International Journal of Engineering & Technology, 7 (4) (2018) 3946-3950

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet. v7i4.21407
Research paper

Identity evaluation based entity cleansing for entities searched from linked open data cloud

Yonglak Sohn *

Department of Computer Engineering, Seokyeong University, Seoul, 02173, Korea
**Corresponding author E-mail: syl@skuniv.ac.kr*

Abstract

Linked open data (LOD) cloud is composed of LODs that assert facts on an entity with various viewpoints. Knowledge expansion, hence, has been an important goal of LOD cloud and achieved by identity links, specified with <owl: sameAs> predicates, among entities in different LODs. After searching the LODs in depth through the identity links, an entity searched from surface LOD would be expanded with various facts obtained from the other LODs. This paper suggests how to evaluate the searched entities as identical to the entity of the surface LOD and then to pick out the entities whose identity levels were sufficiently high compared to the criteria specified in a user query. For entity identity evaluation, LODs' reputations and agreements on the identity assertions have been considered. Identity evaluation based enti-ty cleansing (IE2C) system and its surroundings have been implemented for experiments. Analysis on the experimental results presented that six or seven identity links would be necessary to an entity in order to achieve the goal of knowledge expansion. IE2C would provide in-depth searching results which were composed of trustworthy entities and their various descriptions to users.

Keywords: *Semantic Web; Ontology; Knowledge Expansion; Identity Evaluation; Linked Open Data Cloud.*

1. Introduction

Today, World Wide Web is 'Web of Pages'. Links in a page simply lead to next pages for navigation and each page has been the unit of information accessing. Within pages, there are little semantics so computers have been suffered from utilizing the pages in detail semantically. Semantic web, on the other hand, allows computers to access information at an individual entity. Each entity is described with RDF (Resource Description Framework) model compliant triples which are composed of {subject predicate object} [1-3]. For example, a subject entity, 'Alice', is described by RDF triples, {Alice Knows Bob} and {Alice BasedNear Wonderland}. Object entities, 'Bob' and 'Wonderland', may be subject entities in other RDF triples. Each component in RDF triple has its own URI [4]. Exceptionally, object is allowed to be a literal. By virtue of the URIs, semantic web provides links among the entities and is called as 'Web of Data' [4], [5]. Since 2007, W3C has supported LOD cloud that is a practical realization of the semantic web. Up to September 2018, 1,163 LODs are participating in LOD cloud [6]. Each LOD is published to LOD cloud together with ontologies which specify semantic structures of the LOD [7].

One of the major goals of LOD cloud is to provide knowledge expansions to users [1]. Fig. 1 presents an example of this. LOD_A provides personal life information about 'Alice' whose URI has been specified as <http://Personal/Alice>. An RDF triple {<http://Personal/Alice> <owl: sameAs> <http://Stdudent/Alice>} in LOD_A asserts that the 'Alice' entity specified as <http://Personal/Alice> in LOD_A is identical to the other 'Alice' entity <http://Stdudent/Alice> in LOD_B in which the 'Alice' has been described as a student. RDF triple with the predicate <owl: sameAs> is defined as an identity link. Since two 'Alice's in LOD_A and LOD_B have been connected by the identity link, searching results of {<http://Personal/Alice> <foaf: Knows>

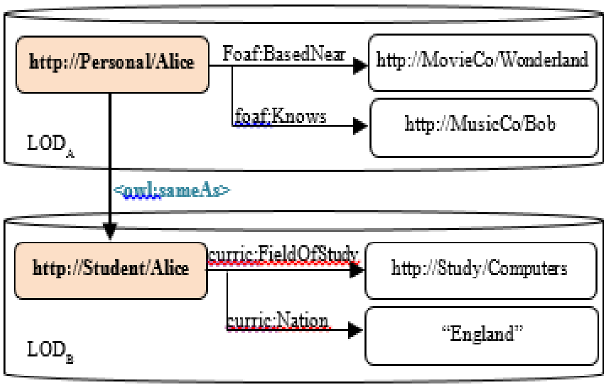

Fig. 1: Example of Knowledge Expansion.

