

# Knowledge Flow-based Semantic Organization of Resources about Foreign Language Learning

Li Li

Shanghai University of Political Science and Law  
Shanghai, China  
e-mail: lily2211@126.com

**Abstract**—With the advent of the big data age, more and more resources about foreign language learning are put on the Web and open to the public, which brings trouble on the foreign language learners in acquiring these unfamiliar learning resources effectively and efficiently. To solve the problem, this paper proposes a method of organizing the learning resources in the form of knowledge flow network and provides learners with learning resources in the form of knowledge flow. Compared with other semantic organizations of resources, the proposed knowledge flow-based organization of resources not only helps the learners obtain the target resources properly but also displays them the study route related to the target resources. Experimental results indicate that the proposed method can improve the utilization rate of learning resources and the efficiency of Foreign Language Learning.

**Keywords**- Foreign Language Learning; semantic organization; recommendation system; knowledge flow; knowledge mining; e-learning.

## I. INTRODUCTION

With the advent of the big data age, more and more resources about language learning are put on the Web and open to the public, such as course websites of universities, learning resources database, MOOCs, etc.. A foreign language learner can get the resources not only from his teachers but also from the Web which contains a large amount of high quality resources from the first language country. There is so much variety that it is difficult for a beginner of foreign language learner to obtain the appropriate resources from such massive online learning resources.

Many methods have been proposed to help a user find learning resources more easily and exactly. Search engines, such as Google, Bing, Baidu, etc., are the most popular tools. Search engines consider all learning resources on the Web as common webpages. Search engines crawl the learning resources from the Web, indexed them by terms, provide a search service based on keyword matching for the users. However, the search engine has trouble with search learning resources effectively for the beginner of foreign learning, which is caused by the implementation of search engine. The realistic scene is as following. When a learner searches a resource by a keyword, the search engine will feedback him a large amount of results containing the keyword, which may puzzle the user how to select an appropriate result from the returned list. To solve the problem of search engine,

recommendation systems are proposed to provide more exact results for the users based on some semantic computing technologies, such as, replacing keyword matching with similar computing, organizing the resources in semantic forms, and recommending the related resources based on association relations. Although recommendation systems work better than traditional search engines, they cannot still meet the requirement of the beginners of foreign language learning. For example, the recommended results are discrete, which is of no help in providing well-organized learning resource for the foreign language learners and improving the efficiency of language learning.

The organization of learning resources determines the final service pattern based on the resources. To the beginners of foreign language learning, they expect to get not only the target resource but also the knowledge relations between the target resource and other related learning resources. For example, what are basic knowledge points he must acquire before starting to learn this knowledge point? What is the next knowledge point he can study after finishing learning this knowledge point? We summarize the following shortcomings of the current organization structure which have to be confronted by the learning resources searching task.

- Lack of the proper sequence of knowledge resources. Both the linear index structure employed by search engines and the semantic index employed by recommendation system do not hold the knowledge sequence of learning resources
- Isolated islands of learning resources. There are some isolated islands in resources utilization rates of which are very low. When a resource is at the tail of index sequence, the possibility that will be accessible is very low.

Based on the above analysis, the knowledge flow is suitable to organize knowledge points of learning resources to support the above requirements. Therefore, this paper proposes a new semantic organization of learning resources, knowledge flow-based semantic organization structure, which organizes the learning resources in the form of knowledge flow network and provide learners with learning resources in the form of knowledge flow. The framework of knowledge flow-based service of learning resources is shown in Figure 1, the main components of which will be discussed in Section III and Section IV. To demonstrate that the proposed methods are

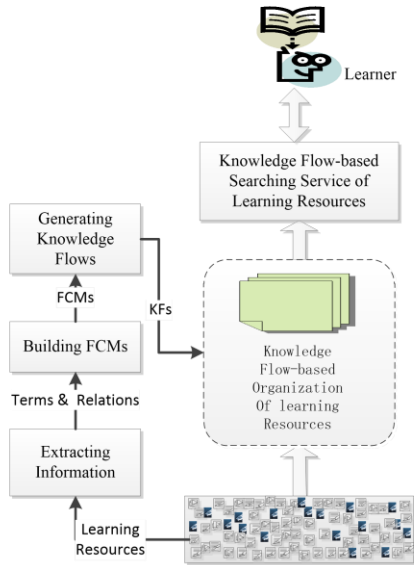


Figure 1. The framework of knowledge flow-based service of learning resources

effective, experiments are conducted in this paper and the results show that our approach works well in improving the utilization rate of learning resources and the efficiency of foreign language learning. The contributions of the paper are as follows:

