Optimizing the Power Station Project Management Using Knowledge Graphs

Yang Cao[1], Linhua Su[1], Guoyi Jiang[1], Haibo Liu[1], Yuanmiao Gui[2,*] and Pengcheng Liu[3,4]

[1]*State Grid East Inner Mongolia Electric Power Supply Co., Ltd, Hohhot Inner Mongolia 010010, China*
[2]*Hefei Institutes of Physical Science, Chinese Academy of Sciences, Anhui 230031, China*
[3]*Institute of Physical Science and Information Technology, Anhui University, Anhui 230039, China*
[4]*Institute of Technology Innovation, Hefei Institute of Physical Science, Chinese Academy of Sciences, Anhui 230088, China*
E-mail: smalltalkman@foxmail.com
[*]*Corresponding Author*

Received 18 March 2021; Accepted 01 April 2021;
Publication 24 June 2021

Abstract

In "The Belt and Road", there are more and more investment projects of hydropower stations abroad in China. This paper proposes to use the semantic interconnection ability of knowledge map to manage the big data of hydropower using knowledge graphs. It solves the problem that a lot of project information is in the state of dispersion. To further understand the development space and potential of hydropower construction investment in different countries.

Keywords: Neo4j, hydropower station project, knowledge graph.

Distributed Generation & Alternative Energy Journal, Vol. 36_2, 113–124.
doi: 10.13052/dgaej2156-3306.3622
© 2021 River Publishers

1 Introduction

Power shortage has always been one of the important reasons hindering the rapid development of developing countries. The Belt and Road along the distribution is rich water resources for the construction of hydropower [1]. However, these hydropower projects are difficult to build, need to invest more money and other reasons. The degree of water resources development is generally not high. And our country has rich experience in hydropower project construction and abundant capital advantage. Therefore, the data of hydropower projects is increasing day by day, and how to efficiently manage hydropower projects becomes a problem. The knowledge graph is used to manage the big data of hydropower plant project. It solves the problem that a lot of project information is in the state of dispersion.

As a result of the development of semantic networks, knowledge graph has the advantages of relational networks. Knowledge graph can be used to organize knowledge in an orderly way. Knowledge graph can integrate various entities and construct entity relationship network by graph composed of nodes and edges [2]. According to Wikipedia, knowledge graph is the knowledge base that Google uses to support the organization of web data from a semantic perspective, thus providing intelligent search services [3]. Knowledge graph stores real-world entities and relationships between two entities in the form of structured triples [3].

Neo4j not only stores nodes and edges, but also supports complex manipulation of relationships. At present, there are more and more knowledge maps constructed by Neo4j, including film knowledge [5], protein knowledge [6], and Asian music knowledge graph [7]. And the hydropower projects knowledge graph can be built by Neo4j in this paper.

Based on the knowledge graph of hydropower projects, the deep relationships among overseas hydropower projects are analyzed and obtained. Through the integration of various information of hydropower project, the distribution of the project can be obtained, which is beneficial to provide the basis for the next development direction. Based on the knowledge graph of hydropower plant project, it can strengthen the management of hydropower plant project and improve the management efficiency. When different hydropower plants are used to generate electricity, the knowledge graph can be used to manage the hydropower plant in this field, and the optimal power generation strategy has been achieved.

2 Background

2.1 Hydropower Project

The Hydropower can convert the potential energy and kinetic energy of water into electrical energy. In "The Belt and Road", Power shortage has always been one of the important reasons hindering the rapid development of developing countries. Although there are abundant water resources along "The Belt and Road", the degree of development of these resources is generally not high due to the difficulty of hydropower project construction and the large amount of investment. And our country has rich experience in hydropower project construction and abundant capital advantage.

As an important industry of exporting renewable energy, the hydropower industry has been supported by the state. From 2016 to 2019, the overall growth rate of hydropower construction investment in China is on the rise. China invested 83.9 billion RMB in hydropower construction in 2019, up 19.86 percent from the same period in 2018. By The end of January to June 2020, China's hydropower construction investment reached 39.9 billion RMB, up 25.3 percent year on year [8].