Owl = web ontology language, foaf = ontology describing person, curric = ontology describing educational process

<http://MusicCo/Bob>} and {<http://Personal/Alice> <foaf: BasedNear> <http://MovieCo/Wonderland>} from LOD_A can be expanded with RDF triples of {<http://Personal/Alice> <curric: FieldOfStudy> <http://Study/Computers>} and {<http://Personal/Alice> <curric: Nation> "England"}.

As presented in Fig. 2, entities in LODs would be linked continuously in LOD cloud and knowledge would expand correspondingly. RDF triples of {S_0 P_A O_A}, {S_0 P_B O_B}, {S_0 P_C O_C} from LOD_0 expand with {S_0 P_D O_D}, {S_0 P_E O_E}, {S_0 P_F O_F} from LOD_1 and {S_0 P_G O_G}, {S_0 P_H O_H}, {S_0 P_I O_I} from LOD_2 due to the identity links among LOD_0, LOD_1, and LOD_2.

We must bear in mind, however, that every fact in web is not axiom but just an assertion [3]. Identity links are assertions as well so their trustworthiness needs to be evaluated in order to refine them and then to pick out sufficiently trustworthy entities.

This paper suggests a new methodology for entity identity cleansing. It takes note of LODs' reputations and LODs' agreements on the entities' identities. To verify usefulness of the methodology,

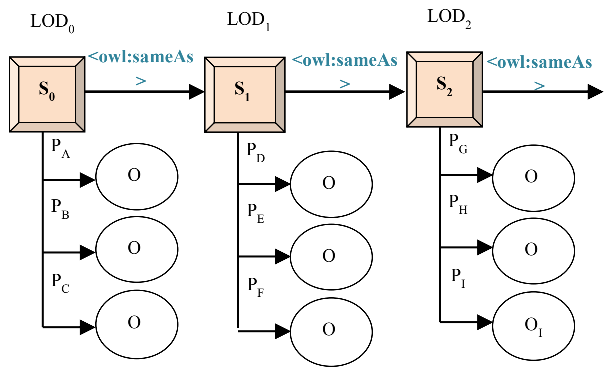

Fig. 2: Continuous Identity links.

S_0 = subject S_0, P_A = predicate P_A, O_A = object O_A

we implemented in-depth searching system that would search and follow LODs in LOD cloud. We also implemented entity identity cleansing system that evaluates the identities of the entities searched at previous stage and then sort out impurities among them. The system is named identity evaluation base entity cleansing, IE²C for short.

2. Related works

Previous works on identity links in LOD cloud concentrated on the creation of the links automatically. [4-9] presented an attention to a predicate of the property <owl: InverseFunctionalProperty>. For example, two triples {A <book: isbn> C} and {B <book: isbn> C}, in which <book: isbn> has been declared to have the property of <owl: InverseFunctionalProperty>, imply that the subject entities, A and B, are identical to each other. The identity link between A and B is very trustworthy due to the property of inverse function. However, the works were confined only to the cases that objects were standard identifiers such as ISBN (International Book Number), GTIN (Global Trade Item Number), ISIN (International Securities Identification Number) and so on.

Other works, SILK[10], LIMES[11], SALE[12], TILE[13], RiiLE[14], focused on the similarity of objects in source and target RDF triples. Common motivation of the works was overcoming the insufficiency of identity links in LOD cloud. Instead of using the identity links which had been explicitly specified with <owl: sameAs> predicate within LODs, they attempted to make identity links automatically. They analysed ontologies of source and target LODs and selected pairs of source and target predicates which were perceived to have the same role in both RDF triples. The predicate pairs would be referenced to find source and target RDF triples which would be candidates of being merged together. They evaluated the similarity of objects of the source and target RDF triples. String comparison method suggested by [15] had been applied in common to the similarity evaluations. If the source and target objects were evaluated to be similar sufficiently, their source and target subjects were accounted identical to each other. [10] was the first try to these approaches.

