Learning Planning and Recommendation Based on an Adaptive Architecture on Data Graph, Information Graph and Knowledge Graph

Lixu Shao[1], Yucong Duan[1](✉), Zhangbing Zhou[2], Quan Zou[3], and Honghao Gao[4]

[1] State Key Laboratory of Marine, Resource Utilization in the South China Sea, College of Information and Technology, Hainan University, Haikou, China
751486692@qq.com, duanyucong@hotmail.com

[2] Department of Information Engineering, China University of Geosciences (Beijing), Beijing, China
zbzhou@cugb.edu.cn

[3] College of Computer Science, Tianjin University, Tianjin, China
zouquan@tju.edu.com

[4] Computing Center, Shanghai University, Shanghai, China
gaohonghao@shu.edu.cn

Abstract. With massive learning resources that contain data, information and knowledge on Internet, users are easy to get lost and confused in processing of learning. Automatic processing, automatic synthesis, and automatic analysis of natural language, such as the original representation of the resources of these data, information and knowledge, have become a huge challenge. We propose a three-layer architecture composing Data Graph, Information Graph and Knowledge Graph which can automatically abstract and adjust resources. This architecture recursively supports integration of empirical knowledge and efficient automatic semantic analysis of resource elements through frequency focused profiling on Data Graph and optimal search through abstraction on Information Graph and Knowledge Graph. Our proposed architecture is supported by the 5W (Who/When/Where, What and How) to interface users' learning needs, learning processes, and learning objectives which can provide users with personalized learning service recommendation.

Keywords: Resource modeling · Knowledge Graph
Service recommendation · Semantic modeling

1 Introduction

Information overload and information confusion caused by exponential increase of Internet information restrict users' efficient use of resources. Knowledge Graph has become a powerful tool for expressing knowledge in the form of a marked

I. Romdhani et al. (Eds.): CollaborateCom 2017, LNICST 252, pp. 323–332, 2018.
https://doi.org/10.1007/978-3-030-00916-8_30

directed graph and can endow text information with semantics. Knowledge base contains a set of concepts, examples and relationships [5]. The author of [10] took topic model as the basis for similarity calculation and obtained the entity catalog from Wikipedia. A large number of supervised learning methods, semi-supervised learning methods [2] based on eigenvectors appeared for achieving relation extraction. The authors proposed an information extraction framework for open domains and released an open information extraction prototype system which was based on self-supervised learning [1]. In [13] the author elaborated concepts of data, information, and knowledge. In [4] the authors divided the information extraction into three levels: entity, relation and attribute.

From the perspective of extending the concept of existing Knowledge Graph, we propose a three-layer architecture composing Data Graph, Information Graph and Knowledge Graph. Through modeling massive resources, users can quickly and accurately search the information they need in the resource processing architecture. With the help of classification of 5Ws (Who, When, Where, What and How) questions [3], it's easy to get the description of users' learning needs, learning processes and learning objectives. The 5Ws are questions whose answers are considered basic in information gathering or problem solving. They are often mentioned in journalism, research, and police investigations.

2 The Establishment of Resources Framework

We divide the expression of Knowledge Graph into three parts including DataGraph$_{DIK}$, InformationGraph$_{DIK}$ and KnowledgeGraph$_{DIK}$ based on the relationship among them. We define typed learning resource elements and the three-layer graph as follows:

Definition 1. *Resource elements. Resource elements include data, information and knowledge. We define resource elements as:*

Elements$_{DIK}$: = <Data$_{DIK}$, Information$_{DIK}$, Knowledge$_{DIK}$>.

Definition 2. *Graphs. We extend existing concept of Knowledge Graph and clarify expression of Knowledge Graph into Data Graph, Information Graph and Knowledge Graph.*

Graph$_{DIK}$: = <(DataGraph$_{DIK}$), (InformationGraph$_{DIK}$), (Knowledge Graph$_{DIK}$)>.

Figure 1 shows the framework of resources processing based on Data Graph, Information Graph and Knowledge Graph. Table 1 shows the explanations of resource type and corresponding graphs.

Figure 2 shows the classification and transformation of 5W questions and relationships between the three-layer graphs. To be exactly, we calculate three frequencies of Data$_{DIK}$ in DataGraph$_{DIK}$. With the help of classification of 5W questions, it's easy to get the description of users' learning needs, learning processes and learning objectives and provide users with personalized learning recommendation service.

