



A Knowledge Graph for Eldercare: Constructing a Domain Entity Graph with Guidelines

You Duan¹, Pin Ji², Liuqi Jin³, Anning Zou³, Jiaoyun Yang^{3(✉)},
Hong Xie⁴, and Ning An³

¹ Institute of Industrial and Equipment Technology,
Hefei University of Technology, Hefei, China

² School of Software, Hefei University of Technology, Hefei, China

³ School of Computer and Information,
Hefei University of Technology, Hefei, China
jiaoyun@hfut.edu.cn

⁴ School of Nursing, Peking University, Beijing, China

Abstract. The sheer size of aging population and the complexity of individual eldercare needs have put great social and economic burdens on various societies worldwide. Particularly, the daunting demands for caring for elders are unmet by the overall supply of caring capacities. There are at least two aspects of this imbalance: first, the number of high quality caring professionals with nursing backgrounds have been insufficient and this trend will continue in the near future; second, most of the current professional and semi-professional caregivers have limited education that in turn limited their caring capacities. To tackle these issues with the rapid development of data science and growing awareness of nursing informatics, we are working on building a knowledge graph for eldercare. In this paper, we propose the concrete first step towards this goal: constructing an eldercare entity graph based on international guidelines, and this is first of its kind to our best knowledge. Protégé is used to construct the entity graph that consists of eight types of entities and four types of relationships. In this graph, a standard eldercare procedure is represented as a path between an entity denoting an eldercare service classification and an entity denoting an eldercare activity. About 230 procedures could be derived according to the guidelines.

Keywords: Knowledge graph · Entity graph · Eldercare · Guideline

1 Introduction

According to the 2017 Revision of World Population Prospects, which was the 25th United Nations official population estimate and projection prepared by the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat in 2015, the global population of people aged 60 and over accounted for 12.3% of the total. The population aged 65 and over accounted for 8.3% of the global total. It is estimated that by 2020, the population aged 60 and over will account for 13.5% of the total population the population aged 65 and over will account for 9.4% of

the total population, which is far more than the standard internationally recognized to enter an aging society. An aging society is one in which the population over the age of 60 accounts for 10% of the total population or in which the population over the age of 65 accounts for 7% of the total population. Aging is a challenging problem to the vast majority of countries. The annual statistics bulletin from the People's Republic of China and the State Statistics Bureau in 2016 shows that as of the end of 2016, the population of people over the age of 60 in Mainland China totaled 230.86 million, accounting for 16.7% of the total population in the country, which is 0.6% higher than the 2015 numbers. Obviously, the problem of aging in China is very serious. As the aging population grows, so does the problem of pension-related issues, and one problem is the difficulty meeting the needs of life care for the large aging population. In China, the pension models include institutional pensions, home-based care, land pensions, social security pensions. More elderly people choose a home for the aged instead of choosing a pension institution for continued care, so this pension model has a larger professional gap with respect to nursing services. To meet the need of the elderly for nursing services, the new pension model, community service for home care of elderly, came into being. This model combines home-based care for the aged and community service to ensure that the elderly, their children, pension service personnel, and the government can meet their own needs. The model employs resources to make up for the lack of home-based care for the aged; it is a new type of pension model strongly advocated by the government.

The key to community service for the home care of the elderly is the provision of nursing services. There are numerous types of nursing services, which can be broadly divided into life care, health care, emergency aid, spiritual comfort, culture and entertainment services, and legal rights protection. Each of these categories contains numerous specific services. Due to current demand, the large and diversified resources, the uneven level of employees and the considerable differences in quality of nursing services, the contradiction between supply and demand for these services is prominent. Effectively integrating service resources and improving nursing levels have become urgent problems. The purpose of the traditional nursing service delivery program is the manual determination of the actual needs of the elderly by professional staff after receiving requests for nursing care. The nursing service is then compared with available community services, suitable services are selected, and then, the patient is sent to the appropriate service personnel. However, with the increase in the material and spiritual needs of the elderly, service item requirements of nursing services have also increased, which were originally simple and general classifications that have become more detailed and numerous. Currently, services simply rely on an artificial dispatch service, which has been unable to meet the demand. However, with the increase in the material and spiritual needs of the elderly, the service items of the nursing services have further increased. Nursing services could be simply divided into several categories, but now, they have become more detailed and numerous and have been unable to meet the demands by relying on an artificial dispatch service.