China attaches great importance to hydropower development, and the Three Gorges Dam shows the importance it attaches to hydropower technology. In 2019, the investment scale of China's hydropower construction reached 83.9 billion RMB. Therefore, the construction capacity of hydropower stations in China has been recognized all over the world [8].

2.2 Knowledge Graph

From the perspective of human cognition of the world, knowledge graph describes information such as entities and relationships in a structured form. This easy to express the form of Internet information, effectively deal with more and more massive Internet data. Knowledge Graph not only differs from traditional methods in its representation, but also strengthens semantic relations. So the essence of the knowledge graph is a semantic network, a network that connects to huge amounts of data on the Internet.

The knowledge graph is divided into data layer and pattern layer. The data layer is the part that stores the knowledge, in which the knowledge is stored in the graph database as a unit of fact. When the triple knowledge is taken as the basic unit of fact, all the data stored in the graph database will form a huge entity relationship network and form the knowledge graph. The pattern layer is the core of the knowledge graph above the data layer. The schema

layer stores the refined knowledge and the ontology library is usually used to manage the schema layer of the knowledge graph.

2.3 Graph Database—Neo4j

Graph database is based on the idea and algorithm of graph theory in mathematics and realize the efficient processing of complex network database. Graphic database is good at efficiently processing a large number of complex, interconnected, changeable data, the computational efficiency is far higher than the traditional relational database.

Neo4j is an embedded, disk-based, fully transaction-enabled Java persistence engine that stores data in graphs (networks) rather than tables [9]. Neo4j focuses on resolving the performance degradation that occurs when querying a traditional RDBMS with a large number of connections. By modeling data around the graph, Neo4j traverses nodes and edges at the same rate, and its traversal rate has nothing to do with the amount of data composing the graph.

In Neo4j, the relationship is the most important element in the database, which represents the mutual relations between nodes [10]. In a relational database, a relationship is used for joining operations between multiple different tables, and the performance degradation of such operations is exponential with respect to the data of the relationship, but in Neo4j, it is used for pointing from one node to another, and its performance is linear.

Neo4j can use Cypher language for data operation [11]. Cypher is a declarative language. In contrast to imperative languages such as Java and scripting languages such as Gremlin and JRuby, the focus is on what to retrieve from the diagram, not how to do it. This makes it possible to focus on optimizing the query in the implementation details that are not disclosed to the user.

In the operation of completion of single data or deletion of redundant data, CYPHER language is also used for corresponding operation. For example, if there is an error in the connection of a data node, the relationship can be replaced directly. Or, after a new piece of knowledge is obtained, new nodes can be directly added to the knowledge graph in the way of adding relations.

Neo4j graph database is used to construct knowledge graph of power project data, which can not only save power project data effectively, but also model the complex relationship in power project data. It is helpful for efficient management of power project relationship and modeling of power project relationship.

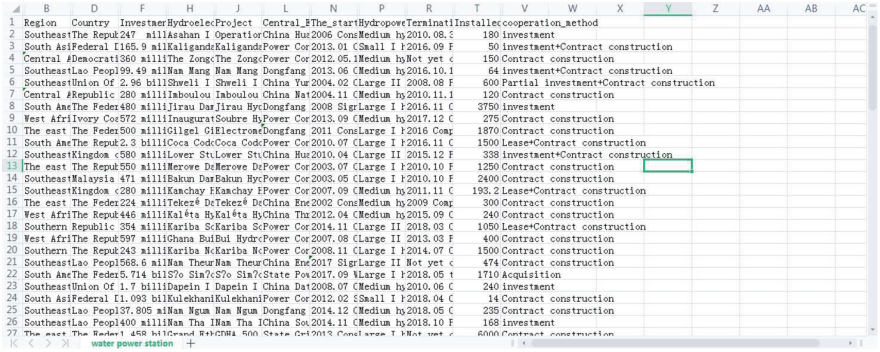
3 Construct Hydropower Project Knowledge Graphs

3.1 Data Set

The data set of hydropower station projects from 2002 to 2019 was collected by Yin et al. [12]. This study collected all available data using network crawler technology from the Belt and Road network, the Belt and Road energy cooperation network, the official websites of major enterprises, and the Ministry of Commerce database. This information was compiled into a data set of One Belt and Road hydropower stations from 2002 to 2019, and collected information on 111 hydropower projects, including very large, large, medium, small, and very small types of hydropower stations. The data set includes the names of the hydropower stations and hydropower projects, their locations, central enterprise groups and secondary units, the initial year of Chinese company construction, year of construction, type of hydropower station, installed capacity, cooperation mode, and the accuracy of positioning on 91 bitmaps.

Through simple processing of the data, we select the data set mainly including the region, country, investment amount, name of hydropower station project central enterprise group, starting time, hydropower station type, termination time, installed capacity and cooperation mode as in Figure 1.

As in Figure 1, each Chinese field has an English counterpart. Therefore, we can add the corresponding Chinese and English as two different attributes of the same entity to the knowledge graph.



	B	D	F	H	J	L	N	P	R	T	V	W	X	Y	Z	AA	AB	AC
1	Region	Country	InvestmerHydroelecProject	Central_E	The_start	HydropowTerminatI	Installec	cooperation_aethod										
2	SoutheastThe Repul	247	milliAsahan I Operatior	China Hu	2006.07	ConsMedium hy	2010.08.2		180	investment								
3	South AsiFederal I	165.9	milKaligandeKaligandePower Cor	2013.01	CSmall I H	2016.09 F			50	investment+Contract construction								
4	Central A	Democrati	360	milliThe Zong	The Zong	Power Cor	2012.05.1	Medium hy	Not yet c		150	Contract construction						
5	Southeast	Lao Peopl	99.49	milNam Mang	Nam Mang	Dongfang	2013.06.0	CMedium hy	2016.10.1		64	investment+Contract construction						
6	Southeast	Union Of	2.96	billShweli I	Shweli I	China Yur	2004.02.0	CLarge II	2008.08 F		600	Partial investment+Contract construction						
7	Central A	Republic	280	milliImboulou	Imboulou	China Nat	2004.11.0	CMedium hy	2010.11.1		120	Contract construction						
8	South Am	The Feder	480	milliJirau Dar	Jirau Hyc	Dongfang	2008.05	SIgrLarge I H	2016.11.0		3750	investment						
9	West Afri	Ivory Coa	572	milliInauguratSoubre Hy	Power Cor	2013.09.0	CMedium hy	2017.12.0		275	Contract construction							
10	The east	The Feder	500	milliGilgel GiElectromeDongfang	2011	ConsLarge I H	2016	Comp	1870	Contract construction								
11	South Am	The Repul	2.3	billiCoca CodeCoca CodePower Cor	2010.07.0	CLarge I H	2016.11.0		1500	Lease+Contract construction								
12	Southeast	Kingdom o	580	milliLower StuLower StuChina Hu	2010.04.0	CLarge II	2015.12.F		338	investment+Contract construction								
13	The east	The Repul	550	milliMerowe DeMerowe DePower Cor	2003.07.0	CLarge I H	2010.10.F		1250	Contract construction								
14	Southeast	Malaysia	471	milliBakun DarBakun HycPower Cor	2003.05.0	CLarge I H	2010.10.F		2400	Contract construction								
15	Southeast	Kingdom o	280	milliKamchay HKamchay HPower Cor	2007.09.0	CMedium hy	2011.11.0		193.2	Lease+Contract construction								
16	The east	The Feder	224	milliTekezé DeTekezé DrChina Ene	2002.0	ConsMedium hy	2009	Comp	300	Contract construction								
17	West Afri	The Repul	446	milliKal é ta HyKal é ta HyChina Thi	2012.04.0	CMedium hy	2015.09.0		240	Contract construction								
18	Southern	Republic	354	milliKariba ScKariba ScPower Cor	2014.11.0	CLarge II	2018.03.0		1050	Lease+Contract construction								
19	West Afri	The Repul	597	milliGhana BuiBui HydrcPower Cor	2007.08.0	CLarge II	2013.03.F		400	Contract construction								
20	Southern	The Repul	243	milliKariba NcKariba NcPower Cor	2008.11.0	CLarge I H	2014.07.0		1500	Contract construction								
21	Southeast	Lao Peopl	1568.6	milNam TheurNam TheurChina Ene	2017.0	SIgrLarge II	Not yet c		474	Contract construction								
22	South Am	The Feder	5.714	billS?o Sla?cS?o Sla?c	State Pov	2017.09.0	VLarge I H	2018.05.t	1710	Acquisition								
23	Southeast	Union Of	1.7	billiDapein I	Dapein I	China Dai	2008.07.0	CMedium hy	2010.06.0		240	investment						
24	South Asi	Federal I	1.093	bilKulekhaniKulekhaniPower Cor	2012.02.2	SSmall I H	2018.04.0		14	Contract construction								
25	Southeast	Lao Peopl	37.805	miNam Ngum	Nam Ngum	Dongfang	2014.12.0	CMedium hy	2018.05.0		235	Contract construction						
26	Southeast	Lao Peopl	400	milliNam Tha I	Nam Tha I	China Soc	2014.11.0	CMedium hy	2018.10.F		168	investment						
27	The east	The Feder	1.458	bilGrand RaHGDHA	500	State Gri	2013	ConsLarge I H	Not yet c		6000	Contract construction						

