Constructing Search as a Service Towards Non-deterministic and Not Validated Resource Environment with a Positive-Negative Strategy

Yucong Duan[1](✉), Lixu Shao[1], Xiaobing Sun[2], Lizhen Cui[3], Donghai Zhu[1], and Zhengyang Song[1]

[1] State Key Laboratory of Marine, Resource Utilization in the South China Sea, College of Information and Technology, Hainan University, Haikou, China
duanyucong@hotmail.com, 751486692@qq.com, 466971642@qq.com, 1464626602@qq.com

[2] School of Information Engineering, Yangzhou University, Yangzhou, China
xbsun@yzu.edu.cn

[3] School of Information Science and Technology, Shandong University, Jinan, China
clz@sdu.edu.cn

Abstract. Internet resources are non-deterministic, non-guaranteed and ultra-complex. We provide a progressive search approach towards problems with positive and negative tendencies aiming at improving the credibility of resources through multi times progressive searching. Meanwhile, we introduce Knowledge Graph as a resource process architecture to organize resources on the network and analyze the tendency of searchers for retrieving information by semantic analysis. We calculate entropy of resources according to searching times and amount of items of each search to represent the reliability of resources with positive and negative tendencies. Resources with ambiguous tendency and false information will be eliminated during the process of progressive search and quality of searching results will be improved while avoiding dead loop of searching towards infinite and complex problems. We apply the searching strategy to a medical resource processing system that provides high precision medical resource retrieval service for medical workers to verify the feasibility of our approach.

Keywords: Bidirectional search · Knowledge graph
Resource modeling · Dynamic equilibrium

1 Introduction

Demand of accuracy for recommended searching results of users is greatly increased. Resources on Internet may be out of time, wrong and fictitious. We propose to improve the quality of resources through performing progressive searches towards problems covering positive and negative tendencies based

I. Romdhani et al. (Eds.): CollaborateCom 2017, LNICST 252, pp. 365–373, 2018.
https://doi.org/10.1007/978-3-030-00916-8_34

in the resource processing architecture. Knowledge Graph is a knowledge base storing both unstructured and structured resources. Knowledge base contains a set of concepts, entities and relationships [4]. Knowledge Graph is a graph representing items, entities and users as nodes and linking nodes through edges. Based on the extension of existing concept of Knowledge Graph, in [9] the authors proposed to express Knowledge Graph in three levels including Data Graph, Information Graph and Knowledge Graph. In [2], the authors clarified the structure of Knowledge Graph according to the progressive relationships among data, information, knowledge and wisdom. In [8], the author proposed to answer 5W (Who/When, Where, What and How) problems by constructing Data Graph, Information Graph and Knowledge Graph. Chaim in [13] elaborated concepts of data, information and knowledge. Data is obtained by observing numbers or other basic individual projects. Information is conveyed through data and data combination. Knowledge is a general understanding and can be used to infer a new background. We introduce Knowledge Graph as a resource processing architecture to organize resources and apply our proposed searching service towards non-deterministic and not validated resource environment with a positive-negative strategy to medical resource management system to provide health care workers with efficient retrieval service.

2 Constructing Search as a Service with a Positive-Negative Strategy

When a user puts forward a noticeable or predisposing problem, there are many non-deterministic, non-fidelity and super-complex resources on the Internet. Faced with the situation, we propose to provide searching service towards non-deterministic and not validated resource environment with a positive-negative strategy. Meanwhile, we apply positive and negative search strategy to medical resource processing system to provide retrieving service for medical workers.

2.1 Investment Allocation of Positive-Negative Searching

According to the description of actual problem searched by searchers, we obtain related resources that satisfy users' information requirements and these resources can be divided into resources with positive and negative tendencies respectively. For example, when a user searches for "Whether changing contact address needs charging?". The positive tendency is that changing contact address needs charging and the negative tendency indicates that changing contact address doesn't need charging. Figure 1 shows progressive search results towards resources with positive and negative tendencies.

We propose to model users' requirements and count weight of searching resources with positive tendency and negative tendency respectively. Weight of resources with positive tendency, denoted as $Weight_p$ and weight of resources

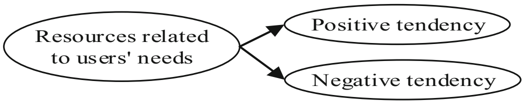


Fig. 1. Resources with positive and negative tendencies

with negative tendency, denoted as $Weight_N$ can be computed according to Eqs. 1 and 2:

$$Weight_P = \frac{Recource_P}{Recource_T} \tag{1}$$

$$Weight_N = \frac{Recource_N}{Recource_T} \tag{2}$$

where $Resource_P$ represents the amount of positive resources and $Resource_N$ represents the amount of negative resources. $Resource_T$ represents the total amount of all resources. We make the hypothesis that expected waiting time of users that denoted as T is already known. Items of each process of progressive search and searching times towards each item are fixed. According to the Eq. 3, we calculate the times of progressive search.

$$S\_times = \frac{T}{S\_items * Pert\_item} \tag{3}$$

where S_times represents times of progressive search determined by users' investment. S_items represents amount of items per search. $Pert\_item$ represents searching time of each item.