Information technology applied to nursing provides a solution. Informatization refers to the historical process that cultivates and develops innovative productivity represented by intelligent computer-based tools that benefit society. Nursing-related information can be comprehensively, quickly and easily transmitted to all who need it

after aggregating care-related information into a database. The Nursing Information System [1] in Taiwan, which has been endorsed by approximately 95% of the local medical institutions, can help nurses provide or acquire accurate and real-time clinical information for patients, physicians or other healthcare providers [2]. Another example is the millions of electronic medical records handled by hospitals in the United States and other countries each year. These records not only contain a variety of clinical information including medical history, physical examination results, laboratory reports, radiology reports, surgery reports and discharge summaries, but they also provide medical information about patients, treatment and clinical processes [3]. However, current nursing informatization mostly focuses on the establishment of a knowledge database, real-time information tracking and a service platform [4]. With the lack of automatic generation tools for nursing plans, a knowledge graph can describe an automated production nursing plan.

A knowledge graph is a large knowledge base of organizational knowledge in the framework of a semantic network, and the scattered knowledge resources are interconnected at the conceptual level [5]. It is the essential composition of many web services [6, 7]. Its basic unit is the “Entity - Relationship - Entity” triple as well as the entity and its related properties. Entities are interconnected through relationships to form a reticular knowledge structure. The structure of the knowledge graph includes the logical structure of the knowledge graph itself and the technical architecture employed to construct the knowledge graph. Logically, the knowledge graph can be divided into the following 2 layers: the data layer and the pattern layer. In the data layer, knowledge is kept in graph databases as units of fact. The pattern layer is at the heart of the knowledge graph, which stores the refined knowledge. The ontology library is utilized to manage the pattern layer of the knowledge graph [8]. With the help of the support for axioms, rules and constraints from the ontology library, we can identify the entities, relations and relationships among the types of entities, attributes and other objects. The status of the ontology library in the knowledge graph is equivalent to that of the mold in the knowledge base. The knowledge base with the ontology library has less redundant knowledge.

Knowledge graph theory is a method often used in the field of natural language processing. Researchers at Twente University and Groningen University initiated the study of knowledge graph theory in 1997. The research work was continued by Professor Hurd of Twente University. The research analyzed text with a knowledge graph [9] and was a new way to describe the language that humans have studied extensively in the semantics of textual forms. The research also examined a different web and the use of computers for field research to understand human languages. The research has an important role in the study of knowledge graphs in information formation [10]. According to the scope of knowledge contained in separate knowledge bases, the current knowledge graph in the Internet domain can be divided into the following three types: the domain-independent knowledge graph, the domain-specific knowledge graph, and the cross-language knowledge graph. The domain-independent knowledge graph is usually not limited to specific areas, including knowledge involved in various fields and industries in various branches with a wide range of content that consist of DBPedia, Yago, Freebase, WikiTaxonomy, etc. The domain-specific knowledge graph is generally targeted to specific areas. Although the breadth of content is not as broad as

that of a domain-independent knowledge graph, it contains more comprehensive and targeted knowledge on the depth of the domain knowledge. Representatives of this type include HowNet, HNC, Geonames, and Linked Movie Database. The cross-language knowledge graph is generally a knowledge base containing two or more languages, and examples of these include XLORE, BabelNet, and WordNet. The knowledge graphs in the medical field belong to the domain-specific knowledge graph type and have many practical applications. For example, Google has established a knowledge graph based on medical information [11, 12] for their Web search engine [13–15]. Many biomedical knowledge graphs have been published online as applicable data [16], such as Drug-Bank [17], ICD-9 [18], Bio2RDF [19] and Chem2Bio2RDF [20]. There is likewise a great deal of research on knowledge graphs and related technologies in the field of medical science. For example, these technologies use basic concept extraction to automatically form high-quality health knowledge graphs directly from medical records [21] and study how to automatically retrieve knowledge from knowledge graphs with high accuracy [22].