Figure 1 The original data.

Figure 2 Attributes of node.

3.2 Construct a Knowledge Graph with Neo4j

After getting the CSV table data, the hydropower station project knowledge graph can be constructed by Neo4j. The CSV table data is importing into the Neo4j graph database by cypher language.

Firstly, hydropower station project data was put into Neo4j diagram database, and CSV table data was imported as knowledge graph node according to columns. And you can import the Chinese name and the English name as different attributes on the same node.

Therefore, different attributes can be selected as the content of the entity during visualization as in Figure 2.

Secondly, according to the visualization requirements, the hydropower station project data is imported into Neo4j.

Finally, the relationship between the two columns of data is added to obtain the knowledge.

Figure 3 shows the information possessed by a hydropower station in this data set. There are 11 nodes and 10 relationships in a hydropower projects knowledge graph.

3.3 Display of Public Safety Knowledge Graph

The knowledge graph of hydropower station project can be obtained by the above steps, and then the knowledge graph can be operated by Cypher language.

First, hydropower stations in the same country are shown in Figure 4.

Figure 3 A hydropower station.

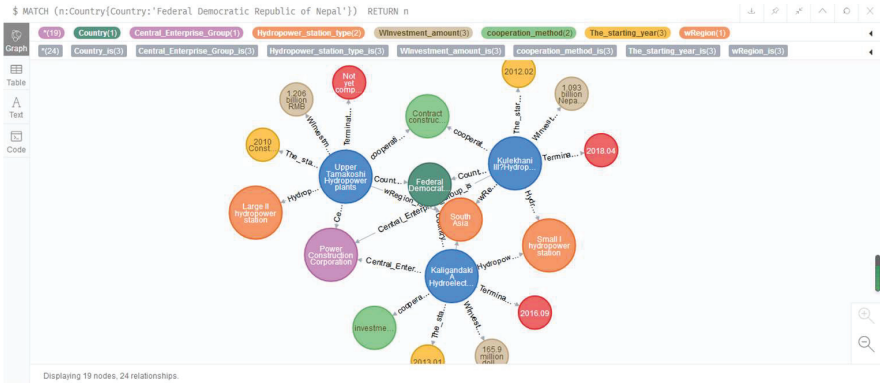


Figure 4 The hydropower project in Nepal.

