

My Health Dictionary: Study on Web Service using Program Information Data-hub as Linked Open Data

Masaru Miyazaki¹, Makoto Urakawa¹, Ichiro Yamada¹, Kikuka Miura¹,
Taro Miyazaki¹, Hiroshi Fujisawa¹, and Toshio Nakagawa¹

Japan Broadcasting Corporation(NHK), Tokyo, Japan,
miyazaki.m-fk@nhk.or.jp

Abstract. With the evolution of the global Internet, it has become increasingly common for companies to automatically exchange various types of data among themselves. In addition, content providers, such as broadcasting stations, are being required to change their content-serving strategy so that the content can be delivered to viewers via various external services. To address this strategy change, this paper proposes program-related information as machine-readable web data that can be used in internal and external services. We report on the construction of a *program information data-hub* using the Linked Open Data (LOD) standard format recommended by the World Wide Web Consortium. Results obtained by prototyping the data-hub and associated web services show that services employing a variety of program information can be realized by representing knowledge about the content as LOD data.

Keywords: linked open data, program information, health

1 Introduction

The circulation of large amounts of content on the Internet, the rapid spread of mobile devices such as smartphones, and the appearance of new viewing styles such as time-shift playback have resulted in a significant change in the behavior of viewers of TV programs. In line with this change, content providers such as broadcasting stations are being required to change their content-serving strategy, so that the content can be delivered to viewers via various external services rather than simply waiting for them to access content via a broadcast and video on demand service. We believe that an effective strategy to achieve this objective is to utilize semantic web and Linked Open Data (LOD) technology to build a “Web of Data.” Consequently, we are currently studying how to build a hub comprising various types of data associated with a TV program by creating LOD program-related information and external data. In this demonstration, we give an outline of the *program information data-hub* and introduce the prototype health information services that use the data-hub.

2 Related Works

The BBC recognized the possibility of improving the utilization and presence of content via LOD technologies at an early stage, and they have been working to build a content data space using LOD. They have consequently developed LOD for content such as program episodes and music artist information, and these can be referenced from a variety of external sources. Moreover, at the web site Wildlife¹, animal species and behavior information are systemized as RDF. Thus, a user can immediately enjoy relevant information and programs. Further, their efforts to connect various pieces of program information as LOD by crowdsourcing using tags are highly appreciated in the academic field[1].

In this study, our aim is to actualize a more advanced data space that can be utilized as an internal or external service of broadcasting stations. As a result, we have constructed a *program information data-hub* that has accumulated not only existing program information but also knowledge about the program contents, and are currently exploring its possibilities.

3 Development of *program information data-hub*

NHK, Japan's public broadcast station, has developed a variety of program-related information in an in-house database for the purpose of providing broadcasting information on their web site. However, much of the information was described in RDB tabular form and, as a result, was underutilized for collaboration with other new services. For example, a vegetable that is introduced as a disease prevention measure in a health program could also be introduced as an ingredient for a recipe in a cooking program. In this way, programs often provide information that share common concepts with another program. By connecting such programs to each other via the common concept, the creation of a new service that connects programs transversely is possible. With the aim of realizing such services, we gathered the program-related information (such as location of image, video, web site) residing in in-house databases, transformed information to RDF and automatically constructed an RDF store called a *program information data-hub*. We used the Programmes Ontology of the BBC as a reference in describing the schema of our data-hub, and expanded it to be able to describe NHK's own program information. Next, in order to realize the cooperation between the various external services, we automatically extracted performer names and important words included in the program information, and added the link information to the vocabulary of the DBpedia Japanese². Further as external knowledge, we automatically added a link to a "knowledge map," which is a program-related knowledge dataset that is currently being built.

The knowledge map comprises two types of data. One is "concept map," which consists of data obtained by analyzing a large text corpus on the web. The map shows the semantic relations between words, such as causal relations

¹ <http://www.bbc.co.uk/nature/wildlife>

² <http://ja.dbpedia.org>

and hyponymy relations[2][3]. The second is “content map,” which was generated by analyzing the summary text of a program. This method is composed of two processes: topic extraction and relation estimation based on TF-IDF statistics and semantic relation in concept map. It shows the relation between words and the associated program.

A structural example of a knowledge map associated with the program information is shown in Fig.1. Concept map included 28 types, approximately 1,012 million word relations in total.

