Knowledge Graph Approach for Complex Relationship: A Study on Kinship in civil servant*

Dinh-Van Phan[1,2] [0000-0002-7015-1432]

[1] University of Economics, The University of Danang, Vietnam
[2] Teaching and Research Team for Business Intelligence, University of Economics, The University of Danang, Vietnam
dvan2707@due.edu.vn

Abstract. Knowledge Graph has been widely applied in many studies to indicate the complex relationships such as social networks, genomes etc. Thereby, It can also detect the relationship between proteins and diseases such as cancer, asthma, etc. Because of the advantages in visual representation and easy access of Knowledge Graph that the relational data system can hardly do. Therefore, this study applies Knowledge Graph to represent and build solutions to quickly retrieve personal relationships in civil servants based on knowledge graph data management system Neo4J and R programming language, combined with the Shiny package. The study showed five relationships of employees including grandparents, parents, spouse, siblings, children. Indeed, the study indicates the applying ability to manage and monitor the work process of employees in organizations.

Keywords: Knowledge graph, Relationship, Civil servant.

1 Introductions

The selection and transparency in the work of administrative agencies is an issue of concern for most countries in the world. The country still has a problem with corruption. Corruption is not only about money but also about the lack of transparency in the selection and appointment of cadres. In Vietnam, the fact that there have been many cases of improper appointment of civil servants and appointment of relatives as officials has caused frustration in the social and the Government [1]. As we know, there were many officers with many families who were recruited and worked in the agencies that the officer managed [2, 3]. These cases not only affected the quality of cadres but also created a wave of anger in the community, eroding people's trust in the government.

In reality, the management of human resources in the government is still difficult. Most of them are using relational databases to manage personnel records. Therefore, the data is often difficult to represent, difficult to access the complex relationships and relationships across many objects. Therefore, it is often difficult to manage, evaluate in general and thoroughly the kinship among personnel in agencies in order to limit the negative cases.

In recent years, Knowledge Graph (KG) is widely used in many fields world wide for representing relationships such as semantics, social networks, and search engines. For example, Microsoft's KG Bing and Google's KG both support searching and answering search queries. It describes and connects people, places, things, organizations, all knowledge in the world. Facebook's KG is the world's largest social network, including contents like music, movies, celebrities, and places of interest etc [4]. eBay's KG represents a product's relationship to real-world entities, identifying them, and determining their value to buyers. Besides, Graph Database is also applied to research in the fields of education and medicine. Typical in the medical industry is “GetNNet”, an integrated genomic analysis that unifies scientific workflows with the Graph Database [5].

Therefore, this study used KG to visually represent the complex kinship among civil servants of organizations. In addition, the study also indicated the ability to strongly trace complex kinship of civil servants in agencies. Because the personal data of officials is confidential, this study used demo data based on regulations on personnel records of the government in Vietnam.

2 Materials and methods

In knowledge representation and inference, KG uses a topology or graph structured data model to integrate data. KGs are often used to store interconnected descriptions of entities, objects, events, situations, or abstract concepts [6]. KG is a model for representing knowledge content in mathematical form. The idea of the model is to reorganize the content knowledge into nodes, and edges. The edges are used to represent connecting between nodes in a graph.

Relational data is organized by rows, columns, tables, but KGs are organized by nodes and edges to represent entities and relationships between entities. Nodes represent entities or instances such as people, businesses, accounts, or any other item that is tracked. They are similar to a record, relation, or row in a relational database. Edges carry the relationship to propagate from this node to other nodes. Edges can be directed or undirected. In an undirected graph, an edge connecting two nodes has a unique meaning, but in a directed graph, the edges connecting two different nodes have different meanings. The nodes and edges may have properties to add more information related to them.

KG will be an important method to the development of semantic understanding [7], turning data into knowledge, creating powerful, user-friendly products and experiences. Most data and databases of all types can be performed by KG. The Knowledge Graph

is fundamentally simpler than the relational model, but it is more expressive, easier to modify and expand into big data.

Graph Database (GD) is a type of database structured in the form of graphs mainly to store data of entities, small or large sets of social entities easily, conveniently and efficiently. It represents nodes and edges based on relations between them together. Along with that is the ability to store big data sets with quick query speeds.

In this study, the data was built from the data of human resource management in agencies of government. The input data was collected from the Resume of civil servants. For this resume sample, the competent authority requires that employees need to declare the background information and the declarant's three generations of relatives. Therefore, this form is considered an important procedure to consider the participation of workers in the offices. According to current regulations, the resumes of civil servants are made according to the form circular number 07/2019/TT-BNV stipulating the regime of statistical reporting and management of employee records. The content of the form includes 32 items [8]. However, this study focuses on the representation and retrieval of personal relationships and personal information, so we only use some information including full name, ID, position title (including start time, end time), and information about kinship relatives with 3 generations including parents, spouse, children, siblings (including in-law).

In this study, we not only consider employees in one agency, but also consider employees in other agencies that are affiliated and have relationships with each other, thereby showing a more complete personal relationship in the way of organizing agencies in Vietnam. In this case, we use agencies under Quang Nam province, Vietnam (pictured). The agencies in Vietnam contant the provincial level, the district level and the commune/ward level, departments, division, etc.

The study used tools such as Neo4j, R. Neo4j is a graph database management system developed by Neo4j, Inc [9]. Described by the developers as an ACID – Compliant transactional database with native graph storage and processing. Neo4j ranks as the most popular GD according to the DB-Engines rankings and ranks 22nd for databases overall [10].

R is a comprehensive programming language, which means it provides services for statistical modeling as well as for software development. R is the primary language for Data Science as well as for developing web applications through its powerful RShiny package [11].

3 Results and Discussion

Demo data is organized as a CSV file and imported into Neo4J. After importing demo data into neo4J, the system has built 1741 nodes, including 742 agency nodes and 999 people (employees) nodes. To access the personal relationships in Neo4J, we can use the Cypher language in Neo4J. As 02 special cases:

(1) Retrieve people who have a kinship with employee 01 (Emp. 01): *Match (c) – [] →(n:* "People{Name: *"Emp. 01"}.) – [] →(a: "People") return n, c, a*. This retrieval shows that there are fours employees have skinship with Emp 01 including Emp. 02

(Spouse), Emp. 03 (Grandparent), Emp 04 (Uncle), and Emp. 05 (Parients) (see Fig. 01)

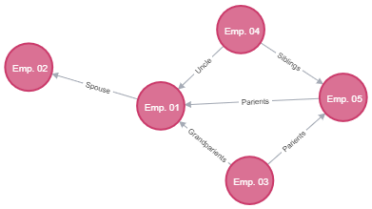

Fig. 1. Kinship of an employee

(2) Retrieve any relationship with 1 node on the left and 3 node levels on the right of Emp. 01: *Match n = () ← [] – (m: People{Name: "Emp. 01"}) – [] → (: 'People") – [] → () – [] → () return n.* This retrieval showed agencies nodes that have relationship with Emp. 01, and also showed and employees (Emp. 02, Emp. 06) with those positions who have relationships with employee 01 (see Fig. 02).

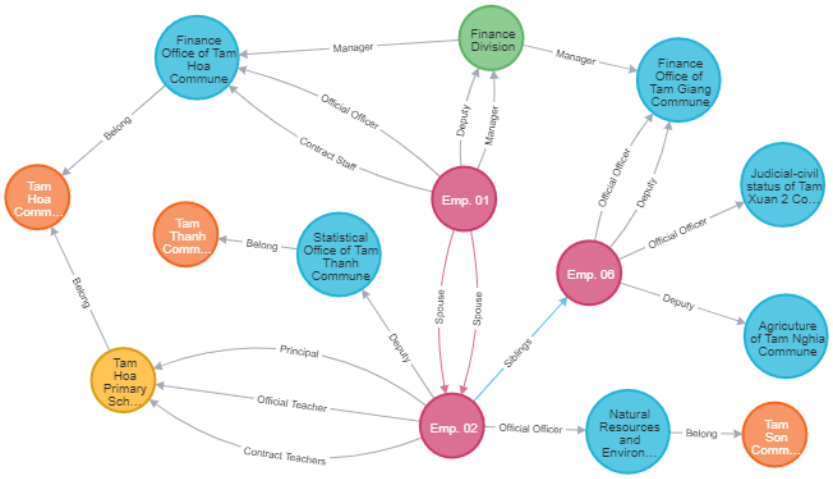

Fig. 2. Employees and agencyies relationships with employee 01

In addition, this study has also built a web-based retrieval system through R language combined with the Shiny package. Through this system, it is possible to support faster and easier access by user interface and selection operation (see Fig. 03). The system may fastly show the nodes and relationships by selecting an employee, or kinship, or agency. When we search by a condition, the system only shows the related nodes and relationships, the others were disable status.

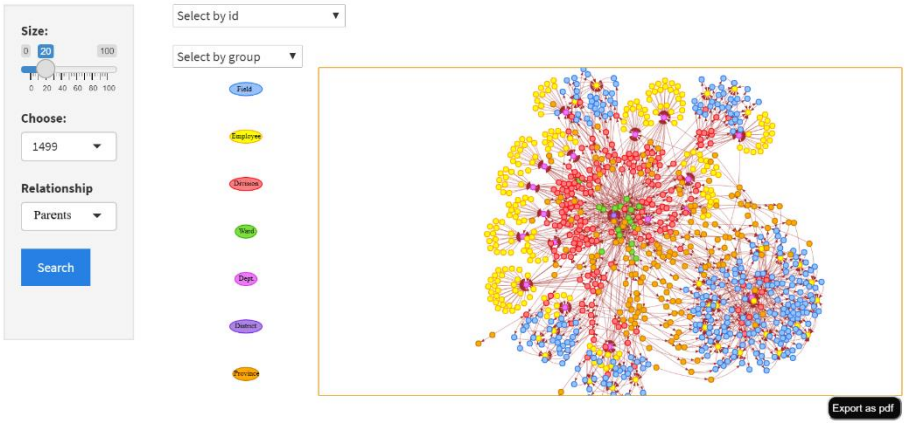

Fig. 3. Web interface of Civil servant graph

In fact, KG has been applied in many studies to find complex relationships [12] and relationships that relational data methods cannot do. As studies find the link of genes, protein with asthma disease [13], cancer [14], Integrated and extensible biochemical [15]. However, there has been no study on applying KG to represent the kinship in employees. Thereby supporting employee management in government agencies in particular and organizations in general.

4 Conclution

The study has presented demo data of civil servants on the Neo4j graph database. That can easily access kinship of civil servants through nodes, relationships with levels. The study also built a web interface to help users access information about personal relationships fastly. However, because the data of civil servants is confidential data, the study only uses demo data and has not used actual data. If the study is supported by the authority agencies, it will indicate an application direction for staff management. Thereby, a complete system can be built to put into practical use in Vietnamese to control personnel thoroughly.

In addition, through this study, we also realize the possibility and feasibility for applying Knowledge Graph to other fields such as applied to represent COVID-19 infections. Specifically, it is possible to represent F0 cases according to localities, F1 cases are related to F0 cases through relationships (edge). Thereby, it is possible to visually represent and easily track the relationship of COVID-19 infected cases. KG also may perform the training programs in Universities. Specifically, KG can visually represent majors, courses and the relationships between them. In addition, KG can represent the output standards of the Majors and courses. Thereby, we can monitor and manage the training programs visually. Thereby, it can support students for tracking academic subjects in the curriculum.

Acknowledgments

The author acknowledges the study team who are Binh-Yen Le, Thi-Man Nguyen, Dinh-Hieu Tran, Thi-Vu-Sa Doan, Thi-Thao-Nhi Nguyen. They have built the demo data and imported it to Neo4J and R.

References

1. http://baochinhphu.vn/Tin-noi-bat/Vi-sao-Thu-tuong-yeu-cau-tim-nguoi-tai-khong-tim-nguoi-nha/283469.vgp
2. https://thanhnien.vn/thoi-su/30-tuoi-lam-giam-doc-so-ke-hoach-va-dau-tu-tre-nhat-nuoc-611811.html
3. https://vov.vn/nhan-su/uu-ai-bo-nhiem-con-trai-den-luot-chu-tich-quang-nam-nhan-an-ky-luat-737595.vov
4. Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., Taylor, J.: Industry-scale knowledge graphs: lessons and challenges. Communications of the ACM 62, 36-43 (2019)
5. Costa, R.L., Gadelha, L., Ribeiro-Alves, M., Porto, F.: Gennet: An integrated platform for unifying scientific workflow management and graph databases for transcriptome data analysis. bioRxiv 095257 (2016)
6. https://en.wikipedia.org/wiki/Knowledge_graph
7. https://towardsdatascience.com/knowledge-graph-bb78055a7884
8. https://www.moha.gov.vn/danh-muc/thong-tu-so-07-2019-tt-bnv-ngay-01-6-2019-cua-bo-noi-vu-quy-dinh-ve-che-do-bao-cao-thong-ke-va-quan-ly-ho-so-vien-chuc-40757.html
9. https://neo4j.com/
10. https://www.wanttolearn.xyz/learn-neo4j/
11. https://en.wikipedia.org/wiki/R_(programming_language)
12. Toure, V., Mazein, A., Waltemath, D., Balaur, I., Saqi, M., Henkel, R., Pellet, J., Auffray, C.: STON: exploring biological pathways using the SBGN standard and graph databases. BMC Bioinformatics 17, 494 (2016)
13. Lysenko, A., Roznovat, I.A., Saqi, M., Mazein, A., Rawlings, C.J., Auffray, C.: Representing and querying disease networks using graph databases. BioData Min 9, 23 (2016)
14. Johnson, D., Connor, A.J., McKeever, S., Wang, Z., Deisboeck, T.S., Quaiser, T., Shochat, E.: Semantically linking in silico cancer models. Cancer Inform 13, 133-143 (2014)
15. Swainston, N., Batista-Navarro, R., Carbonell, P., Dobson, P.D., Dunstan, M., Jervis, A.J., Vinaixa, M., Williams, A.R., Ananiadou, S., Faulon, J.L., Mendes, P., Kell, D.B., Scrutton, N.S., Breitling, R.: biochem4j: Integrated and extensible biochemical knowledge through graph databases. PLoS One 12, e0179130 (2017)