2017 IEEE 11th International Conference on Semantic Computing

Public Cultural Knowledge Graph Platform

Yi Yang[1], Guigang Zhang[1], Jian Wang[1], Shufeng Ye[1]
[1] Institute of Automation
Chinese Academy of Sciences
Beijing, China
{yangyi, guigang.zhang, jian.wang, yeshufeng2015}@ia.ac.cn
Corresponding author: Jian Wang

Jianqiao Hu[1,2]
[2] Department of Automation
Beijing Institute of Technology
Beijing, China
songshansls@sina.com

Abstract—**Recently the online public cultural services have grown fast. Massive cultural knowledge can be extracted from the big data generated by the services. The discovery, management, and analysis of the public cultural knowledge become critical tasks for improving the public cultural services. To address the issues, this paper proposes the public cultural knowledge platform that builds and maintains a knowledge graph by exploring the outside data and mining the internal cultural knowledge. The knowledge graph structures the public cultural information as a graph that consists of attributed cultural entities and relations of the entities. The public cultural knowledge can be discovered, fused, and analyzed. This paper also presents the architecture of the public cultural knowledge platform. With the help of the platform, the public cultural applications can provide users rich cultural knowledge instead of simple data. The platform help enhance the intelligent services of the applications as well. This paper shows example scenarios for depicting how the platform facilitates the public cultural recommendation.**

Keywords- knowledge graph; public culture; big data; semantic web; architecture.

I. Introduction

Recent years the online public cultural services have developed rapidly in China. The services involve various fields, e.g., education, cultural resource sharing service, and cultural research. The services have generated massive data, where the public cultural knowledge can be discovered. The public desires to learn not simple data but rich knowledge from the resources provided by the public cultural services, e.g., the relations of roles in a novel, and the history of a famous work of art and the information of the related collection museums. Fragmentation knowledge can be linked according to semantic meanings. From one knowledge point users can discover the linked knowledge points. This characteristic can help users find interesting knowledge. Providing high-quality knowledge now is a significant goal of the online public cultural services. Individual techniques for the goal have been developed to deal with cultural knowledge data, e.g., NLP approaches, graph-based algorithms. However, there is no suitable platform that makes the individual techniques work together well as a system. Therefore, constructing an integrated platform for discovery, management, and analysis of the massive Chinese public cultural knowledge become an important task of the online public cultural services. The platform restricts the knowledge in the field of public cultural services instead of common sense knowledge. The existing knowledge management systems for the public cultural service applied the traditional techniques for storing and analyzing public cultural knowledge. In the case of public cultural big data, the existing knowledge management systems do not work efficiently, particularly the knowledge acquisition from big data and analysis of massive knowledge. The solution is the knowledge graph technology that is good at big knowledge extraction, management, and analysis. In this paper, we propose a public cultural platform using a public cultural knowledge graph as the core. The platform collects and explores public cultural data for building and updating the knowledge graph. In order to enhance the cultural applications in the aspect of the intelligent services, e.g. resource recommendation, the platform provides service APIs for analyzing the knowledge graph.

The rest of the paper is organized as follows. Section 2 introduces the related work of knowledge graph. The concepts of the public cultural knowledge graph platform are proposed in Section 3. Section 4 presents the architecture of the platform. Section 5 shows an example scenario of the platform. Finally, conclusion and future work are given in Section 6.

II. Related Work

Sophisticated technologies have been applied for improving the online public cultural services. The project EDL integrated data of public cultural institutions over 20 European countries and provides sharing service for the public. The project OCLC[1] allowed libraries to upload their catalog data to World Cat Cloud Platform and share the data with each other. With the explosive data increment, the public cultural services have to face the big data. Zhang et al. [1] built a big data platform for integrating and sharing the Chinese public cultural resources. Yang et al. [2] introduced a novel visualization method for presenting the topic analysis for massive public cultural data. Zhai et al. [3] proposed a recommender system that provided adaptive methods for recommending public cultural resources. The above techniques can handle the massive public cultural data. However, the data were not converted to knowledge.

Knowledge graph represents massive knowledge as a graph structure. The knowledge graph consists of attributed entities and relations of entities. Google Knowledge Graph [4], which was powered by Freebase [5], was proposed to enhance the search engine in order to provide knowledge search instead of the traditional keyword search. Dong et al. [6] developed

Sponsors: (1) the Support Program of the National ‘12th Five-Year-Plan of China’ under Grant No. 2015BAK25B04; (2) the Support Program of the National ‘12th Five-Year-Plan of China’ under Grant No. 2015BAK25B03.

