Integrated Growth Media Database by Standardizing Ingredient Information

Shuichi Kawashima[1], Toshiaki Katayama[1], Yuki Moriya[1], Shinobu Oka-moto[1], Yasunori Yamamoto[1], and Susumu Goto[1]

[1] Database Center for Life Science, 178-4-4 Wakashiba, Kashiwa-shi, Chiba 277-0871, Japan

kwsm@dbcls.rois.ac.jp

Abstract. Microbial culture collections around the world provide valuable information resources about microorganisms, but their metadata are independently maintained. To enhance the interoperability of such resources, we have developed RDF datasets from the metadata of two culture collections in Japan. Among the efforts, here, we report the RDFized microbial growth media information and a new database of growth media . Currently, the database includes 1,644 media and available at http://growthmedium.org/.

Keywords: Semantic Web, RDF, Ontology, Microbiology, Growth media.

1   Background

There are hundreds of microbial culture collections maintained by biological resource centers (BRCs) in the world and currently 787 culture collections are listed in the Culture Collections Information Worldwide (CCINFO) provided by World Data Centre for Microorganisms (WDCM). These culture collections are not only essential infrastructure as the basis of carrying out microbial research but also valuable rich information sources such as microbial physiological and phenotypic information, which are described in the accompanying metadata of each microorganism. So far, as our collaborative efforts with BioResource Research Center, Riken and Biological Resource Center, National Institute of Technology and Evaluation (NITE), metadata of their culture collections, henceforth JCM and NBRC, have been converted into RDF datasets. Both datasets are available at the NBDC RDF portal [1].

Each metadata consists of the minimum data set description defined by the World Federation of Culture Collections (WFCC) Global Catalogue of Microorganisms (GCM) [2], which includes one or more links to the Web documents that describe how to prepare growth media for the corresponding microorganism. Along with the development of the RDF above mentioned, we have also independently developed an RDF dataset focusing on a set of ingredients of microbial growth media whose recipes are provided by JCM and NBRC. Based on the RDF datasets, recently, we have developed a new database of microbial growth media [3]. In the report, we present the RDF model of growth media and the current status of the database.

2 Results and Discussion

Initially, we have developed RDF data of growth media provided by the JCM and NBRC, and then we also started to collect media recipes for microorganisms of which genome sequences are available and not included in both JCM and NBRC. As of October 2019, our database contains 1644 growth media (of which 649 are from NBRC, 843 from JCM, 152 by the manual collection from research papers and the Web). The database system was implemented using OpenLink Virtuoso triple store, SPARQList API, TogoStanza, and TogoDB. One of the useful features of the RDF data is that all ingredients are described using the Growth Medium Ontology (GMO). Different BRCs or even single BRC sometimes use different naming for the same ingredients that hinders the integrated use of the growth media data. The new RDF will enable us to carry out integrated analyses such as an alignment of ingredients and design a similarity measure between growth media. Figure 1 shows a TogoStaza application to visualize the alignment of ingredients of growth media used for the specified microorganisms. We have a plan to implement a functionality for retrieving similar media to a user-specified medium.
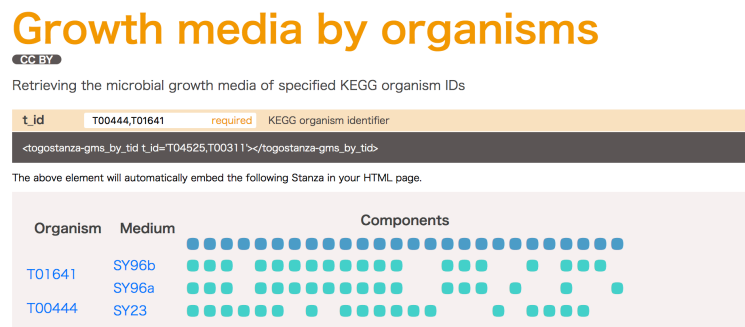

Fig. 1. A TogoStanza application to visualize the alignment of ingredients among growth media. Each colored dot represents an ingredient and linked to the detailed information.

References

1. Kawashima, S., Katayama, T., Hatanaka, H., Kushida, T., Takagi, T.: NBDC RDF portal: a comprehensive repository for semantic data in life sciences. Database (Oxford) doi: 10.1093/database/bay123 (2018).
2. WFCC GCM Minimum Data Sets Description: http://gcm.wfcc.info/datastandards/, last accessed 2019/10/6.
3. DBCLS Growth media database: http://growthmedium.org/, last accessed 2019/10/6.