An Approach to Constructing a Knowledge Graph Based on Korean Open-Government Data

Jae-won Lee [1] and Jaehui Park [2,*]

[1] Department of Open Data, National Information Society Agency, Daegu 41068, Korea; jaewonlee@nia.or.kr
[2] Department of Computer Science and Engineering, Incheon National University, Incheon 22012, Korea
*   Correspondence: jaehui@inu.ac.kr; Tel.: +82-32-835-8494

check for updates

Abstract: A data platform collecting the whole metadata held by government agencies and a knowledge graph showing the relationship between the collected open-government data are proposed in this paper. By practically applying the data platform and the knowledge graph to the public sector in Korea, three improvements were expected: (1) enhancing user accessibility across open-government data; (2) allowing users to acquire relevant data as well as desired data with a single query; and (3) enabling data-driven decision-making. In particular, the barriers for citizens to acquire the necessary data have been greatly reduced by using the proposed knowledge graph, which is considered to be important for data-driven decision-making. The reliability and feasibility of constructing a metadata-based open-data platform and a knowledge graph are estimated to be considerably high as the proposed approach is applied to a real service of the public sector in Korea.

Keywords: open-government data; metadata; knowledge graph; linked data

1. Introduction

Due to the advances of information and communication technologies (ICTs) such as the Internet and mobile phones, large volumes of data, commonly referred to as "big data", are now being produced in public and private sectors around the world. Recently, some services, such as weather forecasting and vehicle navigation, have intended to improve the reliability (or quality) of services by exploiting open-government data (OGD), and several governments are actively disclosing their OGDs. As a result, governments have become one of the largest data producers in various areas, and there is no limit to the sharing and redistribution of OGDs. According to an Organization for Economic Co operation and Development (OECD) report [1], data-driven innovation has become a key part of 21st-century growth that can significantly improve productivity and social well-being. OGDs are expected to offer a variety of benefits to the community, such as increasing government transparency and accountability, stimulating innovation, and contributing to economic growth [2,3].

According to Kalampokis et al. [4], since the availability of OGDs in the private sector has increased, numerous OGD portals have been launched to provide a single access point to the government data in various domains. As the number of OGDs and the diverse needs of users have increased sharply, citizens have begun to complain that it is difficult to get proper OGDs in the desired form whenever they need it. This is because only OGDs that governments intend to disclose to citizens are registered with their OGD portals. In other words, ordinary citizens cannot know which government agencies produce and hold openable data (not yet provided to their citizens) and to whom they can request the desired OGDs.

Moreover, most citizens want to get a set of relevant OGDs with a single query across heterogeneous domains. For example, suppose a user looking for a set of OGDs submits a query "car camping" to the OGD portal to build the web service related to car camping. Currently, most of the

OGD portals are based on keyword-based search engines, so they can return only search results (i.e., OGDs) that contain "car", "camping" or "car camping" as index terms. In fact, a set of OGDs containing business hours, the weather of camp sites, the pet admission and the location of pet hospital are more useful for building his/her own web service.

The most significant factors that enhance user accessibility to OGDs are for citizens to (1) exactly identify the existing state of OGDs held by government agencies and (2) make it easier to find the relevant OGDs with a single query. To resolve these issues, this paper proposes an open-data platform for auto-collecting the metadata of entire OGDs (including openable data in the future) held by government agencies. In addition, the collected metadata are used to construct a knowledge graph for accessing the entire collected OGDs. To improve the citizens' accessibility to governmental data, we have built the metadata-based open-data platform for real service (www.data.go.kr) with a large-scale OGD of the public sectors. An OGD is linked to semantically relevant ones in the graph, so citizens can easily find the relevant data set as well as the desired data. Our proposal has been applied to the Korean OGD portal (www.data.go.kr), and the OGD-based knowledge graph has also called "OGD map". The word "map" comes from the fact that citizens can see the locations and open-up status of the entire OGDs. To avoid the confusion with geographical maps, this paper refers to the structure of the knowledge graph as "OGD graph" instead of "OGD map", hereafter. To the best of our knowledge, this is the first work that constructs a knowledge graph with large-scale OGDs in a real service of the public sector.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3.1 describes an overview of metadata-based open-data platform. Section 3.2 explains how to automatically collect and standardize metadata from government agencies. Section 3.3 describes the detailed modelling for an OGD graph in the proposed open-data platform. Section 4 shows the proposed OGD graph on the Korean OGD portal. Section 5 concludes this paper and discuss future work.

2. Related Work

The OGD platform has been considered throughout the nationwide with high expectations for benefits to society. Recent studies in this research field has covered many facets of topics including sociology, humanities, cognitive sciences, education, as well as information science. Recent work [5] have addressed the topic of OGD by identifying the problems in developing an ecosystem model of stakeholders, policies, practices, relationships, and influences. The ecosystem has been applied to two case studies of New York a,d St. Petersburg. The study [6] have investigated the evaluation methods to assess the quality of OGD. It identified common problems in OGD as lack of metadata, incompleteness, and loss of updates. Qualitative analysis has focused mainly on the policy guidelines for data quality control. Several other works [6–8] have introduced the portals as their flagships initiatives of OGD. However, they usually consider the strategic perspectives distributing OGD to society rather than its computational meanings resided in the large-scale data sets. This paper is different from the above in that we have developed a practical data search system, whose technical structures are adopted from information science, to support further data analysis for the strategic social applications.

Until now, a popular information retrieval (IR) paradigm is the exact matching of index terms (or keywords) derived from documents and users' needs (such as queries) because of its simplicity and efficiency [9–11]. The most successful keyword-based IR systems used in users' daily lives are keyword search systems such as Google, Bing, and others. Typically, the keyword search system estimates the degree of relationship between the query and the target documents (i.e., search results), and the search results are ordered and displayed in a list of snippets consisting of titles, URLs and descriptions of the web pages. Even if a user wants to know the relationship between search results, he/she will not be able to apprehend the relationship explicitly in existing keyword search systems. Because of the growing number of users interested in getting a set of search results with a single query, some studies

try to take advantage of explicit relationships in documents. This paper classified these studies into two categories; logic-based search, link-based search.

