

## Research on Model of Network Information Currency Evaluation Based on Web Semantic Extraction Method<sup>1</sup>

Mo Chen<sup>1,2</sup>, Xiao-Ping Yang<sup>1</sup>, Meng Sun<sup>2</sup> and Yun Zhao<sup>2</sup>

<sup>1</sup>School of Information, Renmin University of China, Beijing 100872, China

<sup>2</sup>Business College of Beijing Union University, Beijing 100025, China

### *Abstract*

*With the arrival of the big data era, it has become a spot for research to evaluate currency of network information so far. This paper proposes a model of network information currency evaluation based on Web semantic extraction method taking Web news as object of study. The author elaborates the method, technology and main functions on every layer of the model in detail, which have been used or completed, and focus on how to extract semantic information efficiently from the contents of Web news, in order to explore a research method for network information currency evaluation. The experimental results show the validity of the model design that plays a very important role in leading network users pay attention to more valuable network information and helping Web site managers build a higher currency Web site.*

**Keywords:** *Web Semantic Extraction; Network Information Currency; Semantic Similarity; Semantic Distance*

### 1. Introduction

The Internet and the field of information technology have entered the era of big data so far [1]. For example, over 300 million users have already registered in sina microblogging, the numbers of updating post contents have also over 100 million [2], the numbers of visiting YouTube that is the largest video site in world have over one billion times every day, Facebook, which is the world's leading social networking site, has over 3.2 billion in the numbers of increased comment every day [3], what's more, the volume of Taobao's day trading and visiting have both over 10 million level, in a word, the data of network are showing a tendency of explosive growth.

In the majority of network data, the size of the Web news has reached the PB level [4], which shows the 4V features of the big data, it is volume, variety, velocity and value [5]. The high currency is an important reached criteria of Web news [6], however, the phenomenon of untimely updating news, more garbage news, unreliable news sources and so on has made network users put forward a large doubt on the Web news currency currently. Therefore, it has become an urgent problem solved to lead network users concern about the more valuable Web news quickly and effectively and help site managers build a higher currency news network.

This paper mainly proposes a model of network information currency evaluation based on Web semantic extraction method taking Web news as object of study, which contains four layers. The author mainly elaborates the research method, application technology and main functions on every layer of the model in detail, which have been used or completed,

---

<sup>1</sup>This work was supported by the National Natural Science Foundation of China under Grant Nos.71271209, the project of science and technology plan of Beijing Education Committee under Grant Nos.KM201311417011, the project of training outstanding talents of Beijing under Grant Nos.2012D005022000013, and the project of Philosophy and Social Science Planning of Beijing Nos. 13JGC090.

and focus on how to extract <Time, Event> two-tuples semantic information innovatively and efficiently from the contents of Web news. This process of research does key contribution for exploring a research method for network information currency evaluation, this experimental results show the validity of the evaluation model design.

## 2. Related Works

In recent years, some scholars have conducted some research of Web semantic extraction method using different theory and method. For example, Zhu Xudong surveyed researches on Web semantic labeling in the literature<sup>[7]</sup> and established a search engine prototype system oriented Deep Web Semantic Extraction. Zhao Liang and others surveyed researches on information personality recommendation method based on Web semantic extraction layer in the literature<sup>[8]</sup>, exported usage document of Web semantic extraction layer and generated Web pages set of personality recommendation. CELIK Duygu and others surveyed researches on semantic extraction method of Web service in the literature<sup>[9]</sup>, proposed Web semantic extraction algorithm, in order to mine similar Web service using the matching process of service quality parameter criterion. Li Kun and others surveyed researches on Web semantic service discovery algorithm based on constraint extraction and structure analysis in the literature<sup>[10]</sup>, proposed the conceptual semantic matching method based on constraint extraction and the algorithm based on structure analysis while matching is failed. Huang Hui and others surveyed researches on automatic model composition based on Web semantic extraction service in the literature<sup>[11]</sup>, analyzed current development of distributed model management and proposed multiple model composition method, those works were related with author's research direction of Web semantic extraction and application.