There are two separate ways to build a knowledge graph, top-down construction and bottom-up construction. Top-down construction refers to the use of structured data sources, such as encyclopedia-based websites, to extract ontology and schema information from high-quality data and to be added to the knowledge base. Bottom-up construction is based on certain technical means to extract the resource model from the publicly collected data. The process then selects the new model with a higher confidence level and adds it to the knowledge base after manual review. In the early stage of the progress of knowledge graph technology, due to the inadequacy of knowledge extraction and processing technology, most participating enterprises and research institutes adopted a top-down approach. However, as the technology is still maturing, most current knowledge graphs use the bottom-up approach. This paper uses the bottom-up construction approach, which is an iterative process that updates itself. Each round of updates includes the following three steps: (1) information extraction to extract entities, properties and the relationships among entities from various types of data sources, forming ontological expression of knowledge; (2) knowledge fusion that integrates new knowledge to eliminate conflicts and ambiguities [23]; and (3) knowledge processing, which is quality assessment done on the fused knowledge, and the qualified parts will be added to the knowledge base to ensure the quality of the knowledge base.

It can be seen from the above steps that the first step in building a knowledge graph in the nursing field is to extract the pertinent information to clarify the entities and the relationships among them. Thus, this article chooses to crawl nursing guidelines, and through entity relation and attribute extraction, a series of different sets of entities and the relationship edge sets of them are obtained. Then, this research uses the visualization tool Protégé, developed by the Stanford University School of Medicine, to build a guideline-based network of nursing entities and to facilitate the implementation of further work for building knowledge graphs.

2 Definition of the Nursing Entity Relationship Network

A nursing entity is a sort of object in the nursing process, such as measures taken by nursing, the performance characteristics of a disorder, the function of human organs, or problems that belong to the field of nursing. The entities are not independent of each other, and most have some connection with each other. For example, “impaired memory results in decreased transient memories and forgetfulness in patients.” This sentence shows a causal relationship between impaired memories and reduced transient memory and forgetfulness. The relationships connect entities to form a nursing network. We crawled various professional medicine literature and analyzed their texts, and we obtained a series of different entity sets V and the connected edges set E between different entities in E . Then, the directed graph $G(E, V)$ is created as a network of nursing entities.

The nodes and edges of graph G represent entities and relationships, respectively. In the relationship network, each node is utilized to represent an entity, and the edges between nodes are used to represent the relationship between two entities. The relationship between entities can be both one-way and two-way.

Nodes in entity set V mainly represent the following types of entities: entities classified in the field of nursing such as “physical areas,” types of nursing problems such as “sensations and pains,” nursing problems such as “hearing impairment,” problem codes such as “A101101,” existing symptoms such as “Listening to normal intonation difficulties,” outcomes of symptom improvement such as “sensory function: auditory,” methods of symptom improvement such as “promoting communication: hearing impairment,” and entities of nursing activities such as “paying attention to the conversation.”

In discrete mathematics, we define the relationship in this way, with set A and set B . For $x \in A$, $y \in B$, and if $(x, y) \in R$, x and y have a relationship R . In our relationship network, we have defined some of the relationships that we consider to be more important in this network, such as “classification,” “identification,” “performance,” “relationship,” “improvement” and “take action.”

We define the “classification” relationship to show that an entity can be divided into multiple sub-entities.

The relationship between the classification of the nursing field entity and the type of nursing problem entity is a classification relationship; for example “physiological field” entities can be subdivided into “perception and pain”, “cognition and consciousness”, “language communication”, “respiratory function” and other entities.

The relationship between nursing measures entity and nursing activities entity is a classification relationship; for example, “promoting communication: hearing impairment” entities can be divided into “noticing speech content”, “using simple words and short sentences” and other entities.

The relationship between the type of nursing problem entity and the nursing problem entity is a classification relationship; for example, “sense and pain” entities can be divided into “auditory handicap”, “visual impairment” and other entities.

