

FLOWER VOICE: VIRTUAL ASSISTANT FOR OPEN DATA

Takahiro Kawamura^{1,2} and Akihiko Ohsuga²

¹Corporate Research & Development Center, Toshiba Corp.

²Graduate School of Information Systems, University of Electro-Communications, Japan

ABSTRACT

Open Data is now collecting attention for innovative service creation, mainly in the area of government, bioscience, and smart X project. However, to promote its application more for consumer services, a search engine for Open Data to know what kind of data are there would be of help. This paper presents a voice assistant which uses Open Data as its knowledge source. It is featured by improvement of accuracy according to the user feedbacks, and acquisition of unregistered data by the user participation. We also show an application to support for a field-work and confirm its effectiveness.

KEYWORDS

Linked Open Data, Plant, Field, Virtual Assistant

1. INTRODUCTION

Open Data is now collecting attention for creation of innovative service business, mainly in the area of government, bioscience, and smart X projects. However, to promote its application more for consumer services, a search engine for Open Data to know what kind of data are there would be of help. Given that the format of Open Data is RDF (as of Dec. 2012, RDF is a candidate of standard formats for Open Data in Japanese Ministry of Internal Affairs and Communications), it would be hard for the ordinary users to use SPARQL, and full-text search is not good for data fragments. Therefore we propose a simple search mechanism which matches triples extracted from the users' query sentences to the triples $\langle S, V, O \rangle$ in RDF. At the same time, it also becomes a registration mechanism of the user-generated triples. For example, if all the tweets are converted to Open Data, it would be useful for mining and linking with the existing data. To this end, this paper presents a voice assistant based on the user-generated Open Data.

Recently, Siri and other are also drawing attention as the voice assistance, in which application call and information search are commonly used [1]. However, the application call will be eventually embedded in OS. On the other hand, a problem of the information search, especially in smartphone, is in SERP (Search Engine Results Page). It is troublesome to find the necessary data from a list of URLs at the smartphone's screen, and also in some case it needs another search within the page. In this regard, Open Data in RDF, that is, Linked Open Data (LOD) is a promising information source for the voice assistant, because it can pinpoint the necessary data, neither the URL nor the whole page (in fact, the famous question answering system, IBM Watson is also using Linked Data in part). Therefore, combination of Open Data and the voice assistant has potential from both sides.

This paper is organized as follows. Section 2 describes problems and approaches to realize the voice assistant using Open Data. Then section 3 proposes its application, Flower Voice which is an information search and a logging tool for the agricultural work using the smartphone. Finally,

we show the related works from several standpoints in section 4, and summarize with the future work in section 5.

2. PROBLEMS AND APPROACHES

According to classification of interactive voice systems, our voice assistant using Open Data is in the same category as Siri, which is a DB-search question answering (QA) system. But, precisely Siri is a combination of closed DB and open Web-search QA system, but our system is positioned as open (LOD) DB-search QA system. Although the detailed architecture is described in the next section, it basically extracts a triple such as subject, verb and object from a query sentence by using morphological analysis and dependency parsing, and then replaces any of WH-phrases with a variable and searches on the LOD DB. In other word, it is an unification of $\langle S, V, O \rangle$ in the LOD DB and $\langle ?, V, O \rangle$, $\langle S, ?, O \rangle$, $\langle S, V, ? \rangle$ of the query. SPARQL is based on graph pattern matchings, and this way corresponds to a basic graph pattern (one triple matching). The order of matching is that firstly S is matched to a Resource with tracing the links of 'sameAs' and 'wikiPageRedirects', then it searches the matching with V and a Property of the Resource. Fig. 1 shows a conversion from a dependency tree to a triple. Also, at the data registration if there is a Resource corresponding to S and a Property corresponding to V of the user statement, a triple, which has O of the user statement as a Value, is added to the DB.

The DB-search QA system without dialog control has a long history, but apart from lots of other systems, there exists at least the following two problems due to the fact that the DB is open.

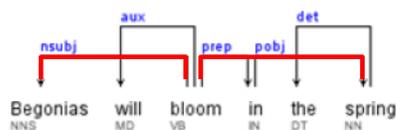
2.1. Mapping of query sentence to LOD schema

The mapping between a verb in the query sentence (in Japanese) and a Property in the LOD schema of the DB is necessary, but in this Open Data scenario both of them are unknown (in case of the closed DB, the schema is given), so that the score according to the mapping degree can not be predefined.