In recent years, some scholars have also conducted certain research for network information currency evaluation technology. For example, Yuan Min and others surveyed researches on network information currency evaluation indexing mechanism for initiative state in the literature<sup>[12]</sup>, proposed a Web service unit based on analyzing current Web service mechanism, which could accommodate currency evaluation indexing mechanism. Shen Yunfei and others surveyed researches on personality recommendation model of network information pages based on currency evaluation in the literature<sup>[13]</sup>, proposed increasing amount mining algorithm based on currency value quotient. Yang Chaojun surveyed researches on improvement of Web usability with UI navigation based on an Information currency auto-linking method in the literature<sup>[14]</sup>, presented a design method to support information currency auto-linking and UI navigation. Ma Cuichang and others surveyed researches on information behavior method based on currency evaluation of information system in the literature<sup>[15]</sup>, presented to further extract the construction model of the general information currency behavior model to replace the Ellis model, and integrates the construction model of the general information currency behavior with the other key elements of the theoretical system to construct a relatively complete theoretical system, including situational element and usability element. Bai Zhibin and others surveyed researches on design and implementation of automated evaluation system for Web system usability in the literature<sup>[16]</sup>, presented the automated evaluation system oriented to Web system based on the large amount of data of real users behavior, those works were related with author's research direction of evaluating network information currency.

Based on the analysis of the related research on Web semantic extraction method and evaluating network information currency, experts and scholars have studied on two directions respectively, but the research of evaluating network information currency using Web semantic extraction method is less. Therefore, this paper mainly proposes a model of network information currency evaluation based on Web semantic extraction method, in order to explore how to evaluate accurately Web currency of network information.

### **3. Theoretical Principles**

Firstly, this section briefly elaborates social role generated by Web news currency that is research object of this paper before doing research for network information currency evaluation model. Secondly, this section briefly analyzes the effect of Web news semantic extraction in the process of currency evaluation, which plays a basic role for the research of evaluation model.

#### **3.1. Web News Currency**

Whether Web news can report on the latest events timely or not, whether it can arouse the common interest of discussion for network users or not, whether it can reflect the current public focus or not, when the core event occurs in Web news content, when Web news will be published on the web, and what is the effect of Web news caused after releasing it, which are idiographic reflection of Web news currency. The evaluation results of Web news currency generate certain role for social influence of news website.

It can be concluded that Web news currency is mainly affected by three distance of time through analyzing the dissemination process of Web news. The first time distance is from actual occurrence of event source to issuance of Web news, the second time distance is from issuance of Web news to attention of network users for Web news, the third time distance is from attention of network users for Web news to public opinion generated by Web news. In addition, if Web news can embody higher currency, then its speed of dissemination will take on direct proportion with value, its time distance should be shorten among several stages in the dissemination process of Web news, and it will be ensure that Web news should be released at a right time, in order to attract the attention of network users in emotion and enhance the effect of Web news report.

#### **3.2. The Semantic Information of Web News**

Web news groups are study objects of network information currency evaluation, whose instance can be the Web news. Unlike news spreading on newspapers and media, Web news are described by using semi-structured hypertext markup language or unstructured documents, read, reviewed and diffused recurring Internet for users. Although Internet can display Web news by way of pages, it really cannot understand the means what Web news express, and cannot automatically process Web news, which makes it difficult for network information currency evaluation, however, Web semantic can transfer the information that Web news represent to pattern that computer can understand and process [17], which provides a effective data analysis method for network information currency evaluation.

