual operation, it is often necessary to select a series of sentences or phrases. Select one or several of them, or not one of them, and classify them all as the selection structure. Select parameter is to select an attribute of a branch node, which is used as a keyword in database query. Each parameter has one more option. If there is no parameter, none is selected.

(4) Special format branch node module

Special format branch nodes are sentences or phrases with special format requirements for text format. Special text formats include font, size and color. These special formats are expressed in the form of coding. Use the format identifier to match the special text format with the format code, and then assign the format code to the fixed value leaf node. Text and text format coding form a fixed value leaf node containing format, which is the node of the document tree.

The second part of the document generation LaTeX file is the process of transforming the document from the node tree form to the intuitive and visible document form. By traversing the document tree, the text in the leaf node and the string formed by the text format code are written into the LaTeX file together with the LaTeX code combination. LaTeX files can be easily converted into pdf files for viewing as the final result.

3. Application Examples

3.1. Data Description

In the test, we used the engineering geological survey data in the construction drawing design stage of Pu'an No. 3 Bridge in the first contract section of the Sandu-Dushan section of the Yuqing-Anlong Expressway in Guizhou Province, including the physical geography, engineering geological conditions, adverse geology and other data. Acquire exploration data and store the data in the database for document generation through preliminary mapping, exploration, experiment and other tasks.

3.2. Experimental Results

This method can meet the requirements of paragraph preparation in most geological exploration reports. Templates have strong flexibility and versatility. Through data and templates, the structure, content and text format of paragraphs can be adjusted.

Two different projects A and B contain different topographic and geomorphological data, and the contents of the report paragraphs are very different. In Figure 1, the site is located on the southeast slope of Yunnan-Guizhou Plateau, and the geomorphic type belongs to the dissolution

type of middle mountain valley. In Figure 1, the site is located in the northern part of the northern Guizhou plateau, and the geomorphic type is erosion-dissolution type low-middle mountain valley. The two projects are also different in terms of maximum altitude, minimum altitude and relative maximum error.

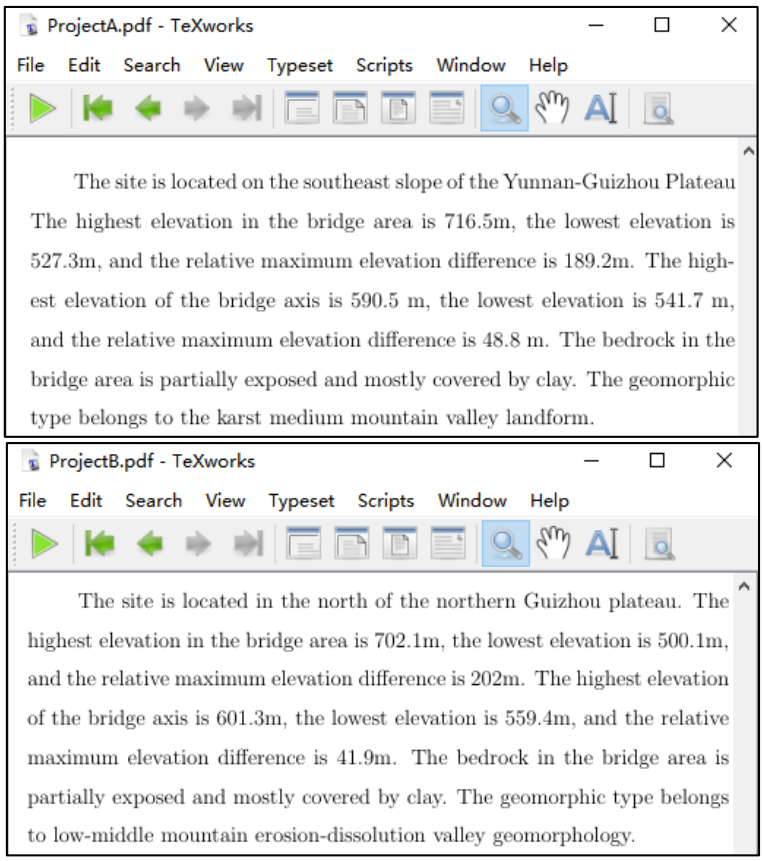
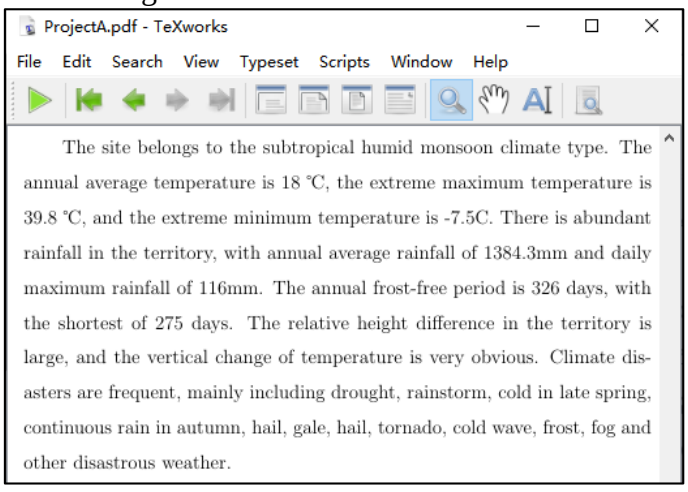