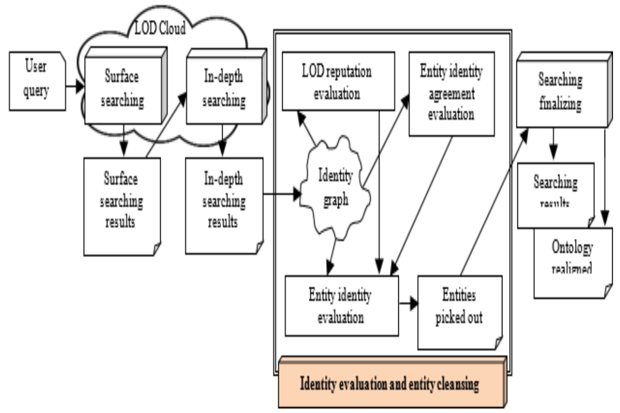

Fig. 3: System Architecture of IE2C and Surroundings.

[11] aimed to reduce the amount of candidate RDF triples proposed by SILK. To improve the degree of the identity, [12] applied syntactic features of RDFS and OWL which were W3C recommended ontology description languages. [13-14] extended SALE by including OWL2. With OWL2, TILE became to consider inferences for selecting candidate RDF triples and then reduced the amount of the triples. From the viewpoint of [16], the works had achieved advances in proving identity links, but insufficiency of trustworthiness of the identity links made automatically was remained.

Trustworthiness of identity links in LOD cloud has not been researched in detail so far. For current WWW, however, Google's page ranking algorithm has provided guidelines on evaluating trustworthiness of searching results [17]. The more references to a page and the higher trustworthiness of the referencing pages, the page being referenced would be decided as more trustworthy one. The page ranking algorithm had been a reference while devising identity evaluation method in this paper.

3. Implementation of identity evaluation based entity cleansing system

We implemented Identity Evaluation based Entity Cleansing (IE²C) system. To supply searching results from LOD cloud to IE²C, we implemented beforehand in-depth searching system as well. Apache Jena 3.1.0 API has been included to process user queries written in SPARQL [18] which is W3C standard query language for accessing RDF triples. Fig. 3 presents architecture of IE²C and surroundings including surface searching, in-depth searching, and searching results finalizing.

3.1. Surface searching and in-depth searching

Searching to LOD cloud begins with a selection of surface LOD that receives user request at the very front. User request is composed of {query, SPARQL endpoint, in-depth level, entity identity level}. Query complies with SPARQL syntax. SPARQL endpoint [11] is URI of a process which receives SPARQL query and returns query results. In-depth level specifies the maximum depth to proceed in LOD cloud. If it were 0, searching would remain in surface LOD. Entity identity level indicates the minimum level of identities of entities, which have been obtained from LODs during in-depth searching, compared to the entity obtained from surface LOD('surface entity' for short hereafter). Entity identity level is between 0.0 and 1.0. An entity whose identity level has been evaluated to be equal or higher than the specified entity identity level, it will join in final searching results. Every surface entity naturally joins in the final results because its identity level has been assigned as 1.0.

Based on each entity in the surface searching results, in-depth searching begins to proceed. From an identity link within the surface searching results, in-depth searching finds a URI of target entity.

With the URI, it composes target LOD's SPARQL endpoint. In-depth searching also organizes SPARQL query to find the target entity, sends the query to the target LOD's SPARQL endpoint, receives query results, and adds them into in-depth searching results. For an identity link in the in-depth searching results, its source entity had been a target entity at the previous searching stage. In-depth searching proceeds to next depth looking for a new target entity in a new target LOD. It proceeds until it meets the depth of in-depth level specified in the user request.

3.2. Evaluations and cleansings

Receiving in-depth searching results, IE²C evaluates identity levels of entities and picks out the entities whose identity levels are equal or higher than the identity level specified before. First, IE²C extracts identity links from in-depth searching results and builds an identity graph for each surface entity. In Fig. 4, an assertion "An entity E_A in LOD_A is identical to an entity E_B in LOD_B" is presented as the identity link (LOD_A, E_A) → (LOD_B, E_B) in bold arrow.
Although assertions in LODs are trustworthy by far compared with those in current web of pages, identity levels of entities, which had been obtained while following the identity links, need to be evaluated. For achieving the identity levels of entities, this paper suggests to go through the following evaluation procedures based on the identity graph.