This picture shows the hydropower project in Nepal. As can be seen from the picture, there are three hydropower stations built in Nepal, all of which are under the charge of China Power Construction Corporation. And the cooperation mode includes construction, indicating that the construction of hydropower stations in Nepal is backward and cannot meet the requirements. Moreover, the construction of the hydropower station started from 2010 to 2013, indicating that Nepal has already met its domestic electricity demand after the construction of these hydropower stations.

Next is to a company to undertake the hydropower station project inquiry.

In Figure 5, China Huadian Corporation was chosen as the focus of the inquiry. As can be seen from the figure, the company is not investing in small hydropower stations, and the investment amount is very large. In terms

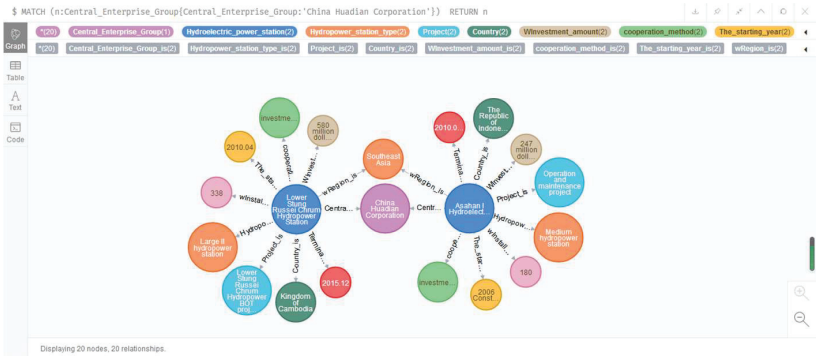


Figure 5 The hydropower project in China Huadian Corporation.

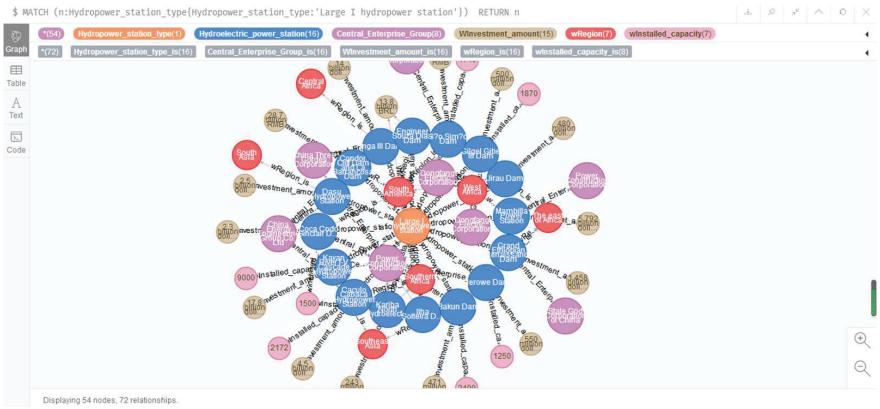


Figure 6 A certain type of hydropower station.

of time, the construction of Asahan I Hydroelectric Power Statio began in 2006 and ended in 2010. Construction of the Lower Stung Russei Chrum Hydropower Station began in 2010 and ended in 2015. This shows that China Huadian Corporation undertakes projects in serial rather than in parallel, so it can be inferred that the company can undertake only one large and medium-sized hydropower station project at a time.

Figure 6 is a query for a certain type of hydropower station.

After the query, there are 16 Large I hydropower stations in the data set. Six of them are located in South America, indicating that there are more basic conditions for large-scale hydropower generation in South America. This further illustrates the lack of large-scale water conservancy construction in South America, which has perfect conditions but does not make use of it.

The most expensive hydropower station is the 1-5 cascade of Kayan River in Indonesia, which cost us $17.8 billion, and the least expensive one is the Bakun Hydropower station in Malaysia, which cost RMB 471 million.