Logic-based search systems [12–14] use domain knowledge to define concepts and their relationships in the given domain. Therefore, a user, who wants to use the logic-based search system, must have background knowledge (such as a query language) to exploit the domain knowledge (or knowledge graph) well. For example, as knowledge graphs such as YAGO [15], and DBpedia [16] typically represent the billions of facts in the form of RDF triples, ontology-based query languages such as SPARQL and OWL-QL are widely used in logic-based search systems. Furthermore, since logic-based search systems provide inferred results via inference engines, users are generally satisfied with the quality of search results. However, learning a specific query language to exploit a logic-based search system is regarded as a barrier for ordinary users. In addition, there are no inference engines available to manage a large knowledgebase with a reasonable period of time [17].

Link-based search systems [18,19] find relevant documents to the given query through the hyperlinks between the web documents. Most link-based search systems consist of two steps. Finding initial documents is based on the exact matching of keywords between the query and documents. In addition, the next step is to find relevant documents that are frequently referenced by the initial documents. As a result, it is possible to expand the search results to find the desired document although it does not contain any index terms derived from the query. However, link-based search systems have two limitations. First, if the hyperlinks between documents do not exist, the results of linked-based search are identical to those of keyword search, i.e., link-based search cannot work well if documents do not contain any hyperlinks. Second, since the links between documents are static, search results cannot dynamically respond to the changes of users' requirements. SimRank [20], PageRank [21] and their variants such as the Personalized PageRank [22] are representative approaches in link-based search. However, they do not consider the various types of relationships in linked data, and all links are regarded as having the same level of importance (i.e., same weight) in the similarity measurement [23].

To overcome the limitations of previous studies, we propose a graph structure that dynamically analyses the relationship between search results for a given query without any background knowledge of query languages.

3. Metadata-based Open-Government Data Platform

3.1. Background and Conceptual Model

Big data technologies have revolutionized many IR areas in recent years, and the make especially decision-making more reliable. Previously, as several decision-making have been significantly influenced by decision makers' intuition, such as past experience or observation, there are no needs to systematically collect and analyze large amount of data. Since the accuracy of data-driven decision-making has been improved and gained a lot of interests, some governments have started to adopt the data-driven policymaking or review its adoption. The central government of Korea has found that making a good decision is started from using good data, but it does not know the exact status of OGDs (such as the number, type and quality of OGDs) held by each government agency. This is because only OGDs, which government agencies intend to disclose to other agencies as well as citizens, have been registered in the Korean OGD portal just like other countries.

To rebuild the data governance from production to disposal, the Korean government has built a metadata-based open-data platform that allows the central government to accurately identify the number, type and location of data. Collecting metadata instead of raw data from government agencies offers several benefits. First, it is possible to reduce a data repository and collect raw data as needed. Second it is possible to increase the collecting interval for keeping the collected data up-to date. Lengthening the collecting interval can significantly reduce the load of network. In this paper, metadata refers to auxiliary data on systems, databases, tables and columns to infer the characteristics

(i.e., structure, owner, location, data types, etc.) of OGDs. To ensure government agencies comply with guidelines of the central government, an instruction has been established by the central government of Korea to standardize database construction and its management. In the instruction, 43 metadata have been defined as national standards, and some of metadata are as follows: agency name, system name, database name, database description, table name, table description, column name, column description, data type, and so on. Due to the size constraints of the paper, the rest of metadata and their detailed descriptions are omitted.

Figure 1 shows the conceptual model of an open-data platform and the data flow of the platform applied to the Korean government. The proposed open-data platform has four key features. First, whenever the metadata of OGDs in a government agency have created or modified, the changed metadata are automatically collected to the agency's metadata system and transmitted to the central metadata system in real time (or periodically according to system settings). Second, by referencing the standard terminology dictionary, the terms, types and formats of the collected metadata are converted into the standardized metadata to share metadata with others and to analyze the relationships of OGDs in the graph. Unless the standardization of terminology is performed, data nodes (i.e., vertices) with synonyms cannot be inter-linked in an OGD graph. Third, by automatically analyzing the relationship of data nodes, the visualized form of the graph can be provided to citizens who want to get useful information. For example, by using the OGD graph, it is possible to find a set of OGDs with a single query by exploring the estimated relationships of OGDs. In addition, it is possible to know the types of OGDs and the number of OGDs that each government agency has. As a result, it may be easier to request OGDs that are not yet available to citizens on their OGD portal. Finally, by collecting metadata and requesting raw data as needed, it is possible to integrate and analyze data models and raw data from heterogeneous domains. The analyzed results can be used for government officials to enhance the accuracy of policy execution, which is typically referred to as the data-driven policymaking.
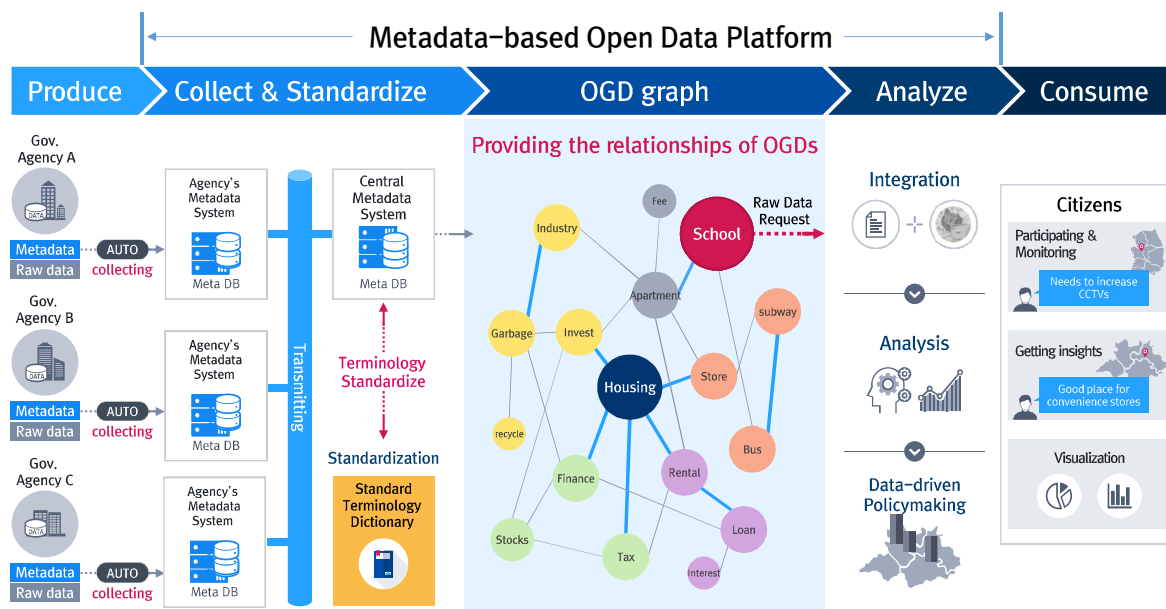