- Propose a method of representing the learning resources based on fuzzy cognitive map, which is efficient for mining the semantic relations among learning resources.
- Propose a knowledge flow-based semantic organization structure to organize the learning resource, which can support new types of knowledge services.

The rest of this paper is organized as follows. In Section II, related work is discussed. In Section III, a method of representing the learning resources based on fuzzy cognitive map is proposed. In Section IV, a knowledge flow-based semantic organization structure is proposed to organize the learning resources. In Section V, experimental results are shown to verify the effectiveness of the proposed methods. Finally, Section VI concludes this paper and offers further research work.

## II. RELATED WORK

### A. Representation of Learning Resource

Most of learning resources on the Web are shown as documents[13], which can be presented by Vector Space Model (VSM), Fuzzy Cognitive Map (FCM), and so on.

Vector space model or term vector model is an algebraic model for representing text documents (and any objects, in general) as vectors of identifiers, such as, for example, index terms [1]. A definition of a document is that it is made of a joint membership of terms which have various patterns of occurrence [2]. These occurrence patterns are often disregarded because of complexity and single word statistics, i.e, document indexing, are used instead [3][5]. The IR pioneer Luhn used the

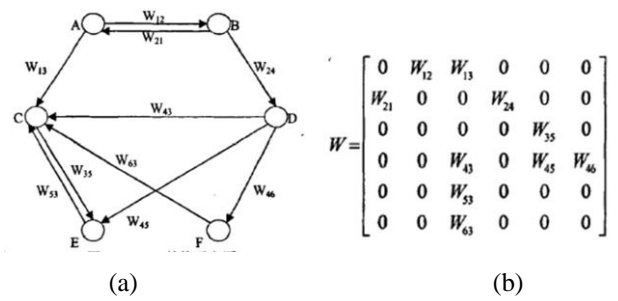


Figure 2. A sample of FCM and its adjacency matrix

term frequency as content descriptive features for documents [4] and it is still today the most used document description method, see e.g. [6][7].

A Fuzzy Cognitive Map (FCM) is a graphical model for knowledge representation and reasoning[8]. Knowledge representation is the basic ability of FCM. Compared with VSM, FCM consists of concepts and relations among concepts. When it is used to present learning resource, it is better than VSM in the aspect of presentation ability. Furthermore, FCM can represent not only causal relations between concepts but also knowledge of different granularity levels. Recent years, FCM has been improved in various aspects [9][10][11][12]. Elpiniki proposed a method to train a FCM, which is based on the nonlinear Hebbian learning and the differential evolution algorithm [8]. Wojciech proposed a novel parallel approach to the learning of FCM in order to deal with large maps due to its high computational complexity [15]. Mateou proposed two algorithms for a multilayer approach developed to expand the capabilities of FCM, one is ML-FCM (Multilayer FCM), and the other is EML-FCM (Enhanced Multilayer FCM) [14]. Luo et al. utilized the Hebbian Learning Rule and the Unbalance Degree to improve a FCM, so as to make it applicable to self-learning [16]. In this paper, FCM is used as the basic presentation model for Foreign Language Learning resources.

### B. Semantic Organization of Learning Resources

Learning resources should be organized in some semantic forms in order to support high efficient searching and recommendation for Foreign Language Learning and Teaching. Because learning resources are usually documents, they can be organized by the same methods with those of Web content organization [24][25]. Semantic Web employs such techniques as URI, XML, RDF, OWL to identify, describe and associate the resources and settles the problem of semantic association of Web/learning resources. But because ontology can't be fully automatic, it is difficult to deal with the increasing number of resources and apply to semantic organization of massive learning resources. In [17], a resource space model is proposed based on semantic link network. [18][19][20] use semantic link network to organize resources efficiently. [19] proposed the construction method of association semantic link network. [18] builds a multi-layer association link network of keywords.[20] proposed the method to build the similar semantic link network on a large scale of web resources. In this paper, semantic link network is used as a basic tool to organize knowledge flows of learning resources.