2.2 Dynamically Balanced Bidirectional Search Service Provision

For questions proposed by searchers, we conduct dynamically balanced bidirectional search service strategy and traverse resource processing architecture to search relevant resources. According to obtained resources, we find the correlated factors with keywords in resources. We calculate $entropy_{T_K}$ of information which ranges from 0 to 1. The larger $entropy_{T_K}$ is, the lower confidence is. We calculate $entropy_{T_K}$ of resources according to Eq. 4 and total entropy of all resources with the same tendency that is denoted as $Entropy$ according to Eq. 5. Timeliness of resources is calculated according to Eq. 6. We use Eq. 7 to compute confidence of resources based on the entropy of resources obtained through progressive search process.

$$entropy_{T_K} = -\sum_{i=1}^{n} p_i log p_i \tag{4}$$

where $entropy_{T_K}$ represents the entropy of resources obtained in time T_K in accordance with a certain factor in the progressive search, pi represents the probability of each type of answer, $Entropy$ represents the weighted average entropy of resource under different aging when the progressive search is performed according to a certain correlation factor.

$$Entropy = \frac{\sum_k^m Timeliness_{T_k} entropy_{T_k}}{m} \tag{5}$$

$$Timeliness = \frac{T_t - T_s}{T_c - T_s} \tag{6}$$

$$Confidence = S\_amount * Item\_amount * \int_1^n Entropy_{item_i} d(item_i) \tag{7}$$

According to calculated confidence of obtained resources, we make a judgement whether the ratio of positive and negative bidirectional resources confidence is greater than the threshold T or less than $1/T$. T can be obtained according to the learning algorithm. If the threshold condition is satisfied, return search results with high confidence to searchers. Otherwise we return to further allocation of users' investment to continue progressive search. Figure 2 shows the flowchart of constructing search as a service with positive and negative strategy.

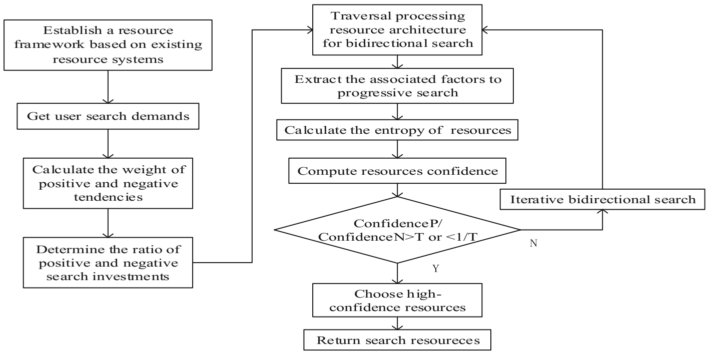


Fig. 2. Flowchart of constructing search with a positive-negative strategy

We clarify that correct resources are unique and divide search resources towards user's requirements into positive and negative tendencies in the non-deterministic, non-fidelity and super complex resource environment. We conduct progressive search strategy towards items related to resources and compute entropy of resources obtained through each progressive search process. Then we sort obtained resources according to their confidence from high to low and ultimately return ordered resources to users.

First of all, we establish a processing resource framework according to existing resource system and access to user search demands and compute weights of positive and negative resources. Secondly, we determine investment allocation according to weights of resources with positive tendency and negative tendency in each progressive search. Then we perform dynamically bidirectional search towards resources with different tendencies. Resources obtained through each progressive search are non-deterministic, not validated and with timeliness. We calculate entropy of resources indicating different tendencies so as to improve the reliability of resources. Resources with high entropy are unreliable. Finally, we determine whether *confidence* of recommended resources is greater than the threshold and recommend resources with higher confidence than the threshold to users.

3 Application of Dynamically Balanced Bidirectional Search Service Provision

3.1 Background of Application

In order to test the accuracy and reliability of recommended resources according to positive and negative bidirectional dynamic balanced search service, we take the background of medical workers' treatment of lung cancer as an example to show rationality of our approach. We apply the strategy to current medical resource processing system to improve the efficiency and accuracy of medical diagnosis and decision-making.

The search engine Bing has recommended 64300000 results when we search resources related to keywords including patients with lung cancer, treatment, chemotherapy. Sources of these resources are different including patients with lung cancer, relevant hospital and some academic institutions. Patients with lung cancer don't need to be responsible for their opinions about treatment of lung cancer released on the Internet which make the credibility of these resources limited. Resources released by relevant hospitals have an obvious tendency for their interests which makes current information more confusing. Information from academic papers is more credible. But because of the timeliness of information, previous information may not be suitable to new patients due to further deterioration of lung cancer. In such a complex resource environment, searchers still need to understand and analyze obtained resources to get reliable and accurate information.