[1]OCLC. https://www.oclc.org/

978-1-5090-4284-5/17 $31.00 © 2017 IEEE
DOI 10.1109/ICSC.2017.37

Knowledge Vault for the web scale knowledge base by combining different web extractions. For representing social network, Facebook built an Entity Graph consisting of person entities and their social relations. Microsoft proposed Satori knowledge base and Concept Graph for supporting Bing search engine and cognitive toolkits. Hoffart el al. [7] presented a method for collective disambiguation of named entity using a coherence graph. Chen el al. [8] developed a system that can extract visual knowledge from Internet using a semi-supervised approach. Perona [9] presented a visual interface of Wikipedia for image query and visual knowledge organization. Kim et al. [10] proposed a probabilistic knowledge graph factorization method using the path structure of the existing knowledge and a modeling method for knowledge increment. Minervini el al. [11] proposed an approach for predicting missing links in large knowledge graphs using latent factor models. Yahya el al. [12] proposed a search engine for extended knowledge graphs containing relational facts and web text. Seufert et al. [13] introduced a system that can visually present the relationship of entity sets in knowledge graphs. Pujara et al. [14] proposed methods for better identifying knowledge graph by combining the ontological information and the distributional information.

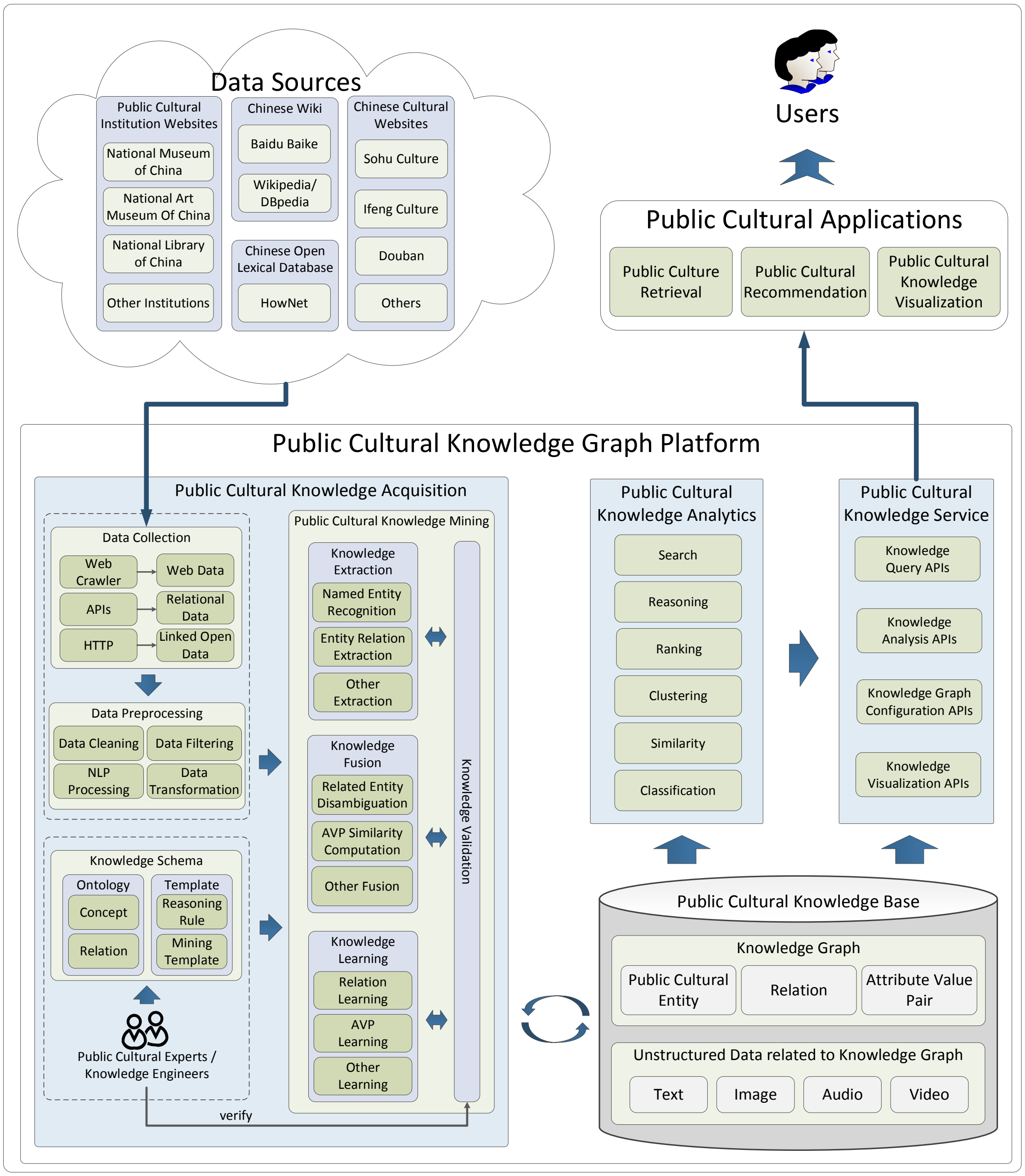

Figure 1 Public Cultural Knowledge Graph Platform

III. PUBLIC CULTURAL KNOWLEDGE GRAPH PLATFORM

In this paper we propose a public cultural knowledge graph platform (see Figure 1). The platform collects data from multiple data sources on the Internet. The platform discovers knowledge from the data and manages the knowledge in a knowledge base. The knowledge acquisition consists of two ways: knowledge population that extracts information from outside data sources, and knowledge completion that learns knowledge from the existing knowledge of the platform. The discovered knowledge is represented as a graph that consists of attributed entities and relations of entities, namely knowledge graph. The platform also provides knowledge analysis services for supporting the public cultural applications.

A. Data Sources

We collect data from multiple of data sources for the public cultural knowledge graph platform as follows.

1) Public Cultural Institution Websites