Finally, we accumulated all schemas, instances, and knowledge map data in an RDF Store, and constructed an environment that is accessible from a variety of services through the SPARQL endpoints and WebAPI. Currently, *program information data-hub* contains 1.89 million pieces of RDF triple created from the experimental accumulation of 6,700 pieces of program data over a period of two months.

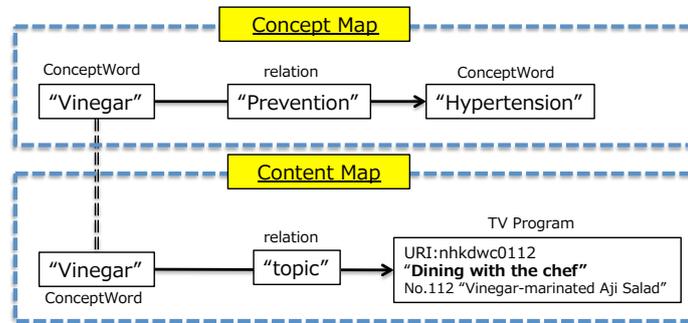


Fig. 1. Structure example of program information and knowledge map

4 *My Health Dictionary* service using *program information data-hub*

We have developed a “*My Health Dictionary*”(Fig.2) as a service that utilizes the *program information data-hub*. *My Health Dictionary* is implemented as an extension of *Google Chrome*. When the user selects a keyword in which he/she is interested while browsing the Web, the program associated with the word is displayed as a popup. Fig.2 shows an example in which the keyword “hypertension” is selected on the text of a Webpage. When the user right-clicks on the selected keyword, a knob describing the keyword string is displayed, and related words linking to the keyword are displayed around the knob. If the user then selects a related word such as “prevention,” a list of programs related to “prevention of hypertension.” is then displayed. By clicking the program, the user can then watch the video of the program or browse the program website. Using the information from the concept maps, the system is also able to list the cooking program that introduced a recipe using “vinegar” which is said to help in the prevention of hypertension. Because the data-hub stores program information about various genres, the cross-sectoral services that connect the various

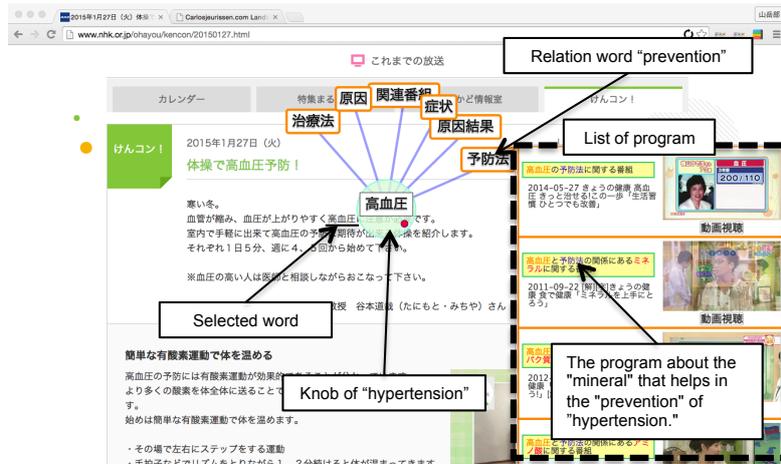


Fig. 2. Screen shot of “My Health Dictionary”

programs can be realized by utilizing program-related knowledge such as from the concept map or the content map.

5 Conclusion and future work

In this paper, we reported on a prototype LOD based *program information data-hub* constructed by linking external knowledge with existing program information. Further, we demonstrated a service example using the data-hub. In future work, we plan to conduct studies on data-hub construction and utilization in areas such as news, culture, and education, and establish an improved and sophisticated data-hub for actual service operation.

References

1. Yves Raimond, Tristan Ferne, Michael Smethurst, Gareth Adams: The BBC World Service Archive prototype, Web Semantics: Science, Services and Agents on the World Wide Web, Vol.27–28, pp.2–9 (2014)
2. Stijn De Saeger, Kentaro Torisawa, Jun'ichi Kazama, Kow Kuroda and Masaki Murata: Large Scale Relation Acquisition using Class Dependent Patterns, In Proceedings of the IEEE International Conference on Data Mining (ICDM'09) , pp.764–769(2009)
3. Stijn De Saeger, Kentaro Torisawa, Masaaki Tsuchida, Jun 'ichi Kazama, Chikara Hashimoto, Ichiro Yamada, Jong-Hoon Oh, Istvan Varga, Yulan Yan, Relation Acquisition using Word Classes and Partial Patterns, In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2011), pp.825-835(2011)