### III. REPRESENTATION OF LEARNING RESOURCE BASED ON FUZZY COGNITIVE MAP

Because most of learning resources are documents (webpages) or can be converted into documents, we use Fuzzy Cognitive Map (FCM) to represent a learning resource document. FCM is a graphical model which is comprised of concepts (nodes) and the relations between concepts (edges). Figure 2(a) shows the logic structure of FCM and Figure 2(b) shows the storage structure of FCM [22].

A document can also be described as a series of keywords and the relations among keywords. According to the structure of FCM, a learning resource is defined as:

**Definition 1. Learning Resource (LR):** A learning resource is formed by attributive keywords and the semantic relations among attributive keywords and denoted as

$$LR = \left\langle P, R^P \right\rangle = \left\langle \begin{matrix} P = \{t_k \mid t_k \in T, 0 \leq k \leq |T|\}, \\ R^P = \{(t_i, t_j, w) \mid t_i, t_j \in P, 1 \leq i, j \leq |P|, 0 \leq w \leq 1\} \end{matrix} \right\rangle, (1)$$

in which,  $T$  is the domain keywords list,  $P$  is the attribution set of Learning Resource  $LR$  and is formed of the keywords  $t_k$  that belongs to  $T$ .  $|T|$  is the size of  $T$ .  $R^P$  is the set of semantic relations in  $P$ , each semantic relation is described by a triad  $(t_i, t_j, w)$ , in which  $w$  denotes the strength of the relation between keywords  $t_i$  and  $t_j$ .  $|P|$  is the size of  $P$ .

Given a learning resource, to represent it in the form of definition 1, two major tasks should be fulfilled.

The first one is to extract concepts from the given learning resource. There are many researches on how to extract keywords/concepts from the document [21]. This paper just takes advantage of the existed methods to do the work.

The second one is to mine the semantic relations among keywords/concepts. It is difficult to mine the causal relations among the concepts based on a single document. This paper uses association relation to replace the causal relation among concept nodes, because the causal relation is a special association relation, easy to mine from document automatically. There are also many methods to mine the association relation, such as the Apriori Algorithm and all kinds of improved algorithms [23].

### IV. ORGANIZATION OF LEARNING RESOURCES BASED ON KNOWLEDGE FLOW

To the learning of foreign language, the amount of learning resources is very large, on the contrary, the knowledge points of the language are much fewer than learning resources. Each learning resource consists of one or several knowledge points and can be mapped to the basic knowledge points. The knowledge flow can be used to organize both the basic knowledge points and the learning resources.

#### A. Knowledge Flow

**Definition 2. Knowledge Flow (KF):** Knowledge Flow is a linear directed graph whose vertices are knowledge points (or

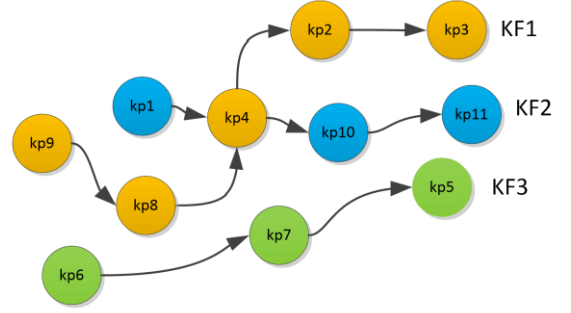


Figure 3. A sample of knowledge flow network

learning resources) and whose edges are the relations among knowledge points. All vertices of KF are listed in the order of  $\{lr_1, lr_2, \dots, lr_n\}$  and  $i = 1, 2, \dots, n-1$ . The edge  $\langle lr_i, lr_{i+1} \rangle$  means the knowledge point (or learning resources)  $lr_i$  should be learnt before  $lr_{i+1}$ .