3.2 Example of Dynamically Balanced Bidirectional Search

We obtain relevant resources that match to users' requirements from Bing and allocate investment to positive and negative search respectively. There are 64200000 results through searching keywords including patients with liver cancer, treatment and chemotherapy in Bing. According to statistical resources, there are 31200000 results with positive tendencies and 28800000 results with negative tendencies. We assume that expected waiting time (T) of users is 240 min and time spent on each search is 30 min. We assume that amount of items in each search is 20 and time spent on searching each item is 1.5 min. Table 1 shows the calculations of positive and negative bidirectional search investment and times of each progressive search process.

Table 1. Investment driven progressive search towards different tendencies

	Resource weight	Investments allocation	Search times
Progressive search according to "feasible"	0.48	116.6 min	4
Progressive search according to "infeasible"	0.44	108 min	3

We carry out progressive search covering positive and negative directions after getting resources. Taking progressive search towards positive tendency sources as an example, we make the assumption that item of the first positive progressive search is "patients with lung cancer, treatment, Chemotherapy, cost?". Through traversing processing resource architecture, we obtain resources illustrating that patients with lung canser need to pay $50,000, $100000 and $150000 of which amounts are 32900000, 35600000 and 29300000 respectively. Resources occurring at different times have timeliness and we can calculate the timeliness of corresponding positive resources. Results are shown in Table 2.

Table 2. Calculations of $entropy_{T_k}$ and *timeliness*

T_k	Resource	Failure resources	$Timeliness$	$entropy$
2015	689000 $50000	366000 (2012–2015)	0.40	0.11
	714000 $100000	194000 (2013–2015)	0.50	
	655000 $1500000	300300 (2013–2015)	0.50	
2016	734000 $50000	235000 (2013–2016)	0.75	0.09
	811000 $100000	220100 (2014–2016)	0.25	
	709000 $150000	403000 (2012–2016)	0.20	
2017	1610000 $50000	676000 (2014–2017)	1.00	0.21
	1850000 $100000	445000 (2016–2017)	1.00	
	1530000 $150000	219000 (2012–2017)	1.00	

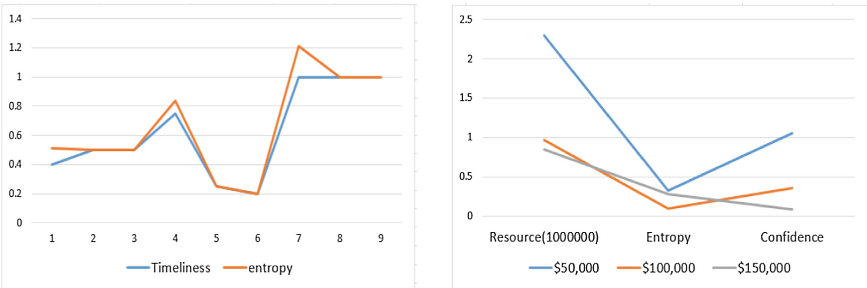
Table 2 shows the entropy of resources at three different time. We calculate *Entropy* of three types of positive resources. It is obvious that the number of progressive search (S_amount) is 1 when we carry out the first positive progressive search. The total number of three types resources of progressive search ($Item\_amount$) are 32900000, 35600000 and 29300000 respectively. We calculate confidence according to Eq. 7 and results we calculated are shown in Table 3.

We sort results according to confidence calculated and recommend results with high confidence degree to searchers, that is paid $50,000. We set the confidence threshold (T) that value is 11000000. The highest confidence of positive search results is 10577350 that it is less than the confidence T. Therefore, we

Table 3. Calculations of *Entropy* and *confidence*

Search results	\$50000	\$100000	\$150000
Entropy	0.3215	0.1	0.283
confidence	10577350	3560000	829190

return a result that searchers should continue to enter the next progressive search to searchers. Search service will be end when carry out positive progressive search at the fourth time. Figure 3 shows the positive progressive search of various indicators of resources calculation. Similar to positive progressive search, we carry out negative progressive search for three times. If the confidence of resources obtained through each progressive search is greater than the threshold we set, we recommend these resources to searchers.



(a) *Timeliness* and *Entropy*

(b) Related indicators of three types of resources

Fig. 3. The calculation of indicators in positive progressive search resources

3.3 Comparison with Traditional Decision Method

Traditional decision-making programs have a low efficiency of resource processing, which make searchers still need to analyze and process constantly to screen some credible resources. For example, for the treatment of patients with lung cancer, searchers use respectively two types of methods to search. Table 4 gives comparison of the positive-negative directions dynamic balanced search strategy and traditional decision-making method for resources obtained on the Internet.