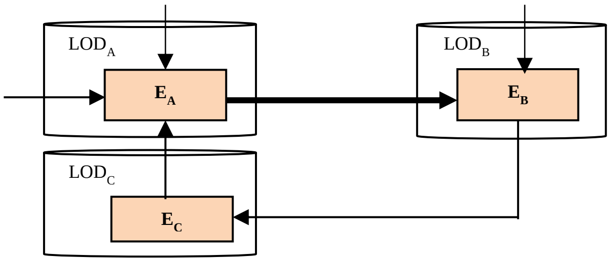

Fig. 4: Example of Identity Graph.

E_A = Entity E_A

3.2.1. LOD reputation evaluation

Predicate and object pairs in RDF triple are assertions about their subject entity based on the viewpoint of an LOD which have opened the RDF triples to the public. Trustworthiness of the assertions, thereby, depends a lot on a reputation of the LOD. For LOD_i, its reputation is denoted by Rep(LOD). Rule 1 suggests basic rules for how to reflect the quantity of identity links coming into LOD_i to Rep (LOD_i).
Rule 1: Identity link reflection

1) Rep (LOD_i) increases as the quantity of identity links coming into LOD_i increases.
2) Rep (LOD_i) increases as Rep(LOD_j), from which the identity links coming into LOD_i had gone out, is higher. ■

From an identity graphs, we can get the coming into and outgoing identity links. Based on the quantity of identity links, reputation of an LOD is calculated as equation (1).

$$Rep\ (LOD_{cur}) = ((Link_{cur} - Link_{min}) / (Link_{max} - Link_{min})) * 0.3 + 0.7 \quad (1)$$

- $Link_{max}$: Quantity of identity links coming into an LOD whose identity links' quantity is the maximum among LODs in the identity graph.
- $Link_{min}$: Quantity of identity links coming into an LOD whose identity links' quantity is the minimum among LODs in the identity graph.
- $Link_{cur}$: Quantity of identity links coming into an LOD whose reputation is currently evaluated.

From (1), every LOD's reputation will be evaluated as between 0.7 and 1.0. If the reputation were evaluated as less than 0.7, identity levels of entities obtained during in-depth searching would be evaluated as low excessively. Realistically, LODs' publishers have sufficient public confidence and thus regarding LODs' reputations as over 0.7 will be adequate supposition. Reputation of surface LOD, LOD_0 hereafter, has been regarded as 1.0 naturally because the user would have accessed LOD_0 based on his confidence. In Fig. 4, if LOD_A were surface LOD, according to (1), Rep(LOD_A), Rep(LOD_B), Rep(LOD_C) wound be 1.0, 0.85, and 0.7 respectively.

3.2.2. Entity identity level evaluation

Identity level of an entity E_j, denoted as Id(E_j), means the extent of identity on the basis of surface entity E_0 which has been a start point of the identity graph being considered. For (LOD_i, E_i) → (LOD_j, E_j), equation (2) is applied to get Id(E_j). Id(E_j) possesses two aspects, one is the identity level of immediately previous entity E_i and the other is the reputation of LOD_i since LOD_i has asserted {E_i <owl:sameAs> E_j}.

$$Id\ (E_j) = Rep\ (LOD_i) \times Id\ (E_i) \quad (2)$$

Likewise Rep(LOD_0), Id(E_0) is regarded as 1.0. By applying (2) on entities in an identity graph, identity levels propagate through identity links. During the propagations, two or more identity links may come into the same entity. IE²C selects one among them in accordance with Rule 2.
Rule 2: Identity Link Selection
For two or more identity links coming into an entity E, to evaluate E's identity level, IE²C selects a link whose identity level is the largest. ■

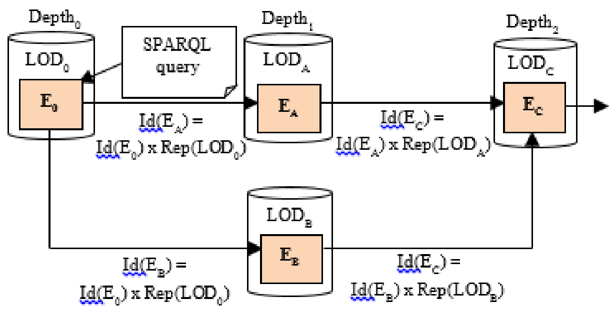

Fig. 5: Propagation and Selection of Entity Identity Levels

In Fig. 5, Rep(LOD_0) and Id(E_0) are 1.0. If Rep(LOD_A) and Rep(LOD_B) were evaluated as 0.9 and 0.8 respectively, with Rule 2, to evaluate Id(E_C), IE²C would select (LOD_A, E_A) → (LOD_C, E_C) instead of (LOD_B, E_B) → (LOD_C, E_C) and then gets it as 0.9.