We use the “identification” relationship to show that an entity is the identity or code of an entity.

The relationships among the signs of existing symptoms is the identification relationship. For instance, the “A101101” entity is used to identify the “difficulty to listen to the normal intonation” entity.

We define the “expression as” relationship to represent symptoms of how a nursing problem behaves; for example, the nursing problem named “hearing impairment” is expressed as symptoms such as “hearing normal intonation difficulties” and “abnormal results of hearing screening test”.

We define “indication” in the relationship to represent the symptoms that can indicate a nursing problem; for example, the symptom of “hearing the normal intonation” reflects the nursing problem of “hearing impairment”.

To simplify the relationship among the three entities that consist of existing statuses/signs, improved outcome and step increase, we divide the entity into 22 relationships.

The relationship between the existing statuses/signs and the improved outcome is called the “improvement” relationship; for example, the symptoms of “listening to normal intonation difficulty” can introduce an improved “sensory function: auditory” outcome.

The relationship between the improved outcome and the step increase is called the “take measures” relationship; for example, the improvement to “sensory function: auditory” outcome requires the adoption of “promoting communication: hearing impairment” measures.

3 Method of Building a Nursing Entity Network

The software that visualizes the oriented graph G of the nursing entity is Protégé, which was developed by Stanford University School of Medicine. We can import the entity set E and the relation set V into the OWL system in Protégé. The software then parses out the entities and the relationships between them before showing the user.

3.1 Entity Generation

Each entity in Protégé is referred to as a class. After an entity is obtained, OWL will have the corresponding class for it. The entity operation is actually the operation on the class, and each class enjoys its own attributes. For example, the name of the entity corresponds to the label attribute of the class. In the visualization process, the name of the entity is of great importance. The appropriate label is inserted for each entity, which is convenient for subsequent review and maintenance. In addition, the relationships between entities are further represented by the entities themselves. A class representing an entity A can have one or more subclasses. After visualization, the user can see other entities that are related to entity A (Fig. 1).

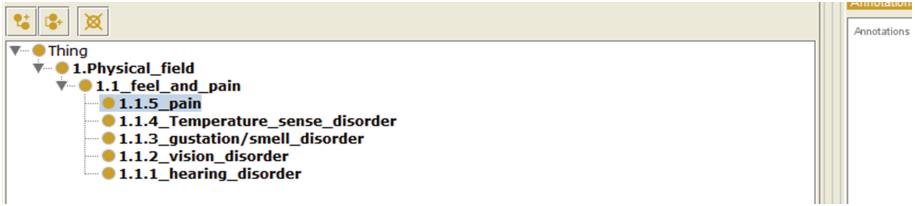


Fig. 1. Examples of multi-level relationship among entities and attributes

3.2 Setting the Relation

The visualization of the entity set E is recognized in multiple classes, and their relationships are embodied as attributes and nests of the class. The default relationship between the classes and subclasses in Protégé is inclusive if, for example, class A has a subclass B . B is displayed as B after visualization. However, there are numerous relationships in the set of relations V other than the inclusion. In addition, such circumstances might occur where “ B is a subclass of A but B leads to A ” (Fig. 2).