Figure 1. Conceptual Model of Metadata-based Open-Data Platform.

3.2. Automatic Collection and Standardization of Metadata

While the agencies of the central government and metropolitan cities have their own metadata systems, most local small cities do not have their own metadata systems as shown in Figure 2. In a rural area, the metadata system of a metropolitan city works as a central metadata system in the area, i.e., a metropolitan city metadata system automatically collects the metadata and raw data of local metadata systems in local small cities. By linking a metropolitan metadata system directly to the metadata

systems of the local small cities, it is possible to grant the local autonomy to the metropolitan area. In addition, by linking only the metadata systems of central government agencies and metropolitan cities to the central metadata system, it is possible to enhance the efficiency of management and reduce network loads.
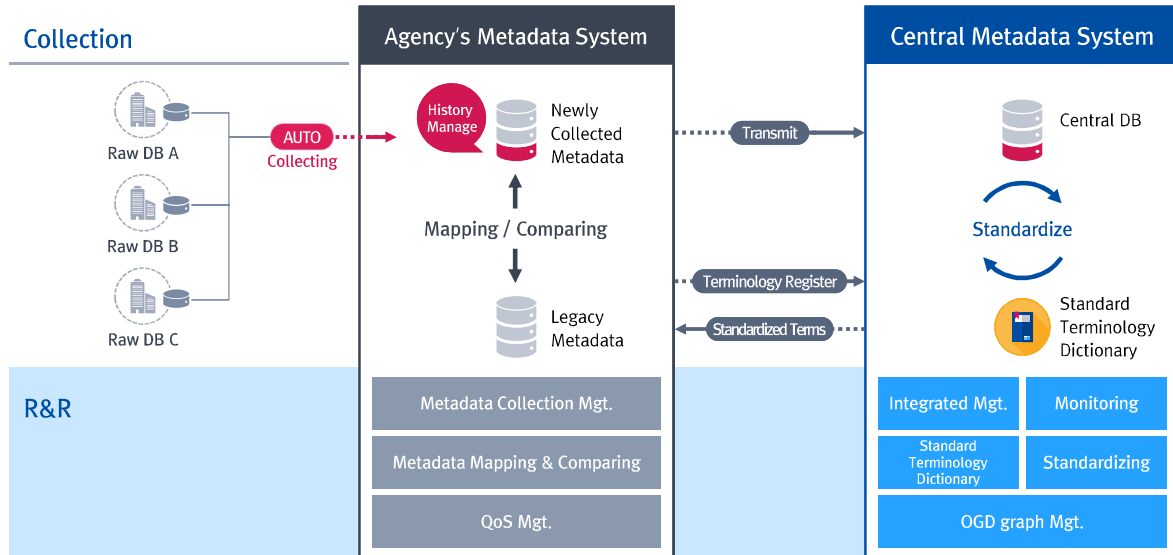

Figure 2. Automatic Metadata Collection by Agency Types.