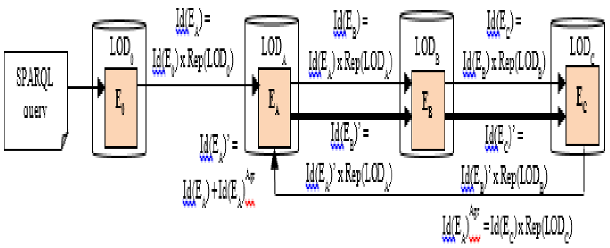

Fig. 6: Example of Entity Identity Agreements.

3.2.3. Entity identity agreements evaluation

As depicted in Fig. 6, identity agreement occurs when identity link leads a cycle in identity graph. Continuous identity links, (LOD_0, E_0) → (LOD_A, E_A) → (LOD_B, E_B) → (LOD_C, E_C), becomes an identity agreement cycle because of (LOD_C, E_C) → (LOD_A, E_A). There already exists (LOD_0, E_0) → (LOD_A, E_A) so that (LOD_C, E_C) → (LOD_A, E_A) becomes an agreement on that E_A is identical to E_0. In Fig. 6, the identity agreement on E_A from E_C is denoted by $Id(E_A)^{Agr}$. If there were an agreement on a fact, naturally, trustworthiness of the fact would increase. $Id(E_A)^{Agr}$, therefore, needs to be a supplement to Id(E_A). Accordingly, IE²C updates Id(E_A) as Id(E_A)′, that is Id(E_A) + $Id(E_A)^{Agr}$. If there were two or more identity agreements, instead of reflecting them all, IE²C would select the largest identity level as the supplement. If the supplemented identity level excessed 1.0, IE²C would set it to 1.0.

After processing the evaluations, entities in an identity graph have been assigned various entity identity levels. For cleansing the entities, IE²C inspects the identity graph closely. It then picks out entities whose identity levels are equal or higher than the identity level that has been specified in the user request. The entities picked out finally are the ones cleansed. IE²C hands them over to searching finalizing stage.

3.3. Searching finalizing

For achieving the knowledge expansion, all pairs of predicates and objects, whose subjects have been evaluated to be identical to the surface subject entity E_0, need to be consolidated. E_0 becomes their representative subject. In addition, the searching results need their own semantic structures. Searching finalizing stage in Fig. 3 collects ontologies from LODs which had provided the entities cleansed by IE²C and their RDF triples. It then builds new ontology specially realigned to the final searching results. The consolidated searching results and their new ontology are submitted as query results to the user. Query results become a small scale LOD which is compliant to RDF model still.

4. Experiments and analysis

Two kinds of data sets were considered for experiments. One is virtual data sets. It is composed of 100 LODs with 1000 entities, hence 100000 entities have participated. The other is five real data sets from DBpedia that is one of the most successful LOD project. Among the LODs in DBpedia, we selected LODs of Korea, France, Italy, Spain, and Portugal in which ‘310,811’, ‘1,591,318’, ‘968,794,, ‘1,120,144’, and ‘865,889’ entities have participated respectively up to September 2018.

4.1. Analysis on influences of entity identity agreements

Virtual data set has been applied for this experiment. Fig. 7 and Fig. 8 present the average identity levels of entities obtained during the in-depth searching. Horizontal axis presents numbers of propagation times. For measuring the average identity levels, four LODs were chosen as surface LODs randomly. One entity from each surface LOD was searched and then entity identity level for each number of propagation times for the each surface entity was measured. The numbers of identity links for each entity were 4, 6, 8, 10, and 12.