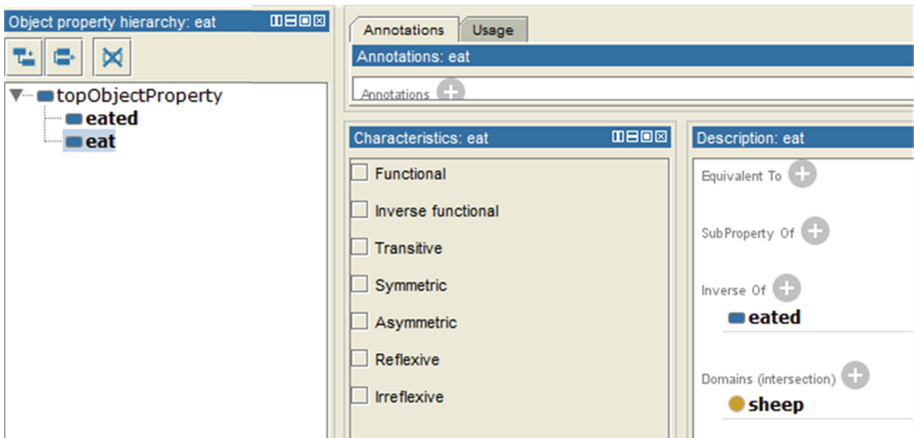


Fig. 2. Relationship properties and relationships among the relations

A relationship is reflected in the creation of the entity, but the specific relationship between classes needs to be created. In Protégé, the relationship properties in Object Properties can be defined, and the relative scope and the relationships among the other relations can be provided. After the visualization process, the relationships among diverse entities are presented (Fig. 3).

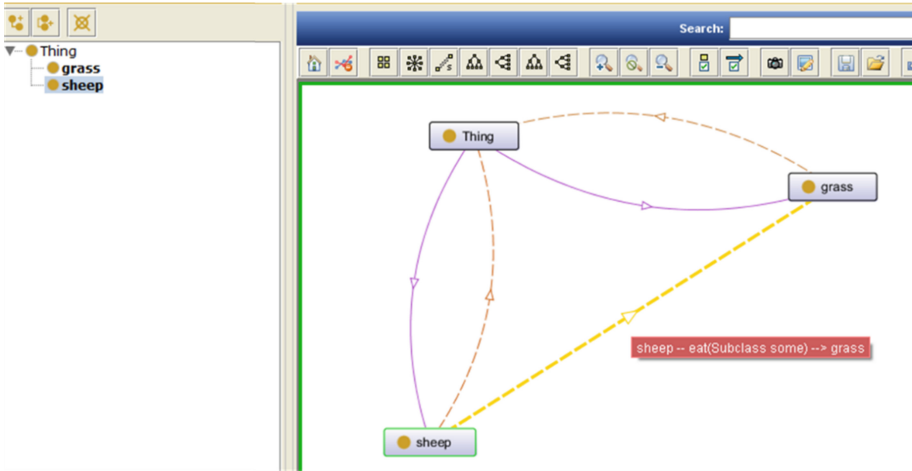


Fig. 3. Example of the relation: sheep eats grass

3.3 Establish Entity Network

After generating all the entities in the entity set E and all the relationships in the relation set V, the network construction of the nursing entities and the visualization accomplished by Protégé are basically complete (Fig. 4).

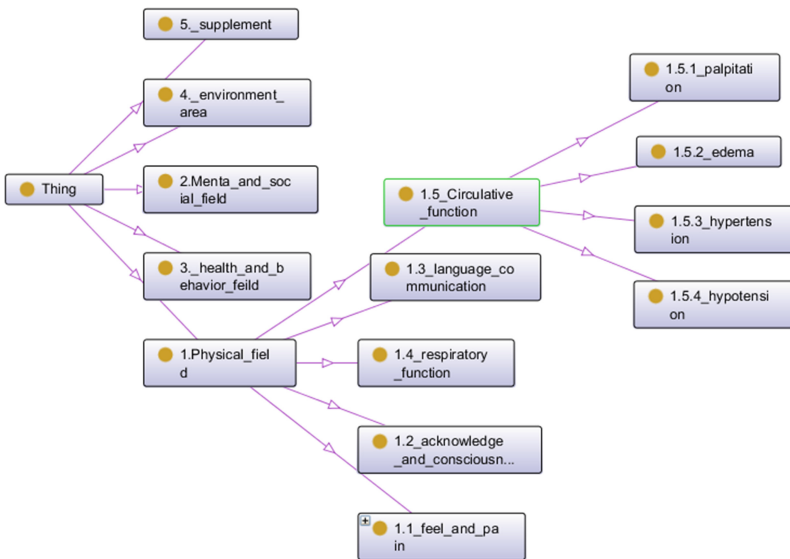


Fig. 4. Nursing entity network example

It is easily acknowledged that the nursing network can be divided into physiological areas, psychosocial areas, etc. according to atypical symptoms of the patients. Each area has corresponding functions. For example, the physiological collar contains breathing, circulation, excretion, feeling, etc. Moreover, each function contains its own analogous symptoms and solutions. The nursing network is presented by these different kinds of solutions.