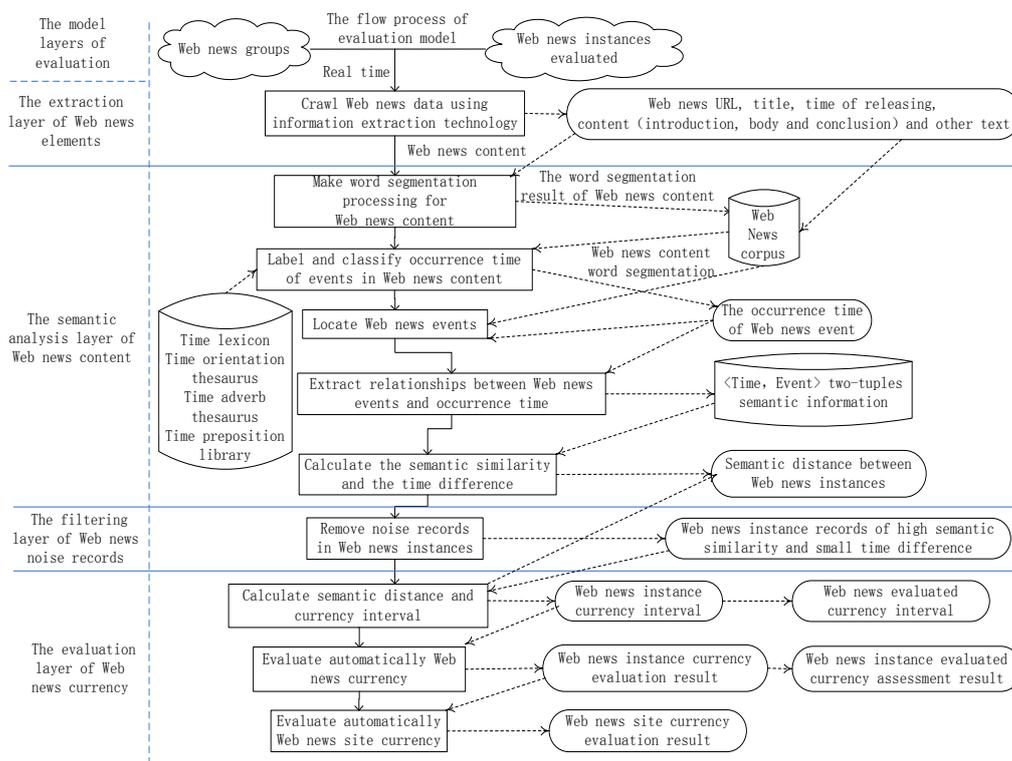
In the Web news instance, its structure mainly includes the news URL, headline, time of releasing, content of guiding, subject, epilogue and many other elements, in the content of guiding, subject and epilogue, the core events and their corresponding occurrence time reported by Web news can combine a two-tuples relation, the extension events arose by the core events and their corresponding occurrence time can also combine several two-tuples relations, which play an important leading role for network users to access theme information of Web news. Therefore, except for extracting numerous elements of Web news, it will be an important assignment in the process of Web news currency evaluation to extract time and event semantic information of Web news contents accurately and perform effective association.

### **4. The Design of Network Information Currency Evaluation Model**

In the era background of big data development, it has become an important research direction of Web text mining to evaluate network information currency through the

process of extracting data, analyzing semantic, filtering noise and so on. Based on the study of two mode, which can perform stream processing and batch processing for big data, this model uses the mode of stream processing for research object of network information according to the characteristic that value of Web news will diminish as time goes by, analyze currency of Web news as soon as possible and come to the results of currency assessment, in order to establish foundation for evaluating currency of network information.

The model of network information currency evaluation based on the Web semantic extraction method taking Web news as object of study is divided into four layers, which include extraction layer of Web news elements, semantic analysis layer of Web news contents, filtering layer of Web news noise records and evaluation layer of Web news currency. As showed in Figure 1, it display flow process and core tasks in every layer of the model.



**Figure 1. The Model of Network Information Currency Evaluation**

#### 4.1. The Extraction Layer of Web News Elements

The extraction layer of Web news elements is mainly responsible for processing the text data structurally, which include the news URL, headline, time of releasing and content contained in the Web news groups or Web news instances. The results extracted are organized into a Web news corpus, which is used for semantic analysis layer of the Web news contents.

In order to improve the extraction accuracy and efficiency of Web news in the design of this layer, this model uses open source library NekoHtml that parses HTML pages, convert the data of Web news pages to plain text format, locate Web news headline in <title> label pertinently through analyzing the organizational structure characteristics of Web news elements, and locate on time of releasing in the next line of Web news headline.

The model uses the corresponding regular expression to determine whether the extracted texts are time information owing to the specification format of Web news

releasing time. Through analyzing the structure of the Web news HTML label including user navigators, floating ads, special theme menus, friendly link embedded in the Web news pages, