# Building and Exploring Marine Oriented Knowledge Graph for ZhouShan Library

Tong Ruan[1,2], Haofen Wang[1,2], Fanghuai Hu[2], Jun Ding[2], and Kai Lu[3]

[1] East China University of Science & Technology, Shanghai, 200237, China
{ruantong,whfcarter}@ecust.edu.cn
[2] Shanghai Hi-knowledge Information Technology Corporation
{hufh,dingjun}@hiekn.com
[3] ZhouShan Library, Zhe Jiang Province, China
{zslukai}@126.com

**Paradigm Shift of Library Industry in China** As more and more readers are in favor of accessing digital resources online, most libraries in China are in their way to build or strengthen their digital libraries. Nowadays, there exist several major content providers like WeiPu[4], WanFang[5], and ChaoXing[6] who not only own a large number of digital contents of journals, books, and magazines, but also run their integrated platforms for search and navigation. Most libraries only act as a consumer or a distributor in the digital content supply chain, which makes them suffer from serious homogenization, lack of content control, and weak competitiveness. The above issues enforce libraries to search for new opportunities.

On the other hand, in early 2013, China Ministry of Culture has issued guidelines to build various resource repositories specified for different sectors. It advocated different regions to develop thematic repositories according to the economic and cultural characteristics of the region. ZhouShan Library takes this chance and becomes a pioneer to make the transition. ZhouShan Islands are listed as the first "state-level new district" around marine economy. With the support of local government, ZhouShan Library starts a project named "Universal Knowledge Repository for Marine Digital Library". The intension is to help inhabitants and travelers know ZhouShan and marine economy, and to support different bureaus of ZhouShan government, such as Fishery Agency or Economic and Information Commission to do queries and statistics about local marine economy. In this way, ZhouShan Library is changing from a content distributor to a content provider of the marine domain. This change also happens to other regional libraries, which leads to a trend of paradigm shift in China's library industry.

**The Role of (Vertical) Knowledge Graph** Regarding the ZhouShan Library project, a marine repository should include fishes, fishing grounds, fish processing methods, related researchers and local enterprises. No single source can cover all aspects of data in the repository. It is also impossible for users to manually integrate knowledge from various sources. In some cases, concepts or facts need to be extracted from semi-structured data (e.g., lists or tables from Web pages) and unstructured data (e.g., documents). In other cases, data from internal database or from LOD are to be extracted, transformed, and loaded to the repository in a unified representation. Moreover, research institutes,

[4] http://oldweb.cqvip.com/
[5] http://www.wanfangdata.com.cn/
[6] http://www.chaoxing.com/

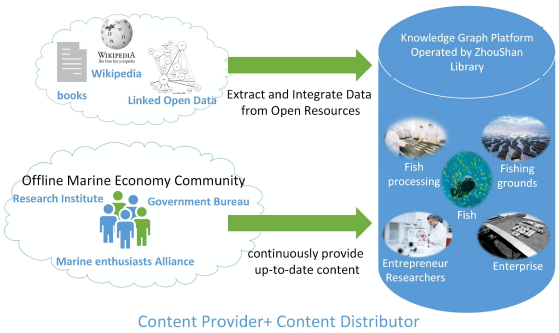


**Fig. 1.** Overview of KG Platform for ZhouShan Library

government bureaus, and marine enthusiasts alliances are allowed to continuously add new knowledge to make the repository up-to-date.

To fulfill the above requirements of repository construction, ZhouShan Library embraces semantic technologies to build a vertical marine oriented knowledge graph (see Figure 1). Knowledge Graph (KG) was first introduced by Google to empower its search. The big success of knowledge graph attracts many attentions from other internet companies as well as traditional industries. The main advantages of semantic technologies include: a) *Incremental schema design and enrichment*. It is difficult to know all concepts during the initial design of KG. Its dynamic extensibility and "schemaless" characteristic enable to add new schemata or revise existing ones later without rebuilding the whole KG from scratch. b) *Easy data integration*. The semantic interoperability of ontologies and the "linked data" principle makes it more efficient to integrate digital contents from different content providers. c) *Existing standards support*. The library can urge the content providers to obey the existing standards like URI, RDF(S), and SPARQL. d) *Expressive semantic search*. Users can ask for entities satisfying semantic constraints when searching on KG. It is more precise than keyword-based retrieval.

**Deployment of KG in ZhouShan Library** We build an integrated tool[7] with three key components namely *knowledge integration* module, *knowledge store* module, and *knowledge access* module. As for knowledge integration, we will describe the technical details of converting relational data from internal databases to RDF triples, facts extraction from user generated contents like Wikipedia, importing marine related ontologies from the Web, and fusion at both schema- and data-level. Moreover, we will introduce the strategies of schema inconsistency and data conflict detection, and the mechanisms for users to extend and validate KG with collaborative editing tools. For the design of knowledge store, we will discuss the choices to select a combination of databases (relational database, NoSQLs, and file systems) for fast access of KG. Regarding knowledge access, we will present different ways including card view, wheel view, and detail view to navigate and browse marine oriented KG. Besides, We will explain how to implement semantic search which supports natural language querying. Finally, we will introduce a list of available Restful APIs for developers to interact with the underlying KG.

[7] You can access the online production test system via `http://202.120.1.49:19155/SSE/`.