A sample of knowledge flow network is shown in Figure 3. In Figure 3, the network consists of 11 knowledge points which are listed in Table I. For example,  $\{kp1, kp4, kp10, kp11\}$  in Figure 3 is a knowledge flow with four knowledge points which indicates that the best learning route should be  $kp1 \rightarrow kp4 \rightarrow kp10 \rightarrow kp11$ .

The procedures of building the knowledge flows are as follows:

- Step 1: Mine knowledge points from learning resource. (This can be realized by concept extraction method.)
- Step 2: Verify the mined knowledge points by domain experts (in FLL they may be the teachers).
- Step 3: Mine the association relations among knowledge points (This can be realized by association relation mining method, such as Apriori.)
- Step 4: Connect the knowledge points by the association relations mined in step 3 and get the initial knowledge flows.
- Step 5: Verify the knowledge flows by domain experts (in FLL they may be the teachers).

When the knowledge flow of basic knowledge points is mined, the knowledge flow of learning resources can be obtained according to the maps from learning resources to knowledge points.

TABLE I. EXAMPLES OF KNOWLEDGE POINT IN FLL

No	Knowledge Point in FLL
kp1	word
kp2	the present tense
kp3	the future tense
kp4	the perfect tense
kp5	attributive clause
kp6	adverbial clause
kp7	object clause
...	...

### B. Knowledge Flow Network

**Definition 3. Knowledge Flow Network (KFN):** Knowledge Flow Network is a set of knowledge points and their knowledge flows. KFN is denoted as

$$KFN = \{KP, KFS\}, \quad (2)$$

where  $KP$  is a set of knowledge points and  $KFS$  is a set of knowledge flows in  $KP$ .  $KFS$  is denoted as

$$KFS = \{kf_1, kf_2, \dots, kf_i, \dots, kf_n\}, \quad (3)$$

where  $kf_i$  is the  $i^{th}$  knowledge flow which is defined by Definition 2.

Knowledge flow network combines knowledge flows to support flexible usage of knowledge flows. For example, in Figure 3, three knowledge flows (shown in different colours) are combined into a network. Knowledge flows KF1 and KF2 have a joint knowledge point and connect together. When  $kp4$  is checked, there are two available learning directions for the next step of learning.

The difference between knowledge flow network and knowledge point network is as follows. The paths in knowledge flow network are limited according to knowledge flows. For example, in Figure 3, when  $kp8$  is the current point, the only path is  $\{kp8, kp4, kp2, \dots\}$ , that is to say that the path  $\{kp8, kp4, kp10, \dots\}$  will not be recommended to the learner. Thanks to this limitation, the efficiency will be greatly increased, especially when the amount of points and knowledge flows are very large, for with the growth of joint points the amount of paths increases rapidly without the limitation.

## V. EXPERIMENTS

In this section, two experiments are designed to validate the proposed methods. One is used to evaluate the accuracy of knowledge flows construction, and the other is to evaluate the effectiveness of the helpfulness of the proposed method in improving the foreign language learning.

### A. Experiment on the Construction of Knowledge Flow Network

The knowledge flow of learning resources is the foundation of learning resource recommendation service, thus it is important to build a highly accurate knowledge flow based on the given learning resources. This experiment is designed to evaluate the accuracy of knowledge flows construction method. In this experiment, we select the College English Learning, which is a required course in Chinese university, as the learning task.

#### 1) Experiment Process

a) *Build the learning resources set:* In this experiment, we collect all learning resources related to the knowledge points appearing in Book One of College English and build the learning resources with them. All kinds of learning resources used in the dataset are described in Table II, which consists of not only the source materials within college but also the materials from the Web.

TABLE II. LEARNING RESOURCE DATASET FOR COLLEGE ENGLISH LEARNING

#	Types of Learning Resources	Description
t1	Intensive reading in book	30 articles
t2	Exercises in book	1000 exercises
t3	Test paper in university	1000 exercises
t4	Words in book	2000 words
t5	Phrases in book	100 phrases
t6	Reading material on the Web	100 articles
t7	Test papers on the Web	500 articles

b) *Extract knowledge points from learning resource:* In this step, all materials in the learning resources set are converted to an unified form which is document in our method. As a result, the learning resource set is a document set in fact. Then, the necessary processes of text semantic mining are implemented on the document set, such as text extraction, word segmentation, keywords/terms extraction, and association rule mining. Next, the keywords/terms semantic network is constructed, whose nodes are keywords/terms and whose edges are assoction rules. Based on the keywords/terms semantic network, the concepts are extracted from the network according to the centrality of each node in the network. All the selected concepts form the candidate set of knowledge points.