1. Original sentence:

“Begonias will bloom in the spring”

2. Dependency parse



3. Extracted triple:

\langle Subject, Verb, Object \rangle

\langle Begonia, bloomIn, spring \rangle

Figure 1. Conversion from dependency tree to triple

Thus, we firstly register a certain amount of the mappings {Verb (in Japanese), Property} as seeds to a KVS (Key-Value Store), and then if the verb is unregistered,

- (1) we expand the verb to its synonyms using Japanese WordNet ontology, then calculate their Longest Common Substring (LCS) as similarity with the registered verbs.
- (2) Also, we translate the unregistered verb to English, and calculate their LCS with the registered Properties.

- (3) If we find a Resource corresponding to a subject in the query sentence in the LOD, then calculate the LCSs of the translated verb with all the Properties belonging to the Resource, and create a ranking of the possible mappings according to the LCS values.
- (4) After that, the user feedbacks showing which Property was actually referred are sent to the server, and the corresponding mapping of the unregistered verb and a Property is registered to the KVS.
- (5) Moreover, the registered mapping is not necessarily true, so that we recalculate each confidence value of the mapping based on the number of the feedbacks, and update the ranking of the mapping to improve N-best accuracy (see section 3.4).

2.2. Acquisition and Expansion of LOD data

Even if the DB is open, it's not easy for the ordinary user to register new triples to the DB. Therefore, we provide a way for the easy registration using the same extraction mechanism of a triple from a statement. As an incentive for the data registration, we first put a registrant's Twitter ID as a Creator with that data. We also prepared the LOD schema not just for the qualitative information, but also for the users' lifelog, that is, personal records (described in the next section). With these efforts, we will further promote the user participatory, social approach.

On the other hand, we have been developing a semi-automatic LOD extraction mechanism from web pages for generic and / or specialized information, which uses CRF (Conditional Random Fields) to extract triples from blogs and tweets. [2] shows it has archived a certain degree of extraction accuracy.

3. DEVELOPMENT OF AN APPLICATION

This section shows an application of our voice assistant and its implementation. Applications of the voice assistant or the QA system include IVR (Interactive Voice Response), guiding system for tourist and facility, car navigation system, and characters in games. But those all are basically using the closed DB, and would not be the best for the open DB. In addition, our system is currently not incorporating the dialog control like FST (Finite-State Transducers), so that problem solving tasks such as product support is also difficult. Thus, we here focus on the information search provided in the previous section, and introduce the following applications.

1. General information retrieval
DBpedia already stores more than one billion triples, so most of the information people are browsing in Wikipedia can be retrieved from DBpedia in one shot.
2. Information record and search for field-work
Because it enables the user registration, the information relevant to a specific domain can be recorded and searched e.g. for agricultural and gardening work, maintenance of elevators, factory inspection, camping and climbing, evacuation, and travel.
3. Information storage and mining coupled with Twitter
If we focus on the information sharing, it is possible that when the user tweets with a certain hashtag (#), the tweet is automatically converted to a triple and registered in the LOD DB. Similarly, when the user asks with the hashtag, the answer is mined from the LOD DB, which stores a large amount of the past tweets. It would be useful for recording and sharing of WOMs and the lifelog.

In the following section, we introduce our voice assistant from the second perspective, which is "Flower Voice" to answer a question regarding the agricultural, gardening work like disease and pest, fertilization, maintenance, etc.

3.1. Flower Voice

Urban greening and agriculture have been receiving increased attention owing to the rise of environmental consciousness and growing interest in macrobiotics. However, the cultivation of greenery in an urban restricted space is not necessarily a simple matter. Beginners without the gardening expertise will get questions and troubles in several situations ranging from planting to harvesting. To solve these problems, the user may engage a professional gardening advisor, but this involves cost and may not be readily available in the urban area. Also, the internet search using the smartphone has trouble with keyword inputs and iteration of tap and scroll in the SERP to find an answer. Therefore, we considered that the voice assistant is suitable for the information search in the agricultural, gardening work, and developed Flower Voice, which is an agent service to answer a question during the gardening work on site. In other words, it is an information search and a logging tool via voice on the smartphone for the agricultural, gardening work (voice control would be good for this work because of soiled hands and no eyes and ears around). We expect it lowers the threshold of the plant cultivation for the users without the gardening expertise. Fig. 2 shows the overview of Flower Voice. It automatically classifies the user's speech intention (Question Type) to the following four types (Answer Type is either one of literal, URI, or image).