4 Conclusion

In this paper, Neo4j is used to construct the hydropower station project knowledge graph. The knowledge graph of hydropower projects will give a clearer picture of the links between projects. It solves the problem that a lot of project information is in the state of dispersion. In addition, the data of hydropower stations in a certain country can be obtained, and further information about the development potential of hydropower stations hidden in the data can be obtained.

Here are the problems to be solved:

Firstly, the amount conversion unit in the data needs to be modified, and this part will be negotiated with the original author of the data set.

Secondly, Neo4j requires that the resulting data set be stored as a CSV table. This will make the data preprocessing phase more difficult and consume a lot of time.

Acknowledgements

This work is supported by the Development Program of China (2018YFB1004600), and the National Nature Science Foundation of China (41871302).

References

[1] https://www.sohu.com/a/279360195_99911830.
[2] Q. Cao, Y. M. Zhao, “Technology implementation process and the related application of knowledge graph”, Information Studies: Theory & Application, 2015, 38(12): 13–18.
[3] Zhao Jun, LIU Kang, He Shizhu, Chen Yubo, Knowledge Atlas [M]. Edited. Beijing: Higher Education Press, 2018:2–6, 20–25.
[4] A.Vukotic, N. Watt, T. Abedrabbo, et al, Neo4j in Action, 2014.
[5] H. Lu, Z. Hong, M. Shi, “Analysis of film data based on Neo4j”, in 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 2017. 675–677.

[6] D. Hoksza, J. Jelinek, "Using Neo4j for Mining Protein Graphs: A Case Study", in 2015 26th International Workshop on Database and Expert Systems Applications (DEXA), Valencia, Spain, 2015. 230–234.
[7] K. Wu, M. Rege, Hibiki: "A Graph Visualization of Asian Music", in 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), Los Angeles, CA, USA, 2019 291–294.
[8] https://bg.qianzhan.com/trends/detail/506/200820-7b28d5e7.html.
[9] J.J. Miller, "Graph database applications and concepts with Neo4j", Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA. 2013, 2324(S 36).
[10] Pang Guoming, Hu Jiahui, Su Liang, Ed. Beijing: Tsinghua University Press, 2017.
[11] Z. Zhang, G.M. Pang, J.H.Hu, et al, "Neo4j authoritative guide", 2017.
[12] Yin fujie, wu Ming quan, xiao jianhua, niu zheng. "2002–2019 "One Belt And One Road" hydropower project information data set." V1. Science Data Bank. http://www.dx.doi.org/10.11922/sciencedb.855. (2019-08-22).

Biographies

Yang Cao received his B.Sc. degree in Electrical Engineering and Automation from Northeast Electric Power University, China. He is mainly engaged in the field of power system research.

Linhua Su received his B.Sc. degree in Electronic Radio Technology from Dalian University of Technology, China. He is mainly engaged in the field of power system research.

Guoyi Jiang received his M.Sc. degree in Thermal Power and Automation of Power Plant from Northeast Electric Power University, China. He is mainly engaged in the field of power system research.

Haibo Liu received his M.Sc. degree in Electrical Engineering and Automation from North China Electric Power University, China. He is mainly engaged in the field of power system research.

Yuanmiao Gui received his B.Sc. degree in Mathematics and Applied Mathematics from Shandong University of Science and Technology, China; M.Sc. and Ph.D. degrees in Control science and Engineering from the University of Science and Technology of China. He is currently an assistant researcher at the Hefei Institute of Intelligent Machinery, Hefei Institutes of Physical Science, Chinese Academy of Sciences, mainly engaged in the theoretical research of deep learning, knowledge graphs, big data, and applications in electric power, water conservancy, biology and other industries.

Pengcheng Liu holds a B.Sc. degree in automation from Anhui University. He is currently a graduate student in material science and information technology at Anhui University. His research interests include knowledge map construction and the application of big data in disaster and power.