Figure 1: Comparison of two different project contents of the same template.

Two different items A and B contain different selection parameter data, and the structure of the report paragraphs is very different. In Figure 2, disastrous weather is a selection structure. Project A has 11 kinds of disastrous weather such as drought and rainstorm, while Project B has only 6 kinds of disastrous weather such as drought and rainstorm.



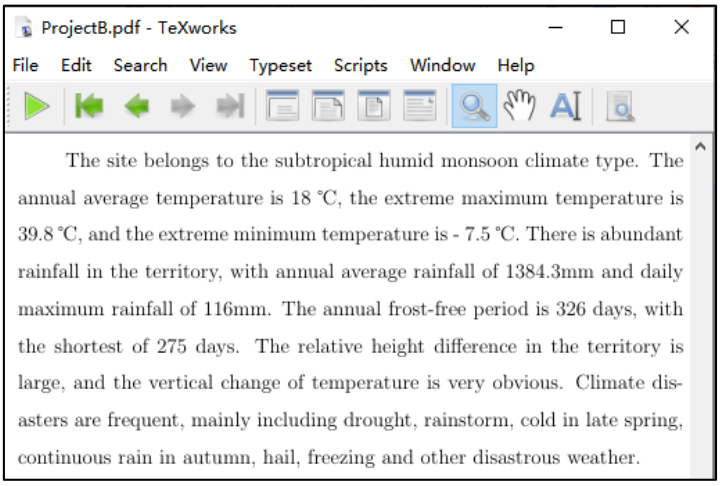


Figure 2: Comparison of two different project structures of the same template.

Two different items A and B have different format codes, and the text format of report paragraphs is very different. In Figure 3, the text in item B is bold, and the text in item A is not bold.

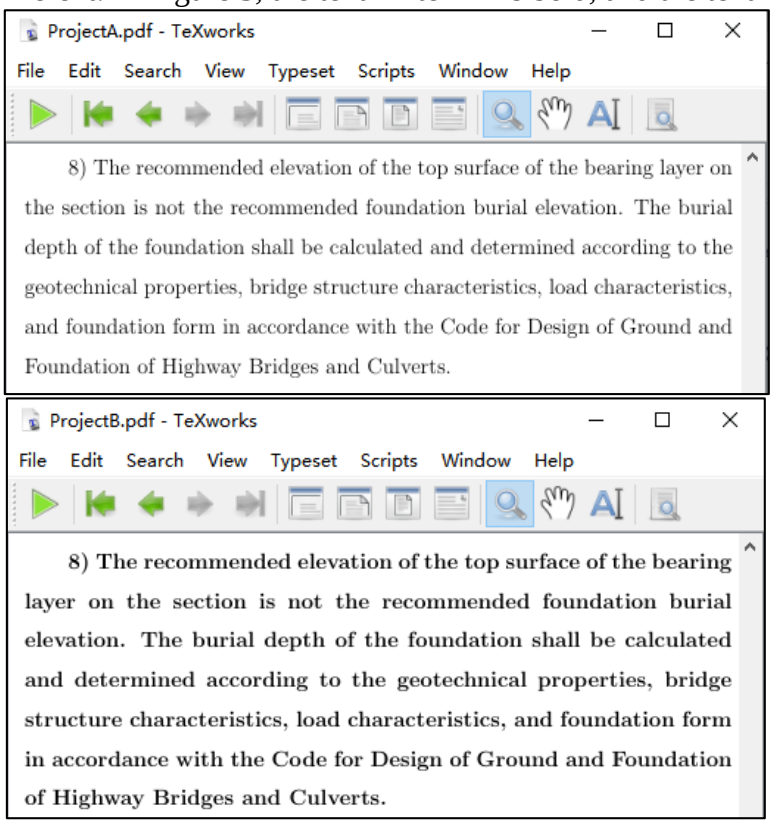


Figure 3: Comparison of text formats of two different items with the same template.