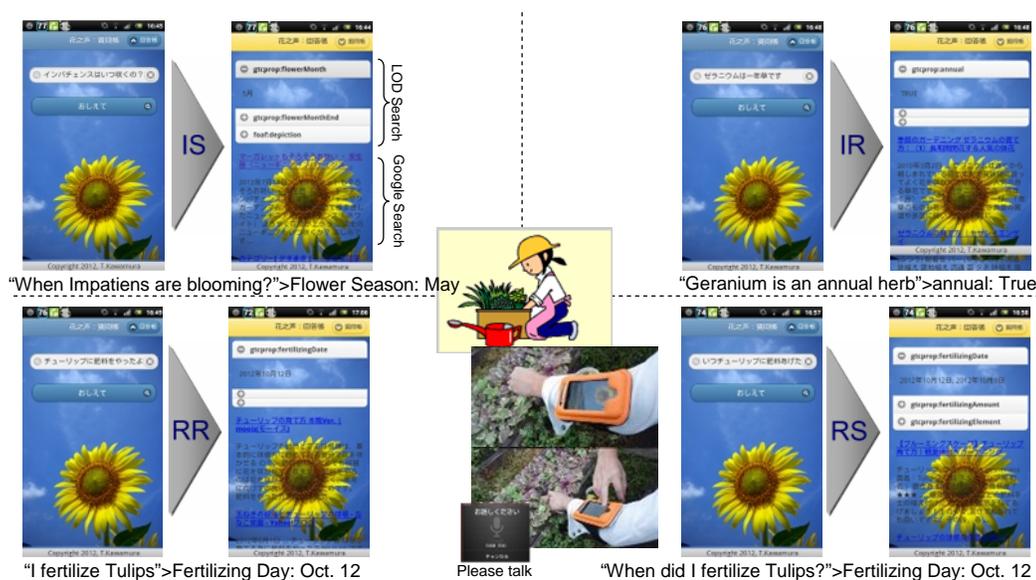


Figure 2. Overview of Flower Voice

1. Information Search (IS)
Search for the plant information prepared in the LOD DB.
2. Information Registration (IR)
Registration of new information for a plant currently not existing in the LOD DB, or additional information to the existing plant.
3. Record Registration (RR)
Record registration and addition for the daily work, and its sharing. According to Japanese Ministry of Agriculture, data logging is the basis of Precision Farming, so the annotation of the sensor information together with the record registration has a high degree of usability. But, the verbs to register are limited to the predefined Properties.
4. Record Search (RS)
Record search to remember the works of the past and refer to ones of other people.

The possible usecases are as follows.

1. Chaining search

This is a case that Flower Voice provides a pinpoint data the user wants to know on site during the work (Fig. 3).

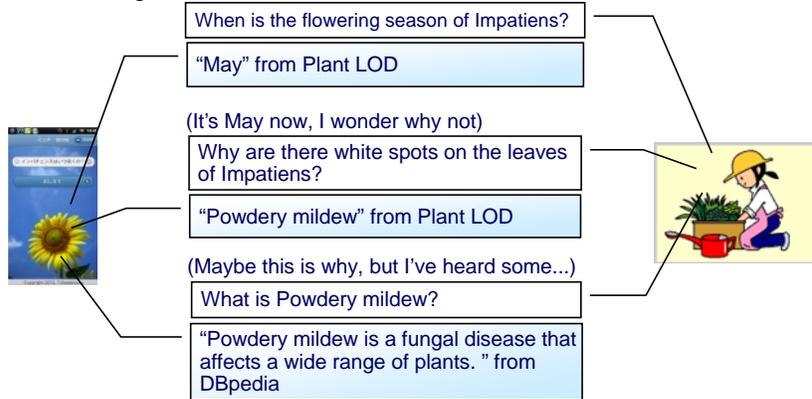


Figure 3. Chaining search

2. Use of user participation

This is a case that the users share a piece of data they learned about by using the registration mechanism. It would be useful, e.g. for ecological survey (Fig. 4 above) and knowledge community (Fig. 4 below). The registrant's Twitter ID is annotated to the data registered.