Fig. 7 presents results in which entity identity agreements have not been considered. Because entities, met in the first half, are irrelevant each other and due to equation (2), entity identity levels have decreased continuously. But for in the middle and in the last half, the levels are almost invariant. Fig. 8, however, presents other results in which entity identity agreements have been considered. Quite unlike Fig. 7, as numbers of propagation times become 5th, 6th, 7th, owing to the entity identity agreements, identity levels of entities have rebounded. As numbers of propagation times become larger, entity identity levels have rebounded earlier. In cases that quantities of identity links are 8, 10, and 12, their entity identity levels have come close to 1.0. Even in case of 6, the entity identity level has approached to 0.9. In case of 4, although the entity identity level has rebounded at 8th and 9th propagations, it remained nearby 0.5. In this experiment, it has been found that entity identity agreements definitely influence entity identity levels and the influences are logically appropriate. It also has been recognized that entities need to have identity links more than six in order to achieve knowledge expansions with sufficient agreements on the entity identity from sufficient LODs.

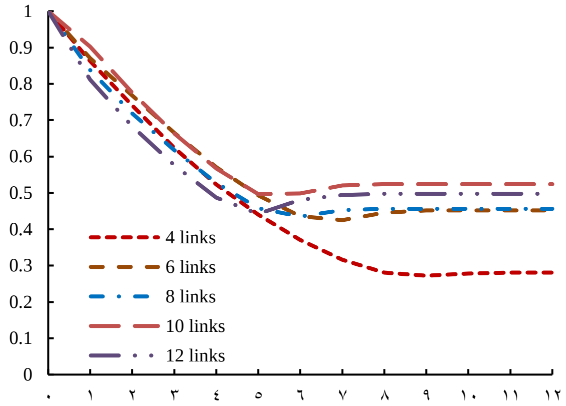

Fig. 7: Average Identity Levels of Entities.

Without entity identity agreements
Horizontal axis: numbers of propagation times

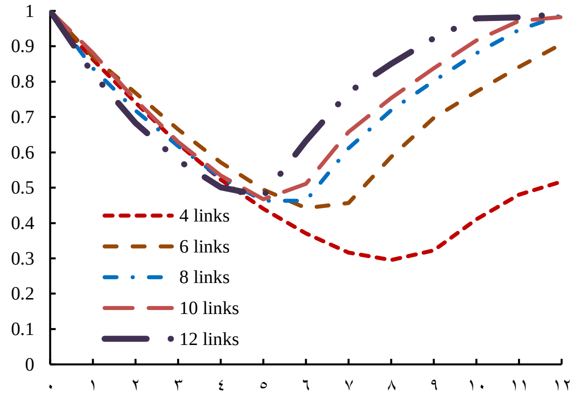

Fig. 8: Average Identity Levels of Entities with Entity Identity Agreements.

Horizontal axis: numbers of propagation times

4.2. Analysis on quantities of identity links

Experiments for this analysis have been carried out on real data of DBpedia LODs. LOD of Korea has been selected as surface LOD. To the LOD of Korea, ‘Iron Man 3’ has been requested to be searched. RDF triples of entities such as ‘fr:Iron_Man_3’, ‘it:Iron_Man_3’, ‘es:Iron_Man_3’, ‘pt:Homem_de_Ferro_3’ have been searched from LODs of France, Italy, Spain, and Portugal.

Fig. 9 presents average identity levels of entities from those LODs searched in depth. They are presented for each quantity of identity links. Compared to conventional LODs in LOD cloud, identity links in DBpedia LODs are very dense. Identity graphs generated from the five LODs have been almost fully connected. On average, entities in DBpedia LODs have 12 identity links. To get the generality, quantities of identity links applied to the experiment have been reduced intentionally. Therefore, 100%, 80%, 60%, 40%, 20% in horizontal axis in Fig. 9 and Fig. 10 correspond with 12, 9.6, 7.2, 4.8, and 2.4 identity links respectively. Entity identity levels in Fig. 9’s vertical axis are averages from the values that had been calculated against all propagations until arrived at the final propagation. As expected, entity identity levels had decreased according as the quantities of identity links decreased.