The public cultural institutions include public museums, e.g., National Museum of China, public libraries, e.g., National Library of China, public art galleries, e.g., National Art Museum of China, and public cultural centers of China. The public cultural institutions have plenty of cultural resources including articles, images, audios, and videos.

2) Chinese Wikis and Open Knowledge Base

Wiki exploration is a good way to collect the available data. In Wikis, the important information is summarized in specific tables, called Info-boxes. The most famous Chinese Wikis are Wikipedia Chinese version[2] and Baidu Baike[3]. Wikipedia is the largest wiki in the world. Baidu Baike is the largest Chinese wiki that contains over 13 million articles. The structured data in the info-boxes of wiki pages are extracted and transformed into Attribute-Value Pairs (AVPs). Open knowledge bases have been constructed in the decade, e.g., NELL [15], DBpedia [16], and YAGO [17]. DBpedia has extracted the structured data of Wikipedia and stored the data in the form of linked data.

3) Chinese Lexical Databases

The lexical databases tag the attributes of words, relations of words and synonyms. WordNet[4] is the most famous one. HowNet[5] is a famous Chinese lexical database that contains inter-conceptual relations and inter-attribute relations of concepts. HowNet is helpful for knowledge fusion, e.g., synonym fusion and entity linking.

4) Chinese Cultural Websites

Websites contain much textual cultural content. We built corpus based on the content for identifying semantics of Chinese words using NLP approaches. Moreover, the content can be used to analyze the relations of the Chinese cultural entities. The public cultural knowledge graph platform collects the famous Chinese culture websites, such as Sohu Culture[6], Ifeng Culture[7], and the vertical website douban[8].

B. Public Cultural Knowledge Graph Platform

1) Public Cultural Knowldege Acquisition

a) Data Collection: the public cultural knowledge graph platform collects data from the data sources in different ways: we developed a web crawler for collecting web pages; certain data may be got using specific APIs; we manually uploaded the data of HowNet and DBpedia to the platform.

b) Data Preprocessing: the platform processes the collected data before the knowledge extraction. Data Cleaning removes noise data. Data Filtering picks out the meaningful data useful for knowledge discovery. NLP Processing works on construting the corpus and building vector space. It provids basic processing methods for natural text, e.g, Chinese word segmentation, and POS-tagging. Data Transformation works on coverting heterogeneous data into required forms, e.g., vector and array.