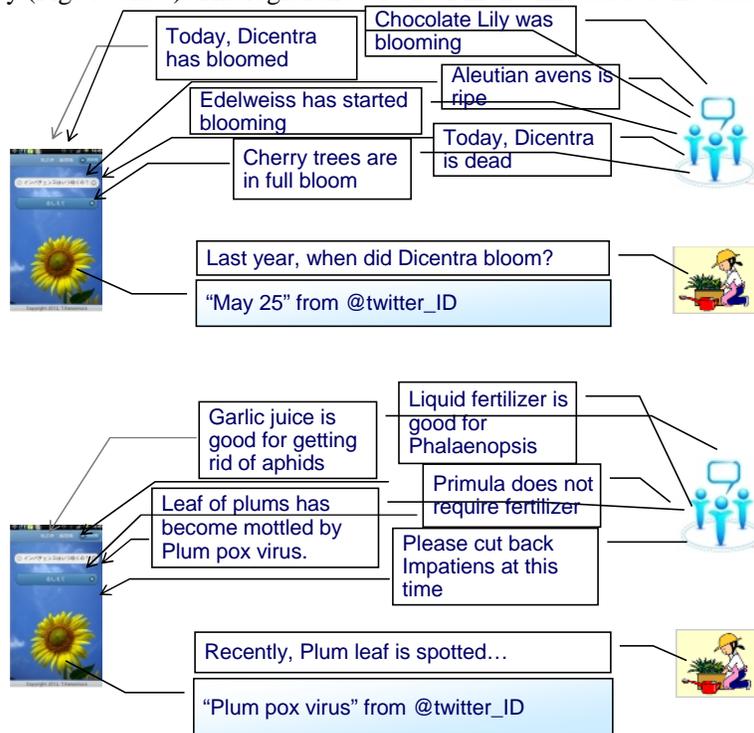


Figure 4. Use for ecological survey (above) and knowledge community (below)

3.2. Plant LOD

The LOD used for Flower Voice is Plant LOD, where 104 Japanese Resources (species) are added to more than 10000 Resources under Plant Class in DBpedia. Also, we added 37 Properties

to the existing 300 Properties from a standpoint of the plant cultivation. In terms of the LOD Schema for the record registration, we prepared the Properties mainly to record dates for flowering, fertilizing and harvesting. Fig. 5 illustrates Plant LOD, which is extended from the one used in Green-Thumb Camera [3] which has been developed for the plant introduction (greening design). Now it is stored and opened in Dydra.com.