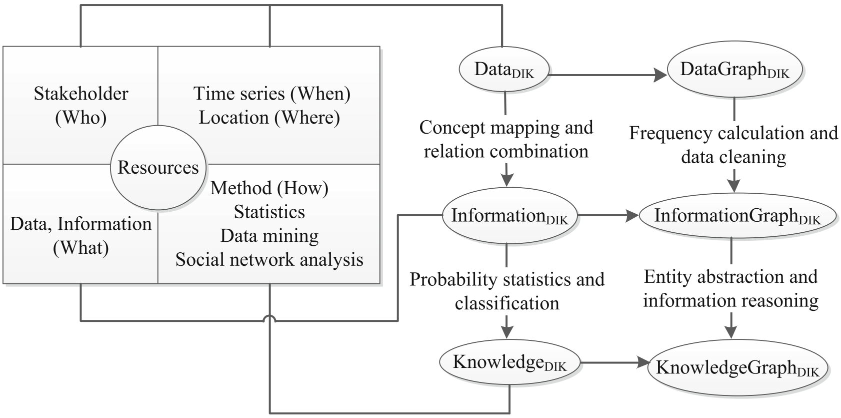

Fig. 1. Resources processing framework towards "5W" based on Graph$_{DIK}$

Table 1. Explanations of typed resources

	Data$_{DIK}$	Information$_{DIK}$	Knowledge$_{DIK}$
Semantic load	Not specified for stakeholders/machine	Settled for stakeholders/machine	Abstracting resources for revelation on unknown
Format	Discrete elements	Related elements	Probabilistic or categorization
Usage	Identification of existence after conceptualization	Communication	Reasoning and prediction
Answer	Who/When/Where	What	How
Graph	DataGraph$_{DIK}$	InformationGraph$_{DIK}$	KnowledgeGraph$_{DIK}$

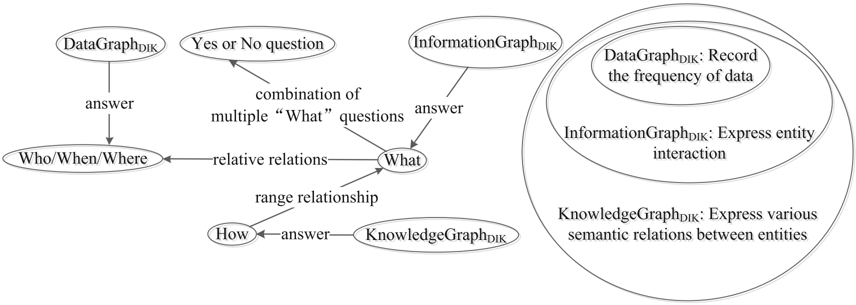

Fig. 2. Classification of "5W" questions and relationship between graphs

2.1 Statistics and Analysis of Data$_{DIK}$ Frequency Based on DataGraph$_{DIK}$

We use DataGraph$_{DIK}$ to record three kinds of frequencies of Data$_{DIK}$ including structural frequency, temporal frequency and spatial frequency. Structural frequency is defined as the number of times that Data$_{DIK}$ appears in different data structures. Temporal frequency is defined as the temporal trajectory of Data_{DIK}. Spatial frequency is defined as the spatial trajectory of Data_{DIK}. We give an example about frequency statistics of Data_{DIK} when modeling Data_{DIK} resource of knowledge points. As shown in Fig. 3, temporal frequency of Data_{DIK} of the knowledge points represents the number of classes of Data_{DIK}. Spatial frequency represents times of Data_{DIK} appeared in different professional systems. Structural frequency represents the educational mode of Data_{DIK}.