3.3. Effect Analysis

Compare the document generation method based on dynamic data template with other document generation methods based on template in the following aspects:

Paragraph content and structure: When there are multiple sentence phrases in a paragraph, dynamic data can realize different choices of sentences and phrases to meet different paragraph needs. Other template-based generation methods are very fixed in structure, and the structure of the

document is completely determined by the template. For example, the template engine in document [12] only modifies the dynamic content, not the structure.

Text format: the text format in the paragraph is completely separated from the content and structure of the paragraph. The text format is controlled by the dynamic data source, and the template only has the function of connection. In this way, the dynamic data can be directly modified to change the text format in the paragraph. Other template-based generation methods: text format and template are bound to each other. The text format must be changed by modifying the template. For example, the document [6] realizes the addition and modification of dynamic data by controlling the fields in the Word template, but the fields cannot control the text format.

Dynamic query: When querying nodes and dynamic data, you can quickly locate the data in the data source through the query parameters. In document [13], different databases and different query methods are used to realize data query, which requires a large number of query methods to face different templates.

In the automatic generation of paragraphs based on dynamic data templates, each part is independent of each other, and each function is connected through the template. As long as the dynamic data changes, you can modify the paragraph. Other template-based document generation methods can only change the document by modifying the template.

4. Conclusion

Aiming at the problem of automatic generation of survey report paragraphs in engineering geological survey, this paper proposes a document paragraph generation model based on dynamic data template after analyzing the professional specification requirements and specific data, and implements the model through the engineering geological survey report system. The results show that the model can realize the change of document paragraphs with the change of dynamic data. This method can meet the specification requirements of engineering geological survey report and improve the work efficiency.

On the basis of this method, there are still many directions to be studied. This method only solves the modification of the internal content and form of a paragraph, and should also study the relationship between paragraphs to realize the automatic generation of documents. In addition, the research did not consider the method of artificial intelligence, and should add the combination of the two to meet the standardization requirements of the engineering geological survey report and realize the intelligent generation of document paragraphs. There are also a large number of charts in the document, and the automatic generation of research charts is also the future research direction.

Acknowledgements

The author thanks Jian Lin and Zhinan Lin for their help in the research process.

References

[1] Du Juan. Preparation of engineering geological survey report [J]. Chinese and foreign entrepreneurs, 2015 (12): 219-220.

[2] Yu Zhenlong. Research and implementation of natural language generation technology based on LSTM [D]. Beijing University of Posts and Telecommunications, 2018.

[3] Li Xueqing, Wang Shi, Wang Zhujun, Zhu Junwu. Overview of natural language generation [J]. Computer Application, 2021, 41 (05): 1227-1235.

[4] Bai Xinyu Design and implementation of automatic generation system of digestive endoscope report based on intelligent template [D]. Shandong University, 2021. DOI: 10.27272/d.cnki.gshdu.2021.004548.

[5] Sun Jingjing Research on Word intelligent report based on label technology [D]. Ocean University of China, 2013.

[6] Dong Gaozhen. Design and implementation of logging report generation system based on. NET [J]. Logging

Engineering, 2014, 25 (01): 70-72+94.
[7] Wang Xiliang. Research and application of public meteorological service document automation [J]. Mid-low latitude mountain meteorology, 2018, 42 (03): 95-98.
[8] Fu Yan, Ji Min, Jia Ning. Design and implementation of oil fingerprint identification report generation system [J]. Ocean Information, 2018, 33 (02): 58-62. DOI: 10.19661/j.cnki.mi. 2018. 02.011.
[9] Jiang Peng, Xu Feng, Qi Rongzhi. Construction of an intelligent generation model for flood control documents based on cloud platform [J]. Water Resources Information, 2013 (03): 25-32. DOI: 10.19364/j.1674-9405.2013.03.008.
[10] Abel Gómez, M. Carmen Penadés, José H. Canós, Marcos R.S. Borges, Manuel Llavador. A framework for variable content document generation with multiple actors [J]. Information and Software Technology, 2014, 56(9).
[11] Jin Liangfeng, Zhou Wenxiang. Automatic generation of WORD test report [J]. China Test Technology, 2007 (04): 112-115.
[12] Luo Rong, Huang Jun, Li Maofeng, Liu Zhiqin. Rapid generation method of complex documents based on Word template [J]. Computer Application and Software, 2020, 37 (10): 57-63.
[13] Lei Yujiao, Deng Jiqiu, Lin Jian, Dick Jeffrey M., Lessani Mohammad Naser, Liu Chaoyue. Research of Automatic Generation for Engineering Geological Survey Reports Based on a Four-Dimensional Dynamic Template [J]. ISPRS International Journal of Geo-Information, 2020, 9(9).