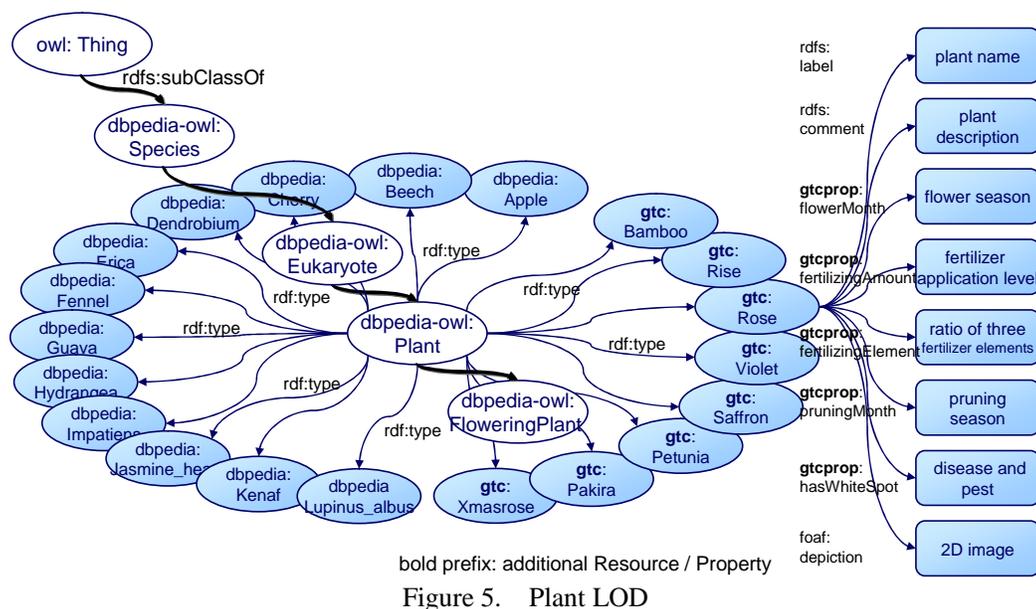


Figure 5. Plant LOD

3.3. System Architecture

Fig. 6 shows our mashup architecture for Flower Voice. The user can input a query sentence by Google voice recognition or a keyboard. Then, it accesses Yahoo! API for Japanese morphological analysis and extracts a triple using a built-in dependency parser. Note that we assumed that the user will not ask a complex question to a machine in the field, so that if a sentence is composed of more than two triples, it must be queried with separated single sentence. Also, the speech intension like search and registration is classified by the existence of question words and usage of postpositional words, not by its intonation. The sentence needs to be literally described whether it's affirmative or interrogative. Then, a SPARQL query is generated by a way of filling in slots in a query template. The search result is received in XML form. Moreover, after searching on {Verb, Property} mappings registered in Google Big Table, and assessing to Microsoft Translator API and Japanese WordNet Ontology provided by National Institute of Information and Communications Technology (NICT), it calculates the LCS value for each mapping by the way described in section 2, and then creates a list of possible answers, which are pairs of Property and Value with the highest LCS value. The number of the answers in the list is set to three due to constraint on a client UI. In the client, it also shows the result of Google search below to clarify advantages and limitations of the QA system by comparison. The user feedback is obtained by opening and closing of an accordion part in the client, which means a detailed look at the Property's Value. At the search time, the feedback increments the confidence value of the registered mapping {Verb, Property}, and register the unregistered mapping. Also, at the registration time, the feedback indicates the Property where O (Value) should be registered. The client UI displays the result and TTS (Text-To-Speech) is not mashed up yet. Currently, the query matches the graph pattern of $\langle S, V, ? \rangle$ alone.