4 Experimental Result

In the final nursing network, we categorized the symptoms and living conditions of 230 different patients and their corresponding nursing interventions into the following five categories: Physiology, Psychology, Health, Environment and Additional Supplements; each category contains its own sub-categories. For example, there are nine sub-categories in the Physiology category including sensation, cognition, language, skin, respiratory function, circulatory function, digestive function and excretion function. Each sub-category contains specific nursing problems, existing symptoms, symptom results, and related nursing measures. These symptoms and measures together constitute the nodes in the nursing network, and the relationships between these nodes are the edges of the network. We have 94 inclusive relationships, 208 symptoms with their problematic relationships, and 208 measures to consider in a relationship. This relationship is actually caused by three types of nodes, which are the consequences of the symptoms, the rules embodied by the results, and the measures taken. The node that contains the rules embodied by the results reflects which symptoms can cause the result and is based on the results respond with appropriate measures. The relationships among these nodes is the edge of the nursing network (Fig. 5).

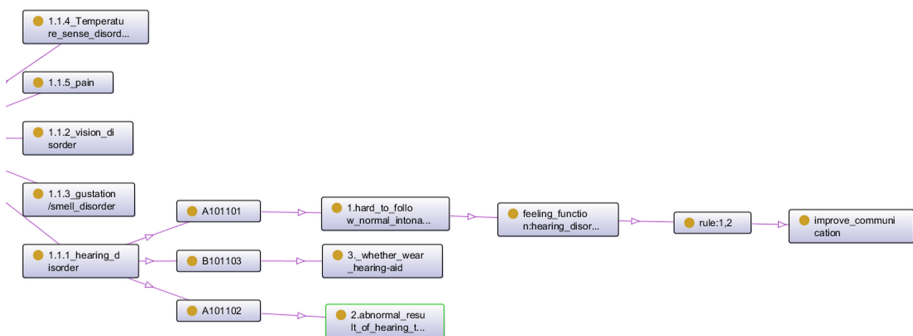


Fig. 5. Nursing knowledge network

The purpose of our core network is to provide corresponding nursing measures based on symptoms. When we decide which measure should be taken in order to address the symptom, we can find the rule node and continue to find the symptom node, which the edge the rule node points to. If the patient has particular symptoms, then we can find the successor of that rule node, which is the nursing measure.

5 Conclusions

In this paper, we propose the establishment of knowledge graphs to address eldercare quality problems, which could be used to generate eldercare procedures to guide the process of eldercare services. To guarantee the reliability of our knowledge graphs, international guidelines are adopted to guide graph construction. To the best of our knowledge, this is the first eldercare knowledge graph to be developed, and it takes the form of an entity graph. Protégé is used to construct the entity graph that consists of eight types of entities and four types of relationships. In this graph, a standard eldercare procedure is represented as a path between an entity denoting an eldercare service classification and an entity denoting an eldercare activity. Approximately 230 procedures could be derived according to the guidelines.

Acknowledgements. This work was supported partially by the National Natural Science Foundation of China (No. 71661167004), National University Training Programs for Innovation and Entrepreneurship (No. 201710359069), and the Programme of Introducing Talents of Discipline to Universities (No. B14025).