Definition 3. *Structural frequency (stru_f). stru_f represents the number of times of $Data_{DIK}$ appearing in different data structures. stru_f of $Data_{DIK}$ should be calculated according to the biggest data structure where $Data_{DIK}$ is appeared. For example, if a $Data_{DIK}$ is displayed as a node of a tree structure in a branch of a graph structure, stru_f of the $Data_{DIK}$ is 1 and the frequency of $Data_{DIK}$ displayed in a tree structure is not calculated repeatedly.*

Definition 4. *Spatial frequency (spat_f). spat_f is the occurring times of $Data_{DIK}$ appearing at different spatial locations describing the relative locations of multiple objects.*

Definition 5. *Temporal frequency (temp_f). temp_f represents times of $Data_{DIK}$ appearing in different time periods. For streaming $Data_{DIK}$ with timeliness, once we observe those $Data_{DIK}$ we should make responses in a timely manner because expired $Data_{DIK}$ will be meaningless.*

Definition 6. *$Data_{DIK}$ comprehensive frequency (DFreq). We define comprehensive frequency of a node as follows:*

DFreq: $= < stru\_f, spat\_f, temp\_f >$.

2.2 Automatic Abstraction Processing of Information_{DIK} Resources Based on $\text{InformationGraph}_{DIK}$

2.2.1 Recording the Interactivity of Nodes Based on $\text{InformationGraph}_{DIK}$

We define $\text{InformationGraph}_{DIK}$ as a directed graph $G\ (V,\ E)$, where V is a set of nodes. E is a set of lines. We measure the importance of nodes on $\text{InformationGraph}_{DIK}$ according to the comprehensive degree. Comprehensive degree, denoted as Com_degree, can be calculated according to Eq. 1:

$$Com\_degree = \sqrt{deg^{+} * deg^{-}} \tag{1}$$

where deg^{+} represents in-degree of nodes and deg^{-} represents out-degree of nodes. Therefore, further importance measurement of nodes on $\text{InformationGraph}_{DIK}$, denoted as "Impor", can be calculated according to Eq. 2:

$$Impor = \alpha DFreq * \beta Com\_degree \tag{2}$$

where α and β represent weight of comprehensive frequency of nodes on DataGraph_{DIK} and comprehensive degree of interaction between nodes on $\text{InformationGraph}_{DIK}$ respectively. Both α and β can be obtained through data training.

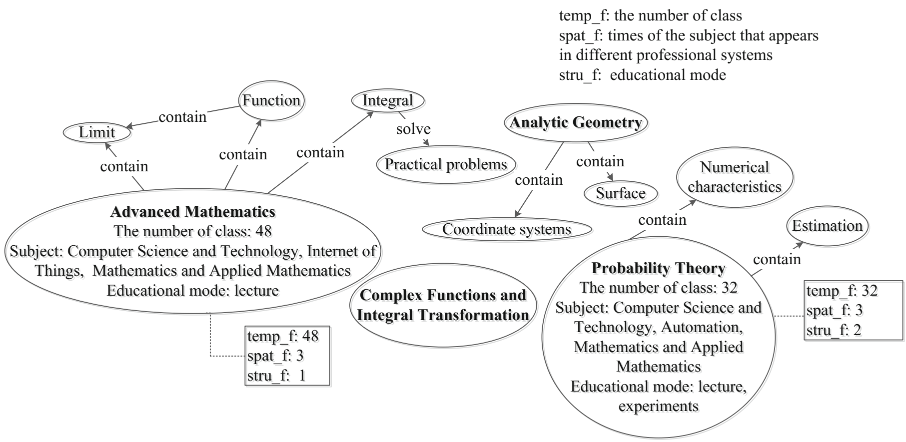

Fig. 3. Statistics about structural frequency, spatial frequency and temporal frequency based on DataGraph$_{DIK}$

2.2.2 Processing Data$_{DIK}$ Integration and Entity Abstraction Based on InformationGraph$_{DIK}$

InformationGraph$_{DIK}$ reflects communications and multiple interactions between entities. As shown in Fig. 4, new concept is generated by integrating Data$_{DIK}$ in Fig. 3. Different Data$_{DIK}$ resources are linked closely with each other in form of data structures including list, queue and tree. In order to improve the expression of resource architecture, we gather closely related Data$_{DIK}$ collection so as to analyze and abstract Information$_{DIK}$ resources. After gathering a certain number of entities as a group, we calculate the ratio of internal and external degree of interaction between entities according to Eq. 3:

$$cohesion = \frac{IFreq_{II}}{IFreq_{EI}} \tag{3}$$

where *cohesion* represents the ratio of internal degree of interaction and external degree of interaction. We restrict the related Data$_{DIK}$ collection must be connected to each other. *Cohesion* is an indicator of the degree of correlation between entities. $IFreq_{EI}$ represents times of external interactions between entities and $IFreq_{II}$ represents times of internal interactions between entities.