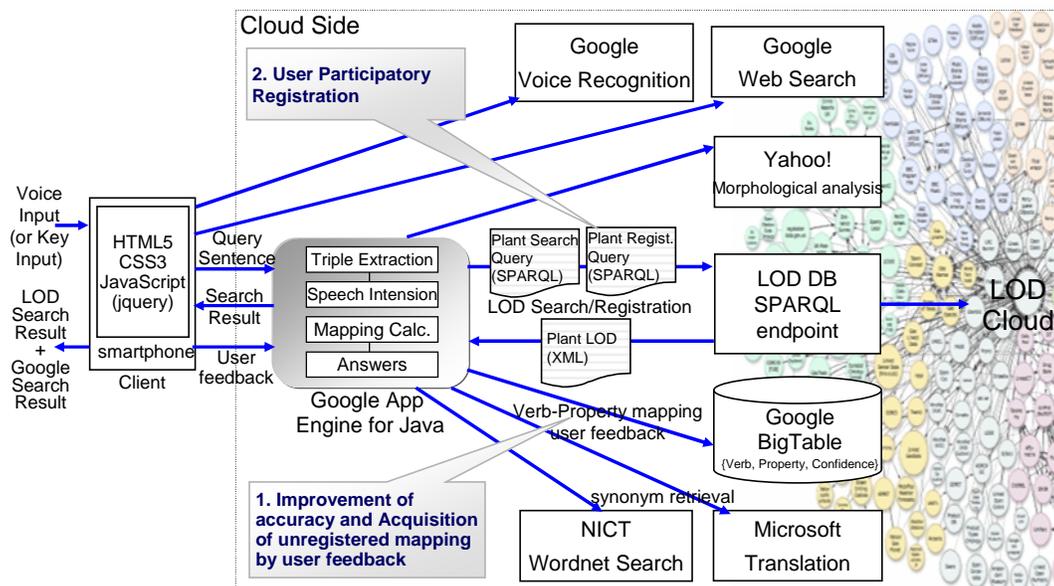


Figure 6. Mashup architecture

We also added an advanced function to change the LOD DB to be searched by the user input of a SPARQL endpoint in an input field of the client UI, though it's limited to the search. Also, every server is not compliant with it, because the query is based on the predefined templates and it receives the result in XML form. Moreover, some servers require attention to its latency. The confirmed endpoint includes Japanese DBpedia (ja.dbpedia.org/sparql), Data City SABAE (lod.ac/sabae/sparql), Yokohama Art LOD (archive.yafjp.org/test/inspection.php), etc. Furthermore, the users can manually register {Verb, Property} mappings. If the Property the user wants does not appear within the three answers, the user can input a {Verb, Property} mapping in the input field. Then, the mapping is registered in the KVS and will be searched at the next query. Either one is targeting for the user who has some expertise to deal with LOD, but we are expecting to find out the exciting usecase by opening it to the user.

Flower voice is available at our site (in Japanese) (www.ohsuga.is.uec.ac.jp/~kawamura/fv.html).

3.4. Evaluation on Mapping Accuracy between Verb and Property

We conducted an experiment to confirm the accuracy of the schema mapping, and its improvement effected by the user feedback mentioned in section 2.1. In the experiment, we first collected 99 query sentences (and their desirable answers) from experienced gardeners. It has no duplication in the sentences, but includes the sentences with the same meaning in semantic level. Then, we randomly made 9 test sets with 11 sentences. We evaluated a test set, and then carried on the next set. Also, we gave the correct feedback, which means the registration of {Verb, Property} mapping and incrementation of the confidence value, to one of the three answers per each query. Here we assume that the query sentence is correctly entered, and do not consider the voice recognition error. After the evaluation of the second test set, we cleared all the effect of the user feedback, and repeated the above from the first set. The result is shown in Table 1.

Table 1. Accuracy of search

	False			True	
	no Res.	no Prop.	triplication error	1-best	3-best
1st Set (ave)	18.2%	0%	9.1%	54.5%	72.7%
2nd Set (ave)	–	–	–	72.7%	72.7%

In this table, “no Res.” means there was no corresponding Resource (a plant) in the Plant DB, and “not Prop.” means no corresponding Property to a verb of the query sentence. Also, “triplication error” means it failed to extract a triple from the query sentence. N-best accuracy is calculated by the following equation;

$$N - best\ precision = \frac{1}{|D_q|} \sum_{1 \leq k \leq N} r_k$$

,where $|D_q|$ is the number of correct answers for question q , and r_k is an indicator function equaling 1 if the item at rank k is correct, zero otherwise. In the case of 3-best, the three answers are compared with the correct answer, and if either one is correct, then it is regarded as correct.

As the result, we found about 20% of the queries was for the unregistered plants, but the prepared Properties covered almost all the queries. Also, the current extraction mechanism is a rule-based, and about 10% of the queries were not analyzed correctly. In this regard, we are planning the extension of the rules and use of the CRF [2]. The N-best accuracy can be increased if we prepare much more data such as Resources and Properties in the Plant LOD and {Verb, Property} mappings, so that the base accuracy for the first set does not have much importance. However, by comparing the result for the first set with the second one, we can confirm the improvement of the accuracy effected by the user feedback (note that 1–best = 3–best means all the correct answers are in the first position).

Also, we expect that the number of the acquired mappings of {Verb, Property} will make a saturation curve according to the number of trials to a domain-dependent value. In this domain, we confirmed it has acquired average 0.09 of the unregistered mappings per a trial (query) initially from 201 mappings in the DB. More detailed analysis will contribute the bootstrap issue at applications to other domains.

On the other hand, we are now considering evaluation on the promotion of the user participatory registration mentioned in section 2.2, including a way of evaluation.

3. RELATED WORK

In the QA system and database researches, there are many attempts on automatic translation from natural language queries to formal language like SQL and SPARQL to help understanding of the ordinary users and even for DB experts. On the other hand, there is also a research to bring back the query and the result to the natural language sentences [4, 5]. Also, we pointed out the difficulty to apply the full-text search to data fragments in section 1, but in this regard there exists a research to convert a keyword list to a logical query [6–8].

In this session, we focus on Linked (Open) Data as a data structure and SPARQL as a query language, and then classify researches on the QA systems which translate the natural sentence to the query into two categories based on whether it needs a deep linguistic analysis or a shallow one.