According to experimental results of above examples we provided, we can see that the dynamically balanced bidirectional search strategy towards problems with positive and negative tendencies can improve the confidence of resources through filtering out the ambiguous and unreliable resources. The significance of dynamically balanced bidirectional search strategy towards problems with positive and negative tendencies we proposed is to provide search

Table 4. Comparison of positive-negative search strategy and traditional methods

	Input	Direction	First search results (C: confidence)	Decision
Traditional program	360 min	Any	Chemotherapy (C = 42%) Radiation therapy (C = 51%)	Cannot make decisions
Bidirectional search program	120 min	Positive negative	"Radiation, Therapy, feasible" (C = 64%) "Radiation, Therapy, infeasible" (C = 17%)	Radiation therapy

and decision-making services toward non-deterministic, non-fidelity and super complex resource environment. Due to the massive resources with timeliness in medical domain, it is difficult for health care workers to choose an optimal program with reliable resources and indicators of patients. Therefore, we apply the proposed search strategy towards resources with positive and negative tendencies to medical resource management system in order to provide high-precision resource retrieval service for health care workers.

4 Related Work

With introduction of statistical machine learning methods, coreference resolution technology of knowledge graph application has entered a rapid development stage. Ontology is used as a standard form of knowledge representation in semantic networks [10]. Probase Microsoft used statistical machine learning algorithm to extract the concept of "IsA" [12]. For complex relationships between entities, TBox and ABox are used to simplify and implement relational reasoning based on the reasoning of descriptive logic for consistency checking [5]. For relationship extraction, there are a large number of semi-supervised learning methods [1]. Sen [7] used theme model as basis for similarity calculation and obtained the entity catalog from Wikipedia. Malin et al. in [6] proposed to use random walk model to carry out physical disambiguation of co-operative network data. Wu et al. in [11] chose Wikipedia as a data source to generate training corpus by automatic extraction. Fader et al. in [3] proposed a method of opening questioning and answering (OQA) system that used planned and extracted knowledge base to extract millions of rules from untagged question corpus.

5 Conclusion

Faced with the non-deterministic and not validated resource environment, we propose a dynamically balanced bidirectional search approach to construct search service provision. In the processing of progressive searching, we build a fuzzy vocabulary to filter out useless resources with ambiguous tendency.

Times of progressive searching and amount of items of each search are depended on users' investment. We propose to use entropy of resource sets to indicate the reliability of searching results. Computation of entropy are related to searching times and amount of items of each searching. We validated the rationality of positive-negative search strategy through applying the approach to medical resource processing system aiming at improving the quality of medical service and reducing medical burden of patients. In the future, we will expand the scale of dataset to further verify the feasibility of our proposed approach.

Acknowledgments. 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. Carlson, A., Betteridge, J., Wang, R.C., Hruschka, E.R., Mitchell, T.M.: Coupled semi-supervised learning for information extraction. In: WSDM 2010, pp. 101–110 (2010)
2. 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)
3. Fader, A., Zettlemoyer, L., Etzioni, O.: Open question answering over curated and extracted knowledge bases. In: 20th ACM SIGKDD, pp. 1156–1165 (2014)
4. Lamba, D.S., et al.: Building, maintaining, and using knowledge bases: a report from the trenches. In: ACM SIGMOD, pp. 1209–1220 (2013)
5. Lee, T.W., Lewicki, M.S., Girolami, M., Sejnowski, T.J.: Blind source separation of more sources than mixtures using overcomplete representations. IEEE Signal Process. Lett. **6**(4), 87–90 (1999)
6. Malin, B., Airoldi, E., Carley, K.M.: A network analysis model for disambiguation of names in lists. Comput. Math. Organ. Theory **11**(2), 119–139 (2005)
7. Sen, P.: Collective context-aware topic models for entity disambiguation. In: 21st WWW 2012, pp. 729–738 (2012)
8. Shao, L., Duan, Y., Sun, X., Gao, H.: Answering who/when, what, how, why through constructing data graph, information graph, knowledge graph and wisdom graph. In: SEKE 2017, pp. 1–7 (2017)
9. 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)
10. Vol, N.: Ontology learning from text: methods, evaluation and applications. Comput. Linguist. **32**(4), 569–572 (2005)
11. Wu, F., Weld, D.S.: Autonomously semantifying Wikipedia. In: 16th ACM Conference on Conference on Information and Knowledge Management, pp. 41–50 (2007)
12. Wu, W., Li, H., Wang, H., Zhu, K.Q.: Probase: a probabilistic taxonomy for text understanding. In: ACM SIGMOD 2012, pp. 481–492 (2012)
13. Zins, C.: Conceptual approaches for defining data, information, and knowledge. J. Assoc. Inf. Sci. Technol. **58**(4), 479–493 (2007)