2.3 Analysis and Processing of Knowledge Based on KnowledgeGraph$_{DIK}$

2.3.1 Information Reasoning and Knowledge Forecasting on KnowledgeGraph$_{DIK}$

Path ranking algorithm uses each of different relational paths as a one-dimensional feature. Through constructing a large number of relational paths in KnowledgeGraph$_{DIK}$ to construct the eigenvector of relation classification

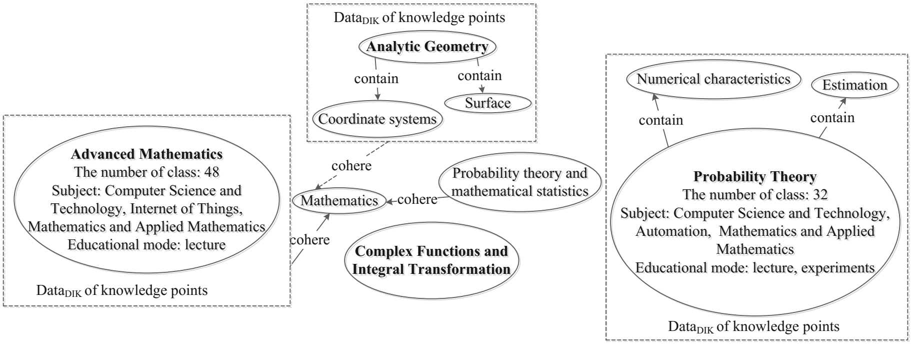

Fig. 4. Automatic generation of entities based on InformationGraph$_{DIK}$

and a relational classifier to extract the relationship between entities. Actual model abstractions should include abstractions from a set of existing relational entities and relationships to one or several new entities and new relationships. We create a truth table to indicate whether the information obtained through entity and relationship reasoning is true. We calculate the rate of correctness of a relationship, denoted as Cr (E1, R, E2), according to Eq. 4:

$$Cr(E1, R, E2) = \frac{\sum_{\pi \in Q} P(E_1 \rightarrow E_2)\theta(\pi)}{|Q|} \tag{4}$$

where *P(E1, E2)* indicates one path between *E1* and *E2*, *Q* indicates all paths starting from *E1* and ending with *E2*, represents the weight obtained by training.

2.3.2 Importance Measurement of Nodes on KnowledgeGraph$_{DIK}$

Knowledge Graph have been widely adopted, in large part owing to their schema-less nature that enables Knowledge Graph to grow seamlessly and allows for adding new relationships and entities as needed. Each semantic relationship on KnowledgeGraph$_{DIK}$ has its own weight that is denoted as θ to represent the importance of the relationship. We evaluate the importance of nodes, denoted as *Final_impor* in KnowledgeGraph$_{DIK}$ according to Eq. 5 through marking and manipulating different types of semantic relationships between entities.

$$Final\_impor = \alpha DFreq * \beta Com\_degree * \gamma \frac{\sum_{i=1}^{n} \lambda_i * Rel_i}{n} \tag{5}$$

where λ_i represents the weight of relationship Rel_i and n represents the amount of relationship types.

3 Characteristics Analysis of Application of DIK Architecture

Through analyzing frequency statistics and calculating of Data$_{DIK}$ based on DataGraph$_{DIK}$, we filter Data$_{DIK}$ with lower integrated frequency and reduce appearance of false and useless Data$_{DIK}$. Unifying expressive forms of Data$_{DIK}$ based on InformationGraph$_{DIK}$, we eliminate redundant Data$_{DIK}$. Through integrating frequently-interactive Data$_{DIK}$ together, we can summarize general rules of Information$_{DIK}$. Relationship of Data$_{DIK}$ and Information$_{DIK}$ is lack of hierarchies and logicality. As a result, through classifying and statistically analyzing Information$_{DIK}$, taking empirical knowledge into consideration, we speculate unknown information and get probabilistic answers.

Because of restriction of relationships between Data$_{DIK}$, Information$_{DIK}$ and Knowledge$_{DIK}$, users cannot obtain Information$_{DIK}$ and Knowledge$_{DIK}$ directly based on DataGraph$_{DIK}$. They cannot obtain Knowledge$_{DIK}$ based on InformationGraph$_{DIK}$ as well. On cross-layer searching there are following cases including some resources cannot be searched, sometimes endless resources are searched or obtained resources don't match with users' information needs. We model these resources and search the integrated resources according to three-layer graphs. We use different resource architectures according to different cases.

4 Application of Learning Recommendation Service Based on DIK Method

Knowledge points are divided into three layers including Data$_{DIK}$, Information$_{DIK}$ and Knowledge$_{DIK}$. It is easy to get description of users' learning needs and learning processes based on architecture of Graph$_{DIK}$ with the introduction of "5W" questions. Through establishing learners' models and acquiring their current learning state, ability level and learning target, we provide users with an effective learning strategy. After getting learner's expected effort and learning objectives, we compute the learner's expected learning efficiency according to Eq. 6:

$$Expected\_effi = \frac{Total\_know}{Expected\_effort} \tag{6}$$

where $Total\_know$ represents total amounts of knowledge points that the learner has to study to achieve his/her target.

Figure 5 shows partial knowledge system of computer science. We assume that a user's learning objective is to master the course of principle of communication. Courses that he or she has been mastered and hasn't been mastered yet have been marked in the resource system. Each number marked with a course represents the average effort that users must put in order to master this course. If a user's expected effort is 90, then we can recommend the three learning programs as Fig. 5 shows.

According to the type of user's learning objectives, we choose the corresponding layer of resource processing architecture to traversal. Marking the knowledge points that the learner has learned and will learn on the resource processing architecture and traversing graphs, we can obtain all preorder nodes of target

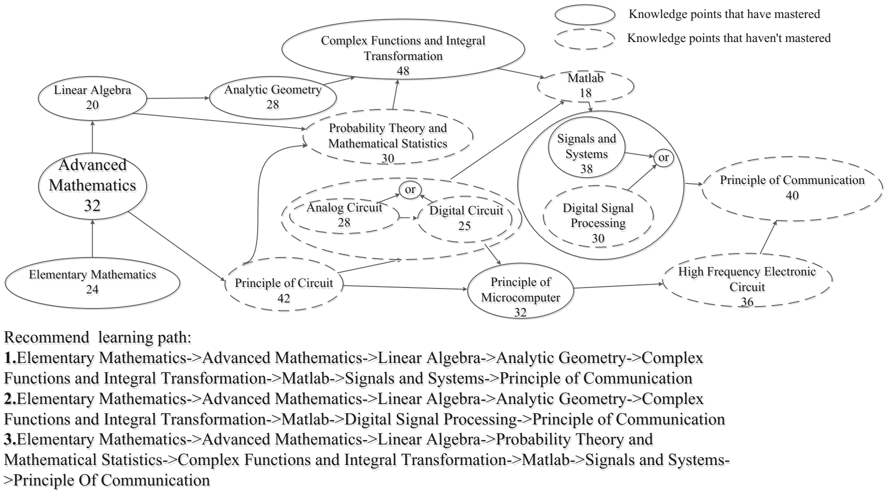

Fig. 5. An example of learning recommendation service based on Graph$_{DIK}$ resource processing architecture

knowledge points and recommend the complete learning path to users. We compute the actual learning efficiency of a user according to Eq. 7:

$$Actual\_effi = \frac{Got\_know}{Actual\_effort} \tag{7}$$

where Got_know represents the knowledge point that the learner has mastered and $Actual\_effort$ represents the learner's actual learning effort. Through analyzing changes of learner's ability, we can update learner model. According to the updated learner model and resource processing architecture, we re-acquire the learner's learning objectives and expected learning effort.