c) Knowledge Schema: our public cultural experts and knowledge engineers build knowledge schemas. The ontology model is defined as

$$(C, A_C, R, A_R, I)$$

where C is concept set; A_C represents attribute set of concepts; R is relation set of concepts; A_R reprensets attribute set of relations; I means instance set. The instances work as seeds for the heuristic data collection methods. The concepts, the attributes of concepts, relations, and attributes of relations can be partialy extracted from HowNet. We use cultural data in DBpedia as priori knowledge to build instances of ontology models. We define the concepts with the following types:

$$(PER, OBJ, LOC, TM, EVT, ORG)$$

where PER represents person, OBJ indicates object, LOC means location, TM is time, EVT is event, and ORG presents organization. For data collection, the ontology models provide seeds that include important public cultural entities. The experts define templates including reasoning rules and mining templates as well. Reasoning rules are used for inference to identify new entities and relations. The reasoning rules are defined in the form of N-triples. Mining templates, e.g., bootstrapping, are applied for automatic knowledge discovery.

d) Public Cultural Knowledge Mining: the main public cultural knowledge mining tasks contains the following point:

- Knowledge Extraction works on extracting available knowledge from the preprocessed data. The core tasks are Named Entity Recognition (NER) and Relation Extraction (RE). To identify the named entities and the associated attributes, the platform applies the probabilistic method Conditional Random Field (CRF) and the template-based method Bootstrapping that uses the template predefined in knowledge schema.
- Knowledge Fusion focuses on merging the same or very similar entities, attributes, and relations. Entities extracted from different data may indicate the same thing, e.g., nicknames of a person. Related Entity Disambiguation that can fuse the same or very similar entities. The platform applies linked data-based matching methods. HowNet provides synonymous tags of Chinese words. We use this information to identify the same or very similar entities. If no

[2]Wikipedia. https://zh.wikipedia.org/
[3]Baidu Baike. https://baike.baidu.com/
[4]WordNet http://wordnet.princeton.edu
[5]HowNet. http://www.keenage.com/
[6]Sohu Culture. http://cul.sohu.com/
[7]Ifeng Culture. http://culture.ifeng.com/
[8]Douban. https://www.douban.com/

matching information exists in HowNet, we apply clustering analysis to entities according to semantic. We apply a binary logistic regression classifier to determine whether the entities should be combined together. In the same way, we fuse the same or very similar relations and AVPs.

- Knowledge Learning looks up potential knowledge in the knowledge found previously. By analyzing the existing relations with rule-based reasoning methods, Knowledge Learning can predict the relations possible to exist. We also use deep learning, e.g., Word2Vec, in order to find the correlation or similarity of entities that may be the potential relations. The knowledge learning can also find new AVPs by heuristic methods.
- Knowledge Validation allows public cultural experts and knowledge engineers to validate the knowledge obtained in the platform. The knowledge may not be completely correct, such as, logical reasoning errors caused by unsuitable rules, or algorithm precision issue. The incorrect knowledge may cause critical problems in the people's knowledge learning. Hence, the platform provides manual knowledge verification.

2) *Public Cultural Knowledge Base*

The public cultural knowledge base is core of the platform. The knowledge base logically manages the knowledge graph and the most related unstructured data. The knowledge graph includes public cultural entities, relations of the entities, and AVPs. The entities and relations build the basic structure of the knowledge graph. AVPs describe the attributes of the entities and relations in the form of structured data. The unstructured data most related to the knowledge graph are stored in distributed file systems and maintain the links of the data in the AVPs. The data include textual documents, images, audios and videos. For other possible related resources, the platform only maintains URLs of the resources in attributes.

3) *Public Cultural Knowldege Analytics*

The platform provides analytics for the knowledge graph. The analytics include analysis methods for the knowledge graph. The platform provides the analysis methods to the third-party applications in the form of RESTful APIs managed by the Public Cultural Knowledge Services described in 4).

a) Search: search of entities. Users are allowed to search entities according to given conditions of attributes. Knowledge filtering and grouping are also available by the search function.

b) Reasoning: logical inference starting with a specified entity. The inference can be available with the rules defined in the knowledge schema and the FOAF (Friend of a Friend) rule. For a given entity, the directly/indirectly related entities can be obtained with FOAF using graph traversal algorithms.

c) Clustering: clustering methods for the public cultural knowledge aggregation. The clustering analysis methods in the platform, e.g., K-Means, can assign entities or relations into different groups according to semantics.

d) Ranking: ranking methods for entities and relations of the knowledge graph. The ranking methods depend on either statistical data or the path ranking methods. The approaches combining the both methods are also possisble.

e) Similarity: entity similarity computation for achieving the semantically similar entities. We build the state-of-the-art text mining methods in the platform, e.g., topic modeling method LDA, word embedding method Word2Vec.

f) Classification: classification methods for entities and relations of the knowledge graph. The methods combine the traditional methods, e.g., KNN, with the conceptual taxonomy defined by the ontology models.

4) *Public Cultural Knowledge Services*

We build service APIs for the analytic methods of the Public Cultural Knowledge Analytics. In this way, the platform provides knowledge analytic for the third-party public cultural applications in the form of RESTful APIs. By the services, the applications possess an ability that effectively provides public cultural knowledge to users.

a) Knowledge Query: entity search with given conditions and information query for the specific entity. This service also provides knowledge aggregation that integrates the knowledge information users may be interested in.

b) Knowledge Analysis: analysis services for analyzing the knowledge graph including reasoning, clustering, ranking, similarity, and classification that are introduced in 3).

c) Knowledge Graph Configuration: configuration for setting the knowledge graph. The requests of a user may only need subgraphs of the knowledge graph. By the configuration services, the platform can focus on the relevant subgraphs instead of the entire graph. It can improve the efficiency of knowledge graph analysis.

d) Knowledge Visualization: services for helping public cultural knowledge applications build visualizations. The visual knowledge is more easy to understand than the textual format. The services provide JavaScript visualization library by extending the third-party visualization library D3.js.

C. *Public Cultural Applications*

The public cultural applications can be beneficial from the public cultural knowledge graph. The applications can become more intelligent and can better meet users' needs. The example scenarios are presented in Section V.

IV. ARCHITECTURE

We design the architecture of the public cultural knowledge graph platform based on Cloud Computing platform (see Figure 2). It includes IaaS, PaaS, and SaaS.

1) *IaaS Layer*

This layer manages the infrastructure of the platform. The IaaS layer manages physical resource, e.g., computers, network devices, and firewall devices. We apply virtualization in this layer for building the virtual resources, e.g., virtual machines and virtual networks. The virtual computer clusters are built using the virtual resources. We deploy the modules of the PaaS layer on the virtual computer clusters.

2) PaaS Layer

The PaaS layer depends on the virtual computer clusters of the IaaS layer and provides services to the applications in the SaaS layer. The PaaS layer has the following main modules.

a) Data Management Module: data store and data access. Relational databases store the raw data extracted from the outside data sources. The public cultural knowledge graph is stored in graph databases. Relevant documents are stored in document databases and related binary data, e.g., images, are stored in distributed file systems.

b) Model Module: manegement of the major models of the platform, including models in public cultural knowledge base, and in knowledge schema that are physically stored in corresponding datastores. The module build, configure, and update the models. The knowledge graph model is represented as a property graph that represents entities as vertices and relations as edges. The vertices and edges may have attributes. The property graph is stored in a graph database. The ontology is modeled using OWL. The reasoning rules and extraction templates are modeled using RDF in the platform.

c) Algorithm Module: management of the algorithms for knowledge acquisition and knowledge analysis, e.g., CRF for entity recognition, Word2Vec for identification of relations of entities, and N-gram for Chinese word segmentation.

d) Computing Module: management of the computing framework. The distributed parallel framework are good at batch processing of big data. We apply Hadoop MapReduce to process the massive data. The graph computing framework works on graph traversal algorithms. We deploy Apache Spark GraphX for graph computing. The In-memory computing and streaming computing frameworks focus on real-time and half real-time computing, e.g., online ranking. We use Apache Spark and its component Spark Streaming as the In-memory and streaming computing framework We apply SparQL and Gremlin for knowledge query. SparQL is a query language for RDF retrieval. Gremlin is a graph-based query language that provides scripts for graph traversal.

e) Knowledge Acquisition: acquisition of knowledge that are introduce in Section III.B.1).

f) Knowledge Graph Analytics: analysis methods for the knowledge graph, which are introduced in Section III.B.3) .

g) System Service Module: management of the platform system services for supporting the major modules. Security component works on the authentication and authorization of data access; Messaging component manages the messaging systems that provides asynchronous data transfer; Log system collects and manages the system logs; Monitoring system provides health-check for the platform by analyzing the logs managed by the log system; Configuration system manages the system configuration; Cache system is able to reduce the performance bottleneck of data transfer in the platform.

h) Public Cultural Knowledge Service APIs Module: management of service APIs. There are four API sets depicted in Section III.B.4): Knowledge Query APIs, Knowledge analysis APIs, Knowledge Graph Configuration APIs, and Visualization APIs. The framework builds an API registry to manage the APIs with a service catalog.

i) Runtime Environment Module: management of web services. We build servers in the platform including web servers, proxy servers, application servers, and API servers.

3) SaaS Layer

SaaS layer supports the public cultural applications. The applications provide advanced functionalities that depend on the services in the PaaS layer.

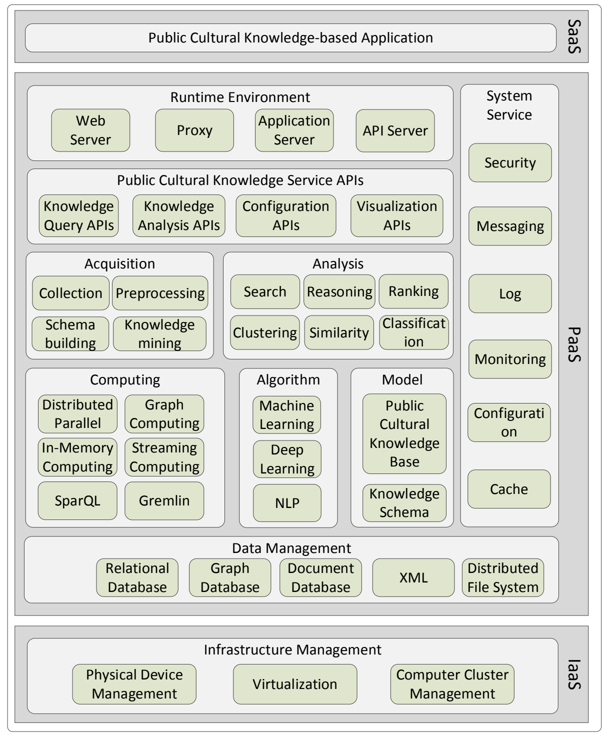

Figure 2 Architecture of Public Cultural Knowledge Graph Platform

V. EXAMPLE SCENARIO

This section presents an example scenario to show how the public cultural knowledge graph facilitates the public cultural applications regarding the public cultural recommendation. The public cultural recommender systems can provide users personalized public cultural information that the users may like. The commonly used recommendation methods include content-based recommendation methods and collaborative filtering methods. The content-based recommendation methods need to extract the important features of the resources. The resources have textual description for the resource content. The application sends the texts to the knowledge graph platform using the RESTful APIs provided by the platform. The texts can be used to build corpus for NLP of the knowledge graph platform. The platform can help discover the hidden features of the cultural resources from the corpus by analyzing the public cultural knowledge graph. The candidate resources can be tagged in terms of the found semantic features that describe the content of the resources. By mining the historical data of the users' visits, we can find the users' interests. The resources have tags complemented by the knowledge graph platform. For a user, the recommender system can analyze the tags of the resources visited by the user in order to form the interests of the user with tags. In this way, the content-based methods can recommend resources to the user by analyzing and matching the resource tags and user interest tags. The knowledge graph platform can strongly support for the content-based methods with respect to tag analysis. The collaborative filtering methods recommend resources by analyzing the co-occurrence of the users' visits. The collaborative filtering methods have the cold-start problem. For a new user with no visit, the collaborative filtering cannot work well. In this case, the commonly used solution is to integrate the resource features and user features into the collaborative filtering algorithms. The both types of features can be represented using tags. The tags can be extracted using the public cultural knowledge graph platform like for the content-based methods.

VI. CONCLUSION

Over the past a few years, the Chinese online public cultural services have developed rapidly. The massive public cultural data contains much knowledge. Constructing an integrated platform for knowledge acquisition and knowledge analysis for online public cultural applications becomes a critical task. In this paper, we propose a platform that discovers, integrates, manages, and analyzes the massive public cultural knowledge. The proposed platform represents the public cultural knowledge as a knowledge graph that consists of attributed entities and relations of entities. The knowledge graph is physically stored in graph databases and logically managed by a knowledge base. The knowledge graph helps the third-party applications for retrieval, recommendation, and visualization of the public cultural service. The platform can improves the knowledge understanding and sharing in the field of the public cultural service. The verification of the proposed knowledge graph platform is an important work in the future. We develop an integrated recommender system that is supported by our knowledge graph. The verification of the platform will be performed based on this system in the future work. In terms of techniques, we will focus on the named entity recognition and relation extraction, particularly in the Chinese culture field. We are going to apply deep learning methods for above tasks. The Recurrent Neural Networks is a worthwhile idea to work on.