c) *Optimize the mined knowledge points by domain experts:* Generally, the candidate set of knowledge points can be used as the knowledge points in the next step. However, improving the accuracy of knowledge points will do good to the accuracy of the final knowledge flow network. In this step, the college English teachers are employed as the domain experts to optimize the mined knowledge points. The optimization results are shown in Figure 4.

d) *Mine the association relations among knowledge points:* Each knowledge point is composed of a kernel term and its attributive terms. Each knowledge point can be seen as a transaction and each attributive term of the knowledge point can be seen as an item of the transaction. All knowledge points form the transaction set. As a result, we can use Apriori algorithm to mine the association rules in the transaction set. Furthermore, the association relations among knowledge

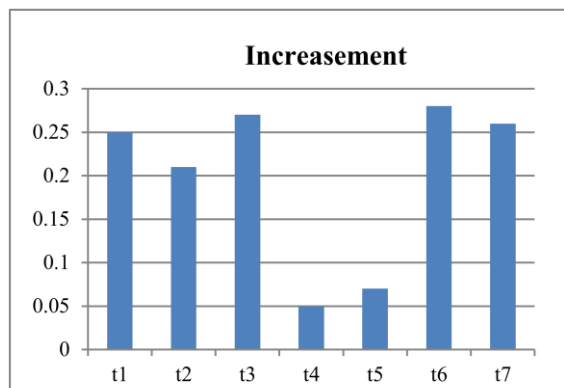


Figure 4. The increase of knowledge points after the optimization.

points can be mined by summarizing the association rules.

e) *Generate the initial knowledge flows:* Connect the knowledge points by the association relations mined in step d) and get the initial knowledge flows. To limit the amount of knowledge flows, a threshold of the weight of the association relation is used to control the association relation to join the construction of knowledge flows. The generated knowledge flows are influenced by the value of threshold, which is shown in Figure 5.

f) *Optimize the initial knowledge flows:* There are some defects in the automatic generated knowledge flows. For example, the lack of association rule will truncate a whole knowledge flow. The threshold used in step e) can also lead to the same question. Some error association rules can lead to the error knowledge flows. In the step, we also use college English teachers as the domain experts to optimize the initial knowledge flows. This step is more difficult than step c). The domain experts check each knowledge flow, revise the error connections in the knowledge flow, split the knowledge flow which is too long, and connect knowledge flows together that should be combined.

g) *Build knowledge flow network:* In this step, knowledge flows are connected by communal knowledge points in knowledge flows. As last, knowledge flows are connected together.

2) *Experiment Results*

Figure 4 shows the effectiveness of knowledge point extraction method. Figure 4 shows the increase of accuracy of knowledge points after they are optimized by domain experts. The results also reflect the accuracy of knowledge point extraction method. According to the figure, the accuracy of knowledge point extraction method is bigger than 0.7 in all seven types of learning resource set.

Figure 5 shows that the amount of knowledge flows is influenced by the threshold of association relation. Obviously, the amount of KFs decreases with the increase of threshold of association relation. When the threshold is bigger than 0.9, all association relations are filtered, which reduces KFs to zero. On the contrary, when the threshold is small, the amount of

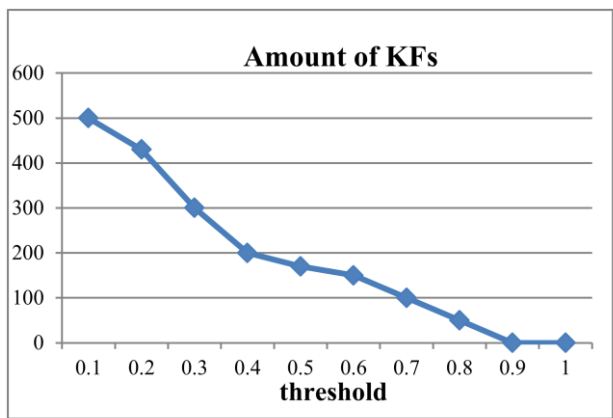


Figure 5. The amount of knowledge flows is influenced by the threshold of association relation.