Figure 3 shows a procedure for standardizing the collected metadata in the metadata systems of agencies. As previously addressed in Section 3, the terms, types and formats of collected metadata with others and accurately compute the relationships of OGDs by eliminating synonym problems. The metadata systems of agencies are responsible for standardizing the collected metadata. To execute the assigned role, the metadata systems in agencies check whether metadata terms are existed as standard terms (or representative terms) in the standard terminology dictionary, which is maintained by the central metadata system. If the terms in the collected metadata do not exist in the dictionary, the terms should be registered as standard terms.
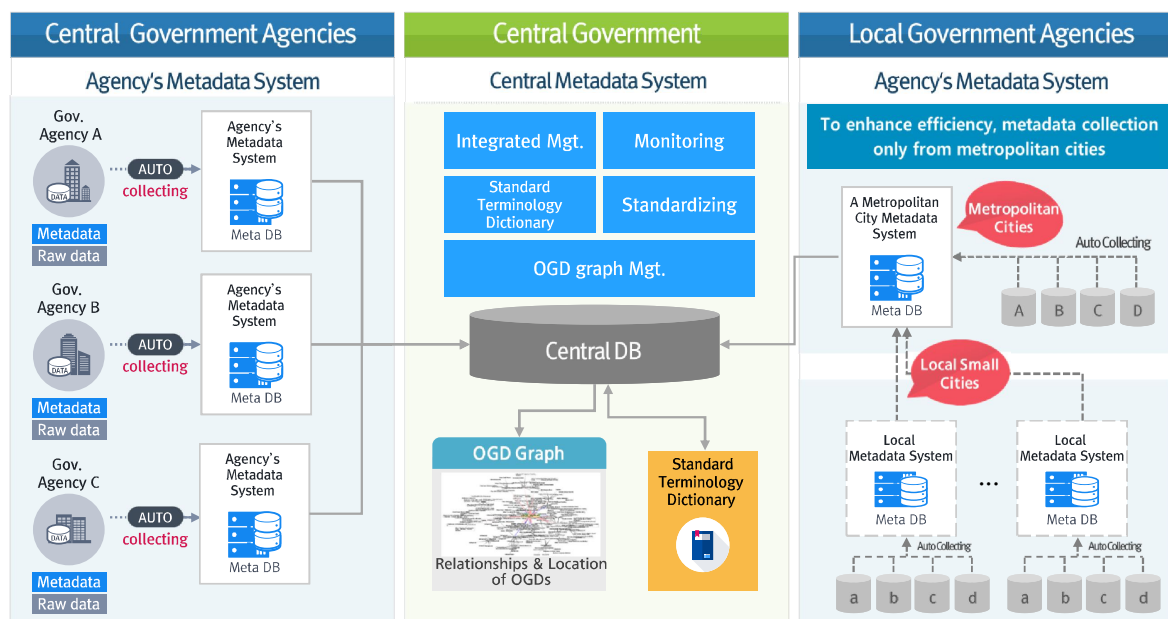

Figure 3. The Process of Metadata Standardization.

3.3. Modelling of Open-Government Data Graph

This section describes an approach for modelling the OGD graph, which is used in the proposed platform. Figure 4 shows an OGD graph structure that consists of data nodes (i.e., vertices) and their relationships (i.e., edges). Vertices represent the collected relations (i.e., tables) in the proposed platform, and a set of terms refers to compound words as well as single words that extracted from the data names (i.e., relation names), attributes and categories. Specifically, the compound words and single words extracted from data names, attributes and categories are obtained by applying Korean natural language processing operations, such as tokenization, stop-words removal and stemming [24] by referencing the standard terminology dictionary. Figure 5 shows an example of natural language processing operations with "locations of public parking lots" extracted from a data name. In fact, the proposed open-data platform executes Korean natural language processing operations, but this paper provides an example of English natural language processing operations to enhance readers' understanding. First, a preposition "of" is removed by stop-words remover. Then, Tokenizer extracts single words and compound words (such as "locations", "parking lots" and "public") by referencing the standard terminology dictionary which contains standard terms and their synonyms. After stemming the extracted plural words, the singular words such as "public", "parking lot" and "location" are provided as the results of natural language processing operations. Finally, the derived words are exploited to model the OGD graph, which is shown in Equation (2).
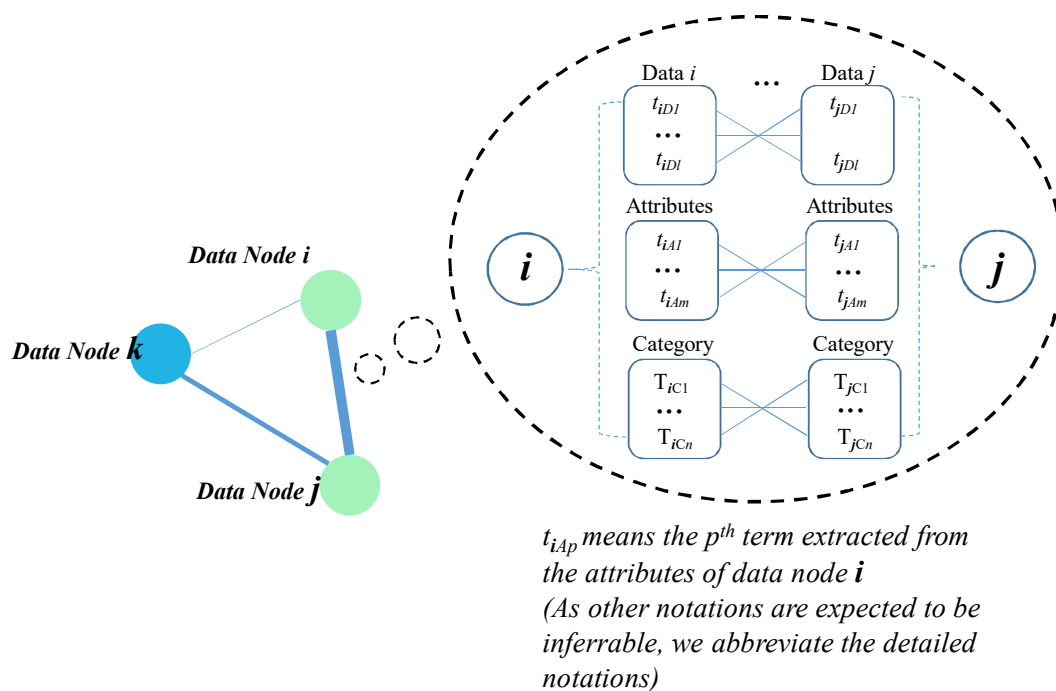

t_{iAp} *means the pth term extracted from the attributes of data node* **i**
(As other notations are expected to be inferrable, we abbreviate the detailed notations)