5 Related Work

Efficiency of search engines depends on a large number of labeled data. With the introduction of statistical machine learning methods, coreference resolution technology has entered a rapid development stage. Using the planned and extracted knowledge base, it extracted millions of rules from unlabeled question corpus and multiple knowledge bases to search a solution to problem analysis and query refactoring [8]. Development of software service system can be divided into stages of data sharing, information transfer, and knowledge creation in terms of data, information and knowledge [11]. In [12] the authors proposed a method to promote cross-language knowledge linking through concept annotations, which enriched cross-language knowledge links, and promoted knowledge sharing in different languages. Regional collaborative medical data centers are established to achieve integration among hospital information. Scientists have proposed numerous models for defining anything "as a service (aaS)", including discussions of products, processes, data and information management as a service [6]. In [9] the authors proposed a semantic-oriented cross-language ontology mapping (SOCOM) framework to enhance the interoperability of ontology-based systems involving multilingual knowledge base. In [7] the authors clarified the architecture of Knowledge Graph as a whole and extended the existing concept of Knowledge Graph into four aspects including Data Graph, Information Graph, Knowledge Graph and Wisdom Graph.

6 Conclusion

With the rapid development of Internet, it is a great challenge to quickly and accurately search resources that satisfy users' requirements. We propose a learning planning and recommendation approach based on an adaptive architecture composing Data Graph, Information Graph and Knowledge Graph. The architecture introduces empirical and theoretical rules to modelling Data_{DIK}, Information_{DIK}, Knowledge_{DIK} and deals with inconsistency, redundancy and lack of resources effectively. We provide users with learning planning and recommendation service based on the adaptive architecture on DataGraph_{DIK}, $\text{InformationGraph}_{DIK}$ and $\text{KnowledgeGraph}_{DIK}$ aiming at improving the effectiveness and efficiency of learning. Through automatic abstraction and dynamic planning, the architecture will recommend optimized service to users after effective and adaptive resources search and self-organization. Currently we have interpreted our work with cases in all segments and in the next stage we will expand the scale of dataset to further verify the feasibility of our approach.

Acknowledgment. This paper is supported by NSFC under Grant (No. 61363007, No. 61662021), NSF of Hainan No. ZDYF2017128 and Hainan University Project (No. hdkytg201708).

References

1. Banko, M., Cafarella, M.J., Soderland, S., Broadhead, M., Etzioni, O.: Open information extraction from the web. In: IJCAI 2007, pp. 2670–2676 (2007)
2. Carlson, A., Betteridge, J., Wang, R.C., Hruschka Jr., E.R., Mitchell, T.M.: Coupled semi-supervised learning for information extraction. In: WSDM 2010, pp. 101–110 (2010)
3. Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Ths, H.: A reference model for learning analytics. Int. J. Technol. Enhanc. Learn. **4**(5/6), 318–331 (2012)
4. Cowie, J.R., Lehnert, W.G.: Information extraction. Commun. ACM **39**(1), 80–91 (1996)
5. Deshpande, O., et al.: Building, maintaining, and using knowledge bases: a report from the trenches. In: ACM SIGMOD 2013, pp. 1209–1220 (2013)
6. Duan, Y., Fu, G., Zhou, N., Sun, X., Narendra, N.C., Hu, B.: Everything as a service (XaaS) on the cloud: origins, current and future trends. In: IEEE International Conference on Cloud Computing, pp. 621–628 (2015)

7. Duan, Y., Shao, L., Hu, G., Zhou, Z., Zou, Q., Lin, Z.: Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph. In: 15th IEEE SERA 2017, pp. 327–332 (2017)
8. Fader, A., Zettlemoyer, L., Etzioni, O.: Open question answering over curated and extracted knowledge bases. In: 20th ACM SIGKDD 2014, pp. 1156–1165 (2014)
9. Fu, B., Brennan, R., O'Sullivan, D.: Cross-lingual ontology mapping and its use on the multilingual semantic web. In: International Workshop on the Multilingual Semantic Web, pp. 13–20 (2010)
10. Sen, P.: Collective context-aware topic models for entity disambiguation. In: International Conference on World Wide Web, pp. 729–738 (2012)
11. Shao, L., Duan, Y., Sun, X., Zou, Q., Jing, R., Lin, J.: Bidirectional value driven design between economical planning and technical implementation based on data graph, information graph and knowledge graph. In: 15th IEEE SERA 2017, pp. 339–344 (2017)
12. Wang, Z., Li, J., Tang, J.: Boosting cross-lingual knowledge linking via concept annotation. In: IJCAI 2013, pp. 2733–2739 (2013)
13. Zins, C.: Conceptual approaches for defining data, information, and knowledge. J. Assoc. Inf. Sci. Technol. **58**(4), 479–493 (2007)