ACKNOWLEDGMENT

We would like to thank all colleagues and graduate students who helped for our work. We also thank the National Key Technology Support Program of the National '12th Five-Year-Plan of China' under Grant No. 2015BAK25B04 and Grant No. 2015BAK25B03 that partially supported our research work.

REFERENCES

[1] G. Zhang, Y. Yang, X. Zhai, W. Huang and J. Wang, "Public Cultural Big Data Analysis Platform," 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), Taipei, 2016, pp. 398-403.

[2] Y. Yang, J. Wang, W. Huang and G. Zhang, "TopicPie: An Interactive Visualization for LDA-Based Topic Analysis," 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), Taipei, 2016, pp. 25-32.

[3] X. Zhai, F. Jin, J. Wang, Y. Yang, Q. Yao, Q. Qiu and G. Zhang, "A Kind of Precision Recommendation Method for Massive Public Digital Cultural Resources: A Preliminary Report," 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), Taipei, 2016, pp. 56-59.

[4] A. Singhal, "Introducing the knowledge graph: Things, not strings," The official Google blog, 2012: http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html accessed on 01-Nov-2016.

[5] K. Bollacker, C. Evans, P. Paritosh, T. Sturge and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," In Proceedings of the 2008 ACM SIGMOD international conference on Management of data (SIGMOD '08). ACM, New York, NY, USA, pp. 1247-1250.

[6] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun and Wei Zhang, "Knowledge vault: a web-scale approach to probabilistic knowledge fusion," In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14). ACM, New York, NY, USA, pp. 601-610.

[7] J. Hoffart, M.A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater and G. Weikum, "Robust disambiguation of named entities in text," In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11). Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 782-792.

[8] X. Chen, A. Shrivastava and A. Gupta, "NEIL: Extracting Visual Knowledge from Web Data," 2013 IEEE International Conference on Computer Vision, Sydney, VIC, 2013, pp. 1409-1416.

[9] P. Perona, "Vision of a Visipedia," in Proceedings of the IEEE, vol. 98, no. 8, pp. 1526-1534, Aug. 2010.

[10] D. Kim, L. Xie and C.S. Ong, "Probabilistic Knowledge Graph Construction: Compositional and Incremental Approaches," In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM '16). ACM, New York, NY, USA, pp. 2257-2262.

[11] P. Minervini, C. d'Amato, N. Fanizzi and F. Esposito. 2016. Leveraging the schema in latent factor models for knowledge graph completion. In Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC '16). ACM, New York, NY, USA, pp. 327-332.

[12] M. Yahya, D. Barbosa, K. Berberich, Q. Wang and G. Weikum, "Relationship Queries on Extended Knowledge Graphs," In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM'16). ACM, New York, NY, USA, pp. 605-614.

[13] S. Seufert, P. Ernst, S. J. Bedathur, S. K. Kondreddi, K. Berberich and G. Weikum, "Instant Espresso: Interactive Analysis of Relationships in Knowledge Graphs," In Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2016, pp. 251-254.

[14] J Pujara, H. Miao, L. Getoor and W. W. Cohen, "Ontology-aware partitioning for knowledge graph identification," In Proceedings of the 2013 workshop on Automated knowledge base construction (AKBC '13). ACM, New York, NY, USA, pp. 19-24.

[15] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E.R. Hruschka Jr. and T.M. Mitchell, "Toward an Architecture for Never-Ending Language Learning," In Proc' AAAI, 2010.

[16] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak and Z. Ives. "DBpedia: A nucleus for a web of open data". In The semantic web, 2007, pp. 722-735.

[17] F.M. Suchanek, G. Kasneci and G. Weikum, "Yago: a core of semantic knowledge," In Proceedings of the 16th international conference on World Wide Web (WWW '07). ACM, New York, NY, USA, pp. 697-706.