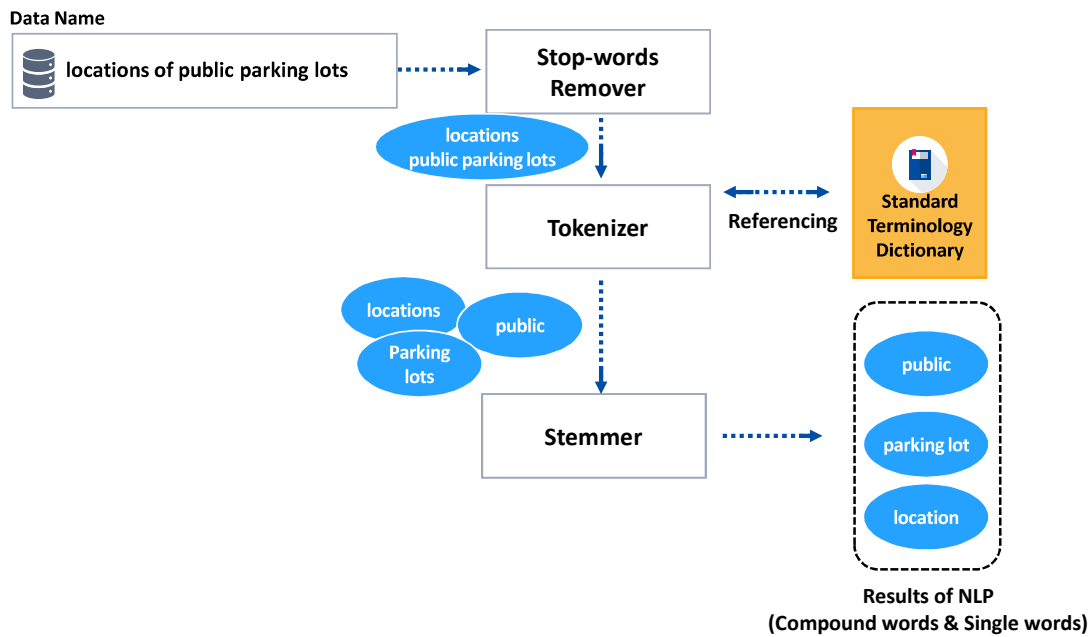**Figure 4.** An Example of the OGD Graph and Detailed Relationship.

Figure 5. An Example of Natural Language Processing.

An edge represents the many-to-many relationships derived from terms of two vertices. Thus, if there is an edge between two vertices, they are assumed to be semantically relevant in this paper. The thickness of an edge indicates the degree of similarity between two vertices, i.e., the thicker the edge, the higher similar between two vertices.

Definition 1. *Given an OGD graph G_{OGD} let V be a set of vertices, and $|V|$ be the total number of vertices in G_{OGD}.*

$$V = \{v_1, v_2, ..., v_{|V|}\} \tag{1}$$

where vertices are described by a set of index terms derived from the metadata such as the data names, attributes and categories. Assume that T_{iD}, T_{iA} and T_{iC} represent sets of index terms extracted from the data name, attribute and category of vertex i, the vertex i is formally defined as follows:

$$v_i = \{T_{iD}, T_{iA}, T_{iC}\} = \{t_{iD1}, ..., t_{i|Dl|}, t_{iA1}, ..., t_{i|Am|}, t_{iC1}, ..., t_{i|Cn|}\} \tag{2}$$

where the xth, y^{th} and z^{th} index terms extracted from T_{iD}, T_{iA} and T_{iC} of vertex i are represented by t_{iDx}, t_{iAy} and t_{iCz}, respectively.

Definition 2. *Given an OGD graph G_{OGD}, let E be a set of edges which are unordered pairs of vertices.*

$$E \subseteq \{(v_i, v_j)|(v_i, v_j) \in V^2 \wedge v_i \neq v_j\} \tag{3}$$

If two distinct vertices v_i and v_j have the same terms, the edge (v_i, v_j) consisting of two vertices has a degree of similarity. Since vertices are modelled as term sets, the Jaccard similarity coefficient is used to measure the similarity between two distinct vertices.

$$sim(v_i, v_j) = \alpha \times \frac{|T_{iD} \cap T_{jD}|}{|T_{iD} \cup T_{jD}|} + \beta \times \frac{|T_{iA} \cap T_{jA}|}{|T_{iA} \cup T_{jA}|} + \gamma \times \frac{|T_{iC} \cap T_{jC}|}{|T_{iC} \cup T_{jC}|} \tag{4}$$

where $\alpha + \beta + \gamma = 1$. T_{iD}, T_{iA} and T_{iC} represent term sets extracted from the data name, attribute and category of i, respectively.

In general, most information retrieval systems estimate the proximity of entities (e.g., queries and documents) by exploiting TF-IDF (term frequency-inverse document frequency), which is the main idea of Vector Space Model. Since the features of the targeted document are represented well when the length of a document is long, TF-IDF-based similarity functions show superior performance compared to other metrics. However, in our approach, the values of term frequencies are low (in fact, most term frequencies are less than two). This is because the index terms are extracted from non-long data name, attributes and categories. Therefore, we choose the Jaccard similarity coefficient to estimate the similarity of two OGDs in Equation (4). The Jaccard similarity coefficient is known to be suitable for estimating similarity in set-based models and is widely used in various information retrieval systems due to its simplicity.

Moreover, in Equation (4), the weight α, β and γ are 0.1, 0.05, and 0.85, respectively. The initial weight values (i.e., 0.1, 0.05 and 0.85) were set by the decision of officials in charge in accordance with their internal policies. However, the weight setting for data name, attribute and category terms in Equation (4) was determined by the user satisfaction survey for vertex associations in the OGD graph. By setting different weights for each data name, attribute and category term, 20 users who understood the purpose of the experiments rated the vertex associations as zero or one (i.e., one indicates the vertices are semantically associated with others). However, we need to validate the weight setting more precisely for non-biased users through experiments on various aspects such as the types of vertices, number of extracted terms, and so on.

4. Results and Discussion: An OGD Graph on the Korean OGD Portal

This section describes a use case of OGD graph applied to the Korean OGD portal. The vertices of the OGD graph consist of two types of OGDs. One is the data already available to citizens on the Korean OGD portal, and the other is data that can be provided whenever citizens request but not yet available on the portal. We refer to the data already available to users as Opened OGD, and data not yet available as Openable OGD to clarify the two types of data in the following description.

In Figure 6, openable OGDs are represented by colored vertices, while opened OGDs are represented by colored vertices with arrows. That is, the OGDs with arrows can be immediately downloadable on the portal. The size of a vertex indicates how relevant the data is to a given user query, and the relevance of the vertex to the query is derived from the applied search engine. When a user clicks a vertex on the graph, he/she can see details of OGD such as the category of the OGD, government agency name, descriptions and keywords. To access the OGD graph with 89,501 vertices on the Korean OGD portal, please visit the portal. Currently, the knowledge graph is only available in Korean [25].

Figure 7 shows a procedure for requesting and delivering raw data. In this example, it is assumed that a user finds raw data of heterogeneous domains to perform decision-making. First, the user searches for OGDs of heterogeneous domains with a single query in the OGD graph, and he/she selects some OGDs for his/her decision-making. Second, the user fills down and submits the raw data request if the selected OGDs are not yet available. Third, a database administrator (shortly, DBA) reviews the raw data request submitted by the user. If the request is appropriate and there is no data limitation, the DBA will extract and upload the raw data to the agency's metadata system. Fourth, after anonymizing personal information such as the social number, mobile number and so on, the raw data is stored in the repositories of the central metadata system. Finally, the raw data is analyzed in a variety of ways to make the correct decision. The data analysis methods such as Text Mining, Statistics and Machine Learning can be selected based on the features of the delivered raw data or the purposes of applications.
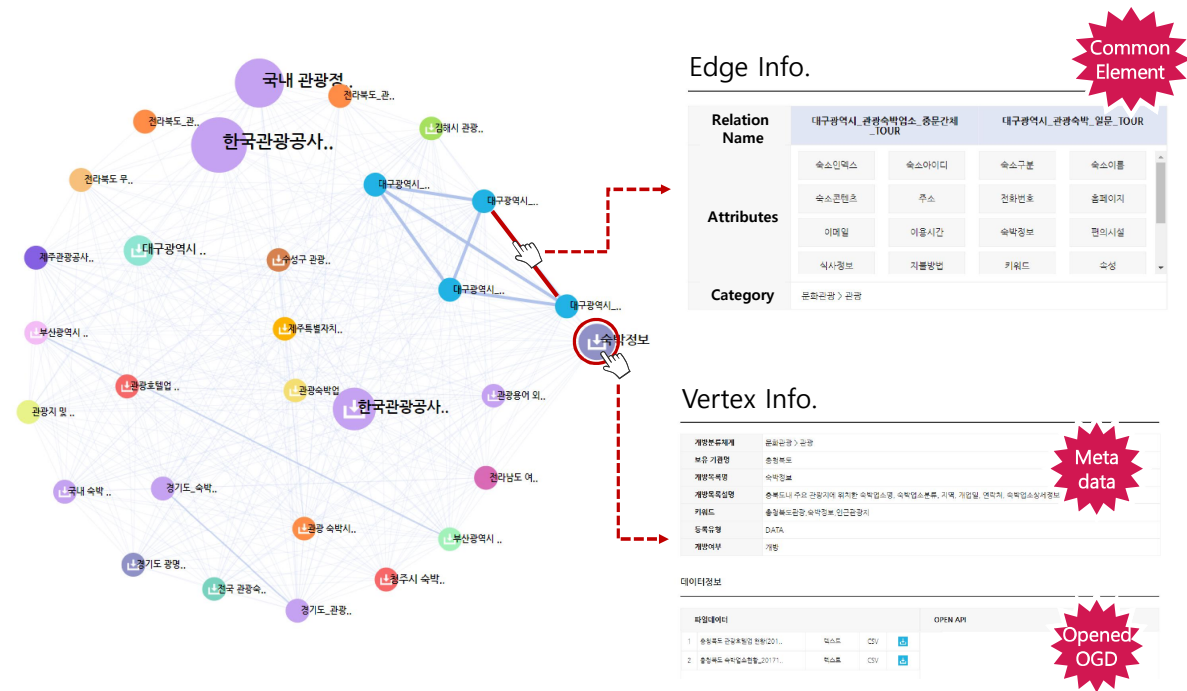
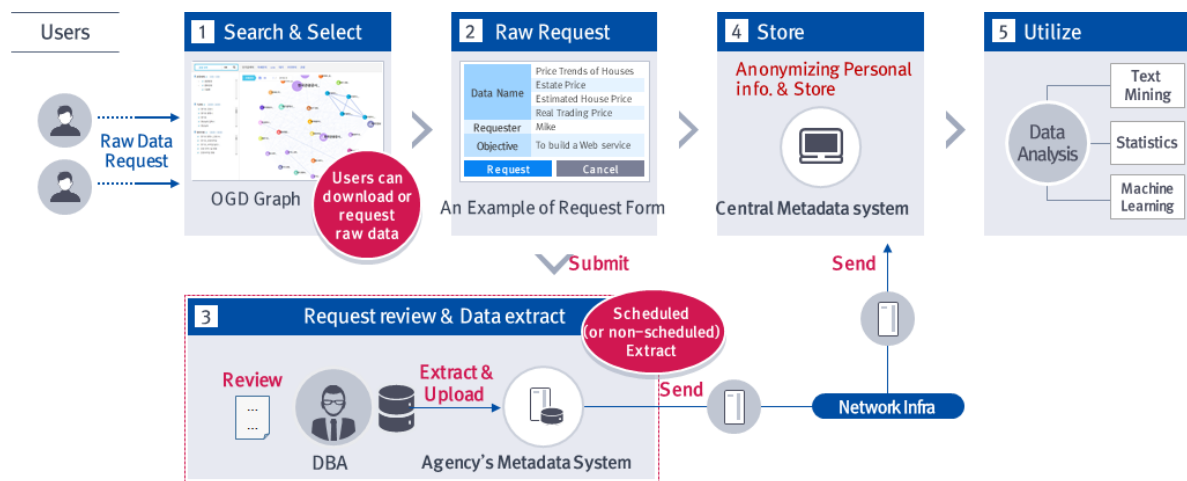**Figure 6.** The OGD Graph Applied to the Korean OGD portal.