Fig. 10 presents numbers of propagation times observed from the aspects of identity links’ quantities. At 100% of identity links, propagations reached 59 times and identity levels of the all entities reached 1.0 at the final propagation. Numbers of propagation times decreased steeply from 100% to 60% in Fig. 10. Entity identity levels, however, declined gradually from 100% to 60% in Fig. 9. This means that 100% and 80% of identity links’ quantities are excessive to usual entities. From Fig. 10, identity levels of all the entities at 40% were founded to be below 1.0 notwithstanding their gradual declinations. For 20%, propagations remained 4 times and all entities were below 1.0 naturally. Consequently, according to the experiments of 4.1 and 4.2, for successful entity identity level evaluations with adequate entity identity agreements, it is recommended that every entity needs to have 6 or 7 identity links. 8 or more identity links are excessive. 5 or less identity links, on the other hand, are deficient. With deficiency in identity links, it will be difficult to achieve the goal of knowledge expansions in LOD cloud.

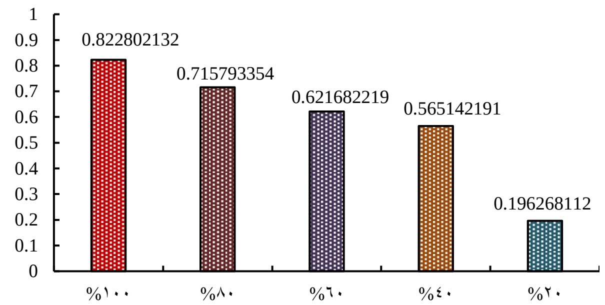

Fig. 9: Average Identity Levels of Entities.

Horizontal axis: 100%, 80%, 60%, 40%, 20% correspond with 12, 9.6, 7.2, 4.8, 2.4 identity links

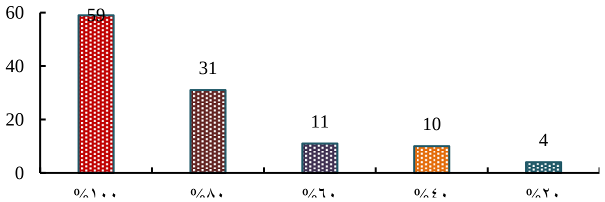

Fig. 10: Numbers of Propagation Times.

Horizontal axis: 100%, 80%, 60%, 40%, 20% correspond with 12, 9.6, 7.2, 4.8, 2.4 identity links

5. Conclusions and further works

This paper implemented identity evaluation based entity cleansing system for LOD cloud. IE^2C receives entities searched from one or more LODs and cleans them to pick out entities that are identical sufficiently to the entity of surface LOD. Minimum limit on the identity level of entities has been specified in the user request. Identity between an entity of surface LOD and entities of other LODs accomplishes knowledge expansion that is an important goal of LOD cloud. For evaluating entity identity level, IE^2C considers LOD's reputation which has been evaluated higher according as the quantity of identity links coming into the LOD is larger. IE^2C combines it with identity level of an entity that has preceded immediately the entity being considered. With these features, entity identity levels propagate as searching proceeds in depth in LOD cloud. The propagations constitute an identity graph. Entity identity agreement would occur if there were a cycle in the identity graph. Identity levels of entities, appeared in the entity identity agreement cycle, are supplemented in order to reflect the agreements. After evaluations, entities whose identity levels are equal or higher than the identity level specified in the user request are picked out as finally refined searching results. Experiments on IE^2C showed that reflecting entity identity agreements to the evaluations of entity identity levels was definitely effective. They also recommended that entities in LOD cloud should be assigned 6 or 7 identity links in order to evaluate the entities' identities successfully with adequate entity identity agreements from LODs.

One contribution of IE^2C is the first attempt to consider LOD's reputation for evaluating trustworthiness of identity links asserted by publisher of the LOD. The reputation has been taken based on the amount of identity links incoming to the LOD. Reflecting identity agreement is another first attempt to evaluate identity levels and its effectiveness has been verified. IE^2C's cleansing method will contribute to improve the trustworthiness of identity links in LOD cloud.