References

1. Lu, C.H., Hsiao, J.L., Chen, R.F.: Factors determining nurse acceptance of hospital information systems. *Comput. Inform. Nurs. Cin.* **30**(5), 257–264 (2012)
2. Ammenwerth, E., Rauchegger, F., Ehlers, F., Hirsch, B., Schaubmayr, C.: Effect of a nursing information system on the quality of information processing in nursing: an evaluation study using the HIS-monitor instrument. *Int. J. Med. Inform.* **80**(1), 25–38 (2011)
3. Goodwin, T., Harabagiu, S.M.: Automatic generation of a qualified medical knowledge graph and its usage for retrieving patient cohorts from electronic medical records. In: 7th IEEE International Conference on Semantic Computing, Irvine, CA, USA, pp. 363–370. IEEE (2013)
4. He, J.: Hospital informatization promotes the nursing management. *Mod. Clin. Nurs.* **4** (2004)
5. Yu, T., Liu, J., Jia, L., Gao, B., Li, J.: Construction and applications of a large-scale knowledge graph for traditional Chinese medicine (2015)
6. Li, X., Liu, G., Ling, A., Zhan, J., An, N., Li, L., Sha, Y.: Building a practical ontology for emergency response systems. In: 2008 International Conference on Computer Science and Software Engineering, Wuhan, China, pp. 222–225. IEEE (2008)
7. Liu, Z.Z., Chu, D.H., Jia, Z.P., Shen, J.Q., Wang, L.: Two-stage approach for reliable dynamic Web service composition. *Knowl. Based Syst.* **97**(C), 123–143 (2016)
8. Liu, Z.Z., Chu, D.H., Song, C., Xue, X., Lu, B.Y.: Social learning optimization (SLO) algorithm paradigm and its application in QoS-aware cloud service composition. *Inf. Sci. Int. J.* **326**(C), 315–333 (2016)
9. Blok, H.E.: Knowledge Graph Framing and Exploitation. Master’s thesis, University of Twente, Faculty of Mathematical Sciences, Enschede, The Netherlands (1997)
10. Hulliyah, K., Kusuma, H.T.: Application of knowledge graph for making text summarization (Analizing a text of educational issues). In: 5th International Conference on Information and Communication Technology for the Muslim World, Rabat, Morocco, pp. E79–E83. IEEE (2010)

11. Paparrizos, J., White, R.W., Horvitz, E.: Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results. *J. Oncol. Pract.* **12**(8), 737–744 (2016)
12. Tang, H., Ng, J.H.K.: Googling for a diagnosis—use of Google as a diagnostic aid: internet based study. *BMJ* **333**(7579), 1143–1145 (2006)
13. White, R.W., Horvitz, E.: Cyberchondria: studies of the escalation of medical concerns in Web search. *ACM Trans. Inf. Syst.* **27**(4), 1–37 (2009)
14. Groopman, J.E.: How doctors think. *JAMA* **298**(6), 685–690 (2007)
15. Barnett, G.O., Cimino, J.J., Hupp, J.A., Hoffer, E.P.: DXplain. An evolving diagnostic decision-support system. *JAMA* **258**(1), 67 (1987)
16. Bizer, C., Tom, H., Berners-Lee, T.: Linked data: the story so far. *Int. J. Semant. Web Inf. Syst.* **5**(3), 1–22 (2009)
17. Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., Tang, A., Gabriel, G., Ly, C., Adamjee, S., Dame, Z.T., Han, B., Zhou, Y., Wishart, D.S.: DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**(D1), D1091–D1097 (2013)
18. Schriml, L.M., Arze, C., Nadendla, S., Chang, Y.W., Mazaitis, M., Felix, V., Feng, G., Kibbe, W.A.: Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res.* **40**(D1), D940–D946 (2011)
19. Dumontier, M., Callahan, A., Cruz-Toledo, J., Ansell, P., Emonet, V., Belleau, F., Droit, A.: Bio2RDF release 3: a larger connected network of linked data for the life sciences. In: *CEUR Workshop Proceedings*, pp. 401–404. Springer (2014)
20. Chen, B., Dong, X., Jiao, D., Wang, H., Zhu, Q., Ding, Y., Wild, D.J.: Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC Bioinform.* **11**(1), 255 (2010)
21. Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S., Sontag, D.: Learning a health knowledge graph from electronic medical records. *Sci. Rep.* **7**(1), 5994 (2017)
22. Shi, L., Li, S., Yang, X., Qi, J., Pan, G., Zhou, B.: Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services. *Biomed. Res. Int.* **2017**(4), 1–12 (2017)
23. An, N., Jiang, L., Wang, J., Luo, P., Wang, M., Li, B.N.: Toward detection of aliases without string similarity. *Inf. Sci.* **261**, 89–100 (2014)