Figure 7. Raw Data Request and Delivery Procedure.

The focus of this paper is how to collect and organize the OGDs held by government agencies for citizens to easily access the desired OGDs (not how to analyze the delivered raw data). Thus, Figure 8 shows an example of the advantages compared to when the proposed open-data platform and an OGD graph do not exist. We assume that users want OGDs related to "real estate" to develop their own applications. Before the proposed open-data platform existed, users could only exploit OGDs that are already opened in the OGD portal and contain the search query (i.e., "real estate"). If the users want additional OGDs (e.g., "real estate with good public traffics"), they need to search other OGD portals based on their knowledge (or intuition) of the desired OGD location. In addition, official servants may approve or reject the OGD provisions in accordance with their internal policies (or the requester). However, providing users with the proposed open-data platform will allow them to search not only opened but also openable OGDs. In particular, by using the OGD graph, it is possible to find the additional OGDs (such as "public traffics", "schools", "Managing fees", and so on) with the given query at one place.

Figure 8. Before and After Comparison of Open-Data Platform Presence.

Moreover, as the whole openable OGDs are shown in the OGD graph, we can expect a significant increase in the number of raw OGDs requests from users. Since the goal of the Korean government is to promote the use of OGDs significantly and to revitalize the data industries in Korea, the OGD graph was planned and released to enhance citizens' accessibility to OGDs. In our work, we assume that improving the citizens' accessibility to OGDs is directly proportional to increasing their OGDs requests and usage. The raw data request log on the open-data platform is analyzed to verify whether our assumption is correct or not. The OGD graph described in this paper was first released to citizens in April 2019, thus 10,150 raw data requests in the same period (i.e., from April to August) of 2018 and 2019 are collected and analyzed to see the effects of an OGD graph release. As shown in Figure 9, for all the given time periods, there are more raw data requests in 2019 than in 2018. Specifically, the monthly average of raw data request in 2019 is about 73% higher than one in 2018, which is highlighted with the red box in Figure 9. This is because the whole OGDs that are already available or not yet available are shown to citizens. In this section, our analysis revealed that showing the whole OGDs held by government agencies (i.e., improving user accessibility to OGDs) could significantly increase the usage of OGDs in private sectors.
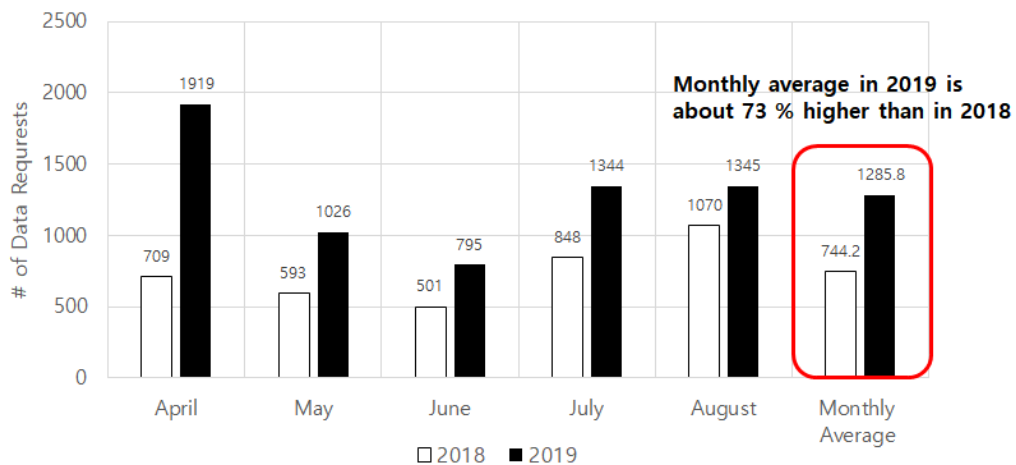

Figure 9. Comparison of Raw Data Requests in 2018 and 2019.

5. Conclusions

In this paper, we proposed the metadata-based open-data platform and an OGD graph. Our approach offers three improvements over the previous OGD portals of other countries: (1) enhancing user accessibility to the entire OGDs held by government agencies, (2) making users find the relevant OGDs as well as the desired OGDs with a single query, and (3) enabling data-driven decision-making. First, ordinary citizens (including government officials) cannot know which government agencies produce and hold openable data if the agencies do not disclose their data. However, by linking the whole government agencies as shown in Figure 2 and automatically collecting the metadata of OGDs held by government agencies, it is possible for citizens to exactly see how many OGDs agencies have in real time. Second, users can browse the OGD graph and find relevant OGDs with a single query by inter-linking OGDs in heterogeneous domains and analyzing their relevance. Occasionally, unexpected discovery may occur while users exploring vertices and edges in the OGD graph. Finally, the barriers for citizens to obtain OGDs for making the correct decisions become significantly reduced. It is worth noting that we have taken an initial step to build a comprehensive data platform of governmental information with openness. As the proposed approach is applied to a real service of the public sector in Korea, the reliability and feasibility for constructing the metadata-based open-data platform and the OGD graph are estimated to be high. However, there remains several limitations in our initial platform despite we have addressed its feasibility in some Korean government agencies. In future studies, it is necessary to execute various experiments to ensure the generality of the proposed platform and OGD graph: First, in measuring the similarity between vertices, parameter settings should be verified by expanding the number of users involved and types of index terms. Among the 43 metadata defined as national standards, only index terms derived from data names, attributes and categories are used in our approach. Second, various similarity measures need to be compared via experiments. As we modelled vertices as a set of index terms addressed in Section 5, we adopted the Jaccard Coefficient due to its simplicity. However, we need to enhance the accuracy of the OGD graph by comparing various similarity measures such as Vector Space Model, Probability Model, and so on. The higher the accuracy of the OGD graph, the easier users' access to the whole OGDs held by government agencies.

Author Contributions: All authors made contributions to this work. J.P. contributed to the organization of the research as well as the final manuscript preparation. J.-w.L. propose d the original idea, conducted data processing, wrote the original draft and discussed the data analysis with experimental results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. OECD. *Data-Driven Innovation Big Data for Growth and Well-Being (Electronic Resource)*; OECD Publishing: Paris, France, 2015.
2. Janssen, M.; Charalabidis, Y.; Zuiderwijk, A. Benefits, adoption barriers and myths of open data and open government. *Inf. Syst. Manag.* **2012**, *29*, 258–268. [CrossRef]
3. Tammisto, Y.; Lindman, J. Definition of open data services in software business. In Proceedings of the International Conference of Software Business, Cambridge, MA, USA, 18–20 June 2012; pp. 297–303.
4. Kalampokis, E.; Tambouris, E.; Tarabanis, K. Linked open government data analytics. In Proceedings of the International Conference on Electronic Government, Koblenz, Germany, 16–19 September 2013; pp. 99–110.
5. Gascó-Hernández, M.; Martin, E.G.; Reggi, L.; Pyo, S.; Luna-Reyes, L.F. Promoting the use of open government data: Cases of training and engagement. *Gov. Inf. Q.* **2018**, *35*, 233–242. [CrossRef]
6. Vetrò, A.; Canova, L.; Torchiano, M.; Minotas, C.O.; Iemma, R.; Morando, F. Open data quality measurement framework: Definition and application to Open Government Data. *Gov. Inf. Q.* **2016**, *33*, 325–337. [CrossRef]
7. Dawes, S.S.; Vidiasova, L.; Parkhimovich, O. Planning and designing open government data programs: An ecosystem approach. *Gov. Inf. Q.* **2016**, *33*, 15–27. [CrossRef]

8. Lourenço, R.P. An analysis of open government portals: A perspective of transparency for accountability. *Gov. Inf. Q.* **2015**, *32*, 323–332. [CrossRef]

9. Lenat, D.B.; Guha, R.V. Building Large Knowledge Based Systems: Representation and Inference in the Cyc Project; Addison Wesley Longman Publishing: Boston, MA, USA, 1989.

10. Xu, S.; Bao, S.; Fei, B.; Su, Z.; Yu, Y. Exploring folksonomy for personalized search. In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore, 20–24 July 2008; pp. 155–162.

11. Lee, J.W.; Kim, H.J.; Lee, S.G. A probability-based unified framework for semantic search and recommendation. *J. Inf. Sci.* **2013**, *39*, 608–628. [CrossRef]

12. Fikes, R.; Hayes, P.; Horrocks, I. OWL-QL—A language for deductive query answering on the Semantic Web. *Web Semant. Sci. Serv. Agents World Wide Web* **2004**, *2*, 19–29. [CrossRef]

13. Kim, H.H. ONTOWEB: Implementing an ontology-based Web retrieval system. *J. Am. Soc. Inf. Sci. Technol.* **2005**, *56*, 1167–1176. [CrossRef]

14. Kumar, V.K. Knowledge Representation Technologies Using Semantic Web. In *Web Services: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2019; pp. 1068–1076.

15. Suchanek, F.M.; Kasneci, G.; Weikum, G. Yago: A core of semantic knowledge. In Proceedings of the 16th International Conference on World Wide Web, Banff, AB, Canada, 8–12 May 2007; pp. 697–706.

16. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. Dbpedia: A nucleus for a web of open data. In *The Semantic Web*; Springer: New York, NY, USA, 2007; pp. 722–735.

17. Lee, T.; Shim, J.; Lee, H.; Lee, S.G. A pragmatic approach to model and exploit the semantics of product information. *J. Data Semant. VII* **2006**, *4244*, 242–266.

18. Balmin, A.; Hristidis, V.; Papakonstantinou, Y. Objectrank: Authority-based keyword search in databases. *VLDB Endow.* **2004**, *30*, 564–575.

19. Rocha, C.; Schwabe, D.; Aragao, M.P. A hybrid approach for searching in the semantic web. In Proceedings of the 13th International Conference on World Wide Web, New York, NY, USA, 17–20 May 2004; pp. 374–383.

20. Jeh, G.; Widom, J. SimRank: A measure of structural-context similarity. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23–26 July 2002; pp. 538–543.

21. Brin, S.; Page, L. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **1998**, *30*, 107–117. [CrossRef]

22. Haveliwala, T.H. Topic-sensitive pagerank. In Proceedings of the 11th International Conference on World Wide Web, Honolulu, HI, USA, 7–11 May 2002; pp. 517–526.

23. Meymandpour, R.; Davis, J.G. A semantic similarity measure for linked data: An information content-based approach. *Knowl.-Based Syst.* **2016**, *109*, 276–293. [CrossRef]

24. Baeze-Yates, R.; Ribeiro-Neto, B. *Modern Information Retrieval*; Addison-Wesley Longman Publishing: Boston, MA, USA, 1999.

25. Ministry of Interior Safety. Available online: https://www.data.go.kr/ndm/view.do (accessed on 30 September 2019).