Official report on LOD cloud [6] points out that 44% of LODs in LOD cloud, currently, have remained as data silos. To enhance the effectiveness of LOD cloud, thus, the amount of identity links needs to be increased. Widespread of OWL2 will provide an opportunity to utilize its inference feature for automatically producing identity links. Up to now, identity links have been attached to LODs explicitly and thus LOD cloud has not been resilient to modifications in LODs. Linkage policy, which lists the predicates for matching RDF triples in source and target LODs, will be a solution to overcome the lack of resilience. In the further works, the researches will concentrate on how to automatically produce identity links in LOD cloud. The approach suggested by IE^2C will contribute to improve the trustworthiness of those identity links.

Acknowledgement

This research was supported by Seokyeong University in 2017.

References

[1] N. Konstantinou, D.E. Spanos, Materializing the Web of Linked Data, first Ed., Springer, USA, 2015. https://doi.org/10.1007/978-3-319-16074-0.
[2] W3C, "Linked Data", https://www.w3.org/standards/semanticweb/data, Revised June 2018, Accessed September 26, 2018.
[3] G. Antoniou, P. Groth, a Semantic Web Primer, third Ed., MIT Press, USA, 2012.
[4] T. Heath, C. Bizer, Linked Data: Evolving the Web into a Global Data Space, Morgan & Claypool, USA, 2011.
[5] T.B. Lee, "Semantic Web Road Map", https://www.w3.org/DesignIssues/Semantic.html, Revised October 1998, Accessed September 26, 2018.
[6] A. Abele and J. McCrae, "The Linked Open Data cloud diagram", 2017, http://lod-cloud.net/, Revised August 2018, Accessed September 26, 2018.
[7] D. Allemang, J, Hendler, Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL, Elsevier, USA, 2011.
[8] J.Z. Pan, G. Vetere, J.M.G. Perez, H. Wu, Exploiting Linked Data and Knowledge Graphs in Large Organizations, Springer, Switzerland, 2017. https://doi.org/10.1007/978-3-319-45654-6.
[9] A. Harsh, K. Hose, R. Schenkel, Linked Data Management, 1st Ed., CRC Press, NewYork, 2014.
[10] J. Volz, C. Bizer, M. Gaedke, G. Kobilrov, "Silk – A Link Discovery Framework for the Web of Data", Proc. of the second Workshop on Linked Data on the Web, pp. 238-247, 2009. http://www.researchgate.net/publication/228638267_Silk-A_Link_Discovery_Framework_for_the_Web_of_Data
[11] A.N. Ngomo, S. Auer, "LIMES - A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data", Proc. of the 22nd IJCAIpp. pp. 2312-2317, 2011. http://svn.aksw.org/papers/2011/WWW_LIMES/public.pdf
[12] J. Park and Y. Sohn, "A Syntax Added Link Evaluation Technique for Improving Trustworthiness of LOD's Linkages", Journal of KIISE: Databases, 41:1, pp. 45-61, 2015. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE02360287
[13] J. Park and Y. Sohn, "Trustworthiness Improving Link Evaluation Technique for LOD Linkages giving Considerations to the Syntactic Properties of RDFS, OWL, and OWL2", Journal of KIISE: Databases, 41:4, pp. 226-241, 2014. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE02457716
[14] Y. Sohn, "Reliability Improving Identity Link Evaluation Technique for Linked Open Data Publication", INFORMATION, 19:9, pp. 4271-4279, 2016.
[15] P.F. Brown, P.V. deSouza, R.L. Mercer, "Class-based n-gram Model for Natural Language", Computational Linguistics, 18:3, pp. 467-479, 1992.
[16] C. Bizer, "Is the Semantic Web what we expected", https://www.slideshare.net/bizer/is-the-semantic-web-what-we-expected-adoption-patterns-and-contentdriven-challenges-iswc-2016-keynote, Revised November 2016, Accessed September 26, 2018.
[17] S. Brin, R. Motwani, T. Winograd., "The PageRank Citation Ranking-Bringing Order to the Web", http:// ilpubs.stanford.edu/422/1/1999-66.pdf. Revised November 1998, Accessed September 26, 2018.
[18] B. Ducharme, Learning SPARQL, O'REILLY, USA, 2013.