The one which needs the deep linguistic analysis is ORAKEL [9, 10]. It first translates a natural sentence to a syntax tree using LTAGs (Lexicalized Tree Adjoining Grammars), and then converts it to F-logic or SPARQL. It's able to translate keeping high expressiveness, but instead requires preciseness and regularity to the original sentence. Also, [11] considers a QA system together with design of a target ontology mainly for event information, and is featured by handling of temporality and N-ary at the syntax tree creation. It assigns words of the sentence to slots within a constraint called semantic description defined by the ontology, and finally converts the semantic description to SPARQL, recursively. But, it requires knowledge of the ontology structure in advance.

However, in terms of the voice assistant for the user, these approaches have difficulty with practical use due to voice recognition errors, syntax errors of the original sentence, and the triplification errors. Also, if the DB is open, as mentioned in section 2 the assumption that the ontology schema is already given is at risk. Thus, aiming at portability and schema independent from the DB, there are approaches with the shallow linguistic analysis. Our system in this paper is also classified in this area.

FREyA [12] has been originally developed for the natural language interface for ontology search. It has many similarities with our system like matching of words of the sentence and Resource / Property using a string similarity measure and synonyms of WordNet, and improvement of the accuracy using the user feedback. But, it performs conversion of the sentence to a logical form using ontology-based constraints (without consideration of the syntax of the original sentence unlike the semantic description), assuming completeness of the ontology used in the DB. On the other hand, DEQA [13] takes an approach called TBSL (Template based SPARQL Query Generator) [14]. It prepares templates of SPARQL queries, and converts the sentence to satisfy slots in the template (not the ontology constraint). Then, as well as our system it applied to a specific domain (real estate search), and showed a certain degree of the accuracy. PowerAqua [15–17] is also for the natural language interface for ontology search in origin, and has similarities with our system such as a simple conversion to the basic graph pattern called Query-Triples, the matching of words of the sentence and Resource / Property using a string similarity measure and synonyms of WordNet, and the use of the user feedback. Moreover, against Open Data, PowerAqua introduced heuristics according to the query context to prevent throughput decrease.

[18] serves as a useful reference to a survey of QA system. The proposed system in this paper is referring to a number of works. However, it can be distinguished by a social approach, that is, the improvement of the accuracy and data acquisition in the user participatory manner by seamlessly combining the search query and the registration statement. Also, the application to the field work (and Japanese sentence) has no similar work.

Recently, the well-known voice assistants have been commercialized like Apple Siri, xBrainSoft Angie. Either one has a voice recognition function with high accuracy and good at some typical tasks like call of terminal capabilities and applications installed, which are easily identified from the query. In terms of the information search, it correctly answers the question in case that the information source like Wikipedia, which is presumably pre-defined for each keyword, is a well-structured site. It, however, often fails the information extraction from an unstructured site, and returns the SERP, so that the user needs to tap a URL in the list as mentioned in section 1. Also, Angie provides a link to Facebook and a development kit. By comparison, our system focuses on the information search and takes the LOD DB as the source enabling the user participatory registration, in order to raise the accuracy.

As applications of the smartphone to the agricultural use, Fujitsu Ltd. provides a recording system in which the user can simply register the work types by buttons on the screen with photos of cultivating plants. Also, NEC Corp. is offering a M2M (Machine to Machine) service aiming at visualization of sensor information and support of farming diary. Both are tackling recording and visualization of the work as well as Flower Voice, but our system has a form of a voice-controlled QA system for Open Data with a social approach by combining the data recording and its reference.

3. CONCLUSION AND FUTURE WORK

In this paper, we proposed a voice assistant which uses Open Data as its knowledge source to facilitate the spread of Open Data in the consumer services. Then, we developed an application for the field-work support, and presented an evaluation. It is intended not to jump into the net to search the necessary data, but to pull out the data from the net in the field. Also it is featured by a mechanism of Human Computation, i.e. improvement of the accuracy according to the user feedback, and acquisition of the unregistered data by the user participation.

In the future, we plan to have further evaluation on the problems mentioned in section 2, especially for the promotion of the user participation. Also, we will measure a response to this application, and apply to other domains.