KFs is big which will increase the complexity of the knowledge flow network and decrease the effective of the learning resources recommendation service. Therefore, the optimal threshold is expected which will be also our future work.

B. *Experiment on the effectiveness of Knowledge Flow-based Foreign Language Learning*

This experiment is designed to evaluate the effectiveness of Knowledge Flow-based Foreign Language Learning. Generally, the planned and ordered learning is more effective than the unplanned and disordered one. We group the participants into two groups. One group learns with the support of knowledge flow and the other group learns at random without the support of knowledge flow. The effectiveness of the proposed method can be got by comparing the scores of two groups.

1) *Experiment Process*

a) *Step 1:* Select 40 non-English major students from different departments in a university. All of the 40 students are divided into two groups randomly: one is experimental group, and the other is control group. Each group has 20 students.

b) *Step 2:* Define five learning tasks and each task consists of five knowledge points. Select five learning resources from the Web for each knowledge point.

c) *Step 3:* Design examination papers for each learning task.

d) *Step 4:* The students of control group will start their learning without any sequence. On the contrary, the students of experimental group start the learning according to the sequence of knowledge flows.

e) *Step 5:* The experiments are repeated on five learning tasks and all the students take three exams. Based on the exam results, the averages of students' scores are calculated and compared.

2) *Experiment Results*

The experimental results are shown in Figure 6. The results show that the experimental student groups concentrate more on knowledge level compared with those of control groups, which means that the knowledge flow-based organization and service

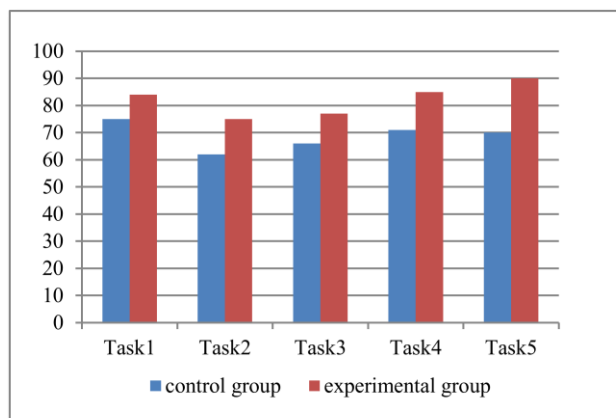


Figure 6. Knowledge flow network of knowledge points

of learning resources can improve the learning efficiency in Foreign Language Learning.

## VI. CONCLUSIONS

To help the foreign language learners to find appropriate learning resources and improve the efficiency of Foreign Language Learning, this paper proposes a new semantic organization of learning resources, knowledge flow-based semantic organization structure, which organizes the learning resources in the form of knowledge flow network and provides learners with learning resources in the form of knowledge flow. Experimental results show that the proposed method works well in improving the utilization rate of learning resources and the efficient of Foreign Language Learning.

The future work is to design a systemic foreign language learning approach to bringing the proposed knowledge flow-based semantic organization structure into full play, so as to improve the efficiency of FLL.

## ACKNOWLEDGEMENT

Research work reported in this paper was supported by the Science Foundation of Shanghai under grant no.16ZR1435500.