REFERENCES

- [1] Ars Technica: ““Siri, does anyone still use you?” Yes, says survey”, <http://arstechnica.com/apple/2012/03/siri-does-anyone-still-use-you-yes-says-survey/>, 2012.
- [2] T. M. Nguyen, T. Kawamura, Y. Tahara, A. Ohsuga: “Building a Timeline Network for Evacuation in Earthquake Disaster”, Proc. of AAAI 2012 Workshop on Semantic Cities, 2012.
- [3] T. Kawamura, A. Ohsuga: “Toward an ecosystem of LOD in the field: LOD content generation and its consuming service”, Proc. of 11th International Semantic Web Conference (ISWC), 2012.
- [4] A. Simitsis, Y.E. Ioannidis: “DBMSs Should Talk Back Too”, Proc. of 4th biennial Conference on Innovative Data Systems Research (CIDR), 2009.
- [5] B. Ell, D. Vrandecic, E. Simperl: “SPARTIQUULATION: Verbalizing SPARQL Queries”, Proc. of Interacting with Linked Data (ILD), 2012.
- [6] P. Haase, D. Herzig, M. Musen, D.T. Tran: “Semantic Wiki Search”, Proc. of 6th European Semantic Web Conference (ESWC), 2009.
- [7] S. Shekarpour, S. Auer, A.N. Ngonga, D. Gerber, S. Hellmann, C. Stadler: “Keyword-driven SPARQL Query Generation Leveraging Background Knowledge”, Proc. of International Conference on Web Intelligence (WI), 2011.
- [8] D.T. Tran, H. Wang, P. Haase: “Hermes: Data Web search on a pay-as-you-go integration infrastructure”, J. of Web Semantics Vol.7, No.3, 2009.
- [9] P. Cimiano: “ORAKEL: A Natural Language Interface to an F-Logic Knowledge Base”, Proc. of 9th International Conference on Applications of Natural Language to Information Systems (NLDB), 2004.
- [10] P. Cimiano, P. Haase, J. Heizmann, M. Mantel: “Orakel: A portable natural language interface to knowledge bases”, Technical Report, University of Karlsruhe, 2007.
- [11] M. Wendt, M. Gerlach, H. Duewiger: “Linguistic Modeling of Linked Open Data for Question Answering”, Proc. of Interacting with Linked Data (ILD), 2012.
- [12] D. Damjanovic, M. Agatonovic, H. Cunningham: “FREyA: an Interactive Way of Querying Linked Data using Natural Language”, Proc. of 1st Workshop on Question Answering over Linked Data (QALD-1), 2011.

- [13] J. Lehmann, T. Furche, G. Grasso, A.N. Ngomo, C. Schallhart, A. Sellers, C. Unger, L. Buhmann, D. Gerber, K. Hoffner, D. Liu, S. Auer: “DEQA: Deep Web Extraction for Question Answering”, Proc. of 11th International Semantic Web Conference (ISWC), 2012.
- [14] C. Unger, L. Buhmann, J. Lehmann, A.N. Ngomo, D. Gerber, P. Cimiano: “Template-based question answering over RDF data”, Proc. of World Wide Web Conference (WWW), 2012.
- [15] V. Lopez, E. Motta, V. Uren: “PowerAqua: Fishing the Semantic Web”, Proc. of 3rd European Semantic Web Conference (ESWC), 2006.
- [16] V. Lopez, M. Sabou, V. Uren, E. Motta: “Cross-Ontology Question Answering on the Semantic Web - an initial evaluation”, Proc. of 5th International Conference on Knowledge Capture (KCAP), 2009.
- [17] V. Lopez, A. Nikolov, M. Sabou, V. Uren, E. Motta, M. DAquin: “Scaling Up Question-Answering to Linked Data”, Knowledge Engineering and Management by the Masses, LNCS 6317, 2010.
- [18] V. Lopez, V. Uren, M. Sabou, E. Motta: “Is question answering fit for the semantic web?”, A survey. Semantic Web J. No. 2, 2011.

Authors

Takahiro Kawamura is a Senior Research Scientist at the Corporate Research and Development Center, Toshiba Corp., and also an Associate Professor at the Graduate School of Information Systems, the University of Electro-Communications, Japan.



Akihiko Ohsuga is a Professor at the Graduate School of Information Systems, the University of Electro-Communications. He is currently the Chair of the IEEE Computer Society Japan Chapter.