## REFERENCES

- [1] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol.18, no.11, pp.613-620, 1975.
- [2] C. T. Meadow, D. H. Kraft, and B. R. Boyce, "Text information retrieval systems," Academic Press, 2007.
- [3] G. Salton, and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol.24, no.5, pp.513-523,1988.
- [4] H. P. Luhn, "The automatic creation of literature abstracts," *IBM Journal of Research & Development*, vol. 2, no.2, pp.159-165,1958.
- [5] R. R. Larson, "Introduction to information retrieval," *Journal of the American Society for Information Science & Technology*, vol.61, no.4, pp.852-853, 2010.
- [6] M. Zorman, V. Podgorelec, and P. Kokol, "Quest for the information: using intelligent search for finding telemedical sites," In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4086-4091,1998.
- [7] N. Vlajic, and H. C. Card, "Categorizing Web pages using modified ART," In *Proceedings of Canadian conference on electrical and computer engineering*, pp.313-316,1998.
- [8] E. I. Papageorgiou, and P. P. Groumpos, "Two-stage learning algorithm for fuzzy cognitive maps," In *Proceedings of 2nd IEEE International Conference on Intelligent System*, vol.1, pp.82-87, 2004.
- [9] C. T. Lin, and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Transactions on Computers*, vol.40, no.12, pp.1320-1336, 1991.
- [10] E. I. Papageorgiou, "Learning algorithms for fuzzy cognitive maps-a review study," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol.42, no.2, pp.150-163, 2012.
- [11] G. Acampora, and V. Loia, "On the temporal granularity in fuzzy cognitive maps," *IEEE Transactions on Fuzzy Systems*, vol.19, no.6, pp. 1040-1057, 2011.
- [12] H. Song, C. Miao, W. Roel, Z. Shen, and F. Catthoor, "Implementation of fuzzy cognitive maps based on fuzzy neural network and application in prediction of time series," *IEEE Transactions on Fuzzy Systems*, vol.18, no.2, pp.233-250, 2010.
- [13] Z. Xu, Y. Liu, L. Mei, C. Hu, and L. Chen, "Generating temporal semantic context of concepts using web search engines," *Journal of Network & Computer Applications*, vol. 43, no. 1, 42-55,2014.
- [14] W. Stach, L. Kurgan, and W. Pedrycz, "Parallel learning of large fuzzy cognitive maps," in *Proc. Int. Joint Conf. Neural Networks*, pp.1584-1589, 2007.
- [15] X. Luo, X. Wei, and J. Zhang, "Guided game-based learning using fuzzy cognitive maps," *IEEE Transactions on Learning Technologies*, vol.3, no.4, pp.344-357, 2010.
- [16] N. Mateou, A. S. Andreou, and C. Stylianou, "A new traversing and execution algorithm for multilayered fuzzy cognitive maps," in *Proc. IEEE Int. Conf. Fuzzy Systems*, pp.2216-2223, 2008.
- [17] H. Zhuge, and Y. Xing, "Probabilistic resource space model for managing resources in Cyber-Physical society," *IEEE Transactions on Service Computing*, vol.5, no.3, pp.404-421, 2012.
- [18] J. Xuan, X. Luo, S. Zhang, et al. "Building hierarchical keyword level association link networks for web events semantic analysis," In *Proceedings of IEEE Ninth International Conference on Dependable Autonomic and Secure Computing*, pp. 987-994, IEEE, 2011.
- [19] X. Luo, Z. Xu, J. Yu, and X. Chen, "Building association link network for semantic link on web resources," *IEEE Transactions on Automation Science and Engineering*, vol.8, no.3, pp. 482-494, 2011.
- [20] X. Luo, J. Ni, J. Zhang, et al. "Building similar link network in large-scale web resources," In *Proceedings of IEEE 16th International Conference on Parallel and Distributed Systems*, pages 87-693, IEEE, 2010.
- [21] X. Wei, and X. Luo, "Concept extraction based on association linked network," In *Proceedings of International Conference on Semantics, Knowledge and Grids*, pp. 42-49, 2010.
- [22] X. Wei, X. Luo, Q. Li, and J. Zhang, "Online comment-based hotel quality automatic assessment using improved fuzzy comprehensive evaluation and fuzzy cognitive map," *IEEE Transactions on Fuzzy Systems*, vol.23, no.1, pp.72-84, 2015.
- [23] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," In *Proceedings of the ACM SIGMOD Conference on Management of Data*, pp.207-216, 1993.
- [24] Z. Xu, X. Wei, X. Luo, Y. Liu, L. Mei, and C. Hu, "Knowle: a semantic link network based system for organizing large scale online news events," *Future Generation Computer Systems*, vol. 43-44, pp.40-50, 2015.
- [25] Z. Xu, Y. Liu, N. Y. Yen, and L. Mei, "Crowdsourcing based description of urban emergency events using social media big data," *IEEE Transactions on Cloud Computing*, 10.1109/TCC.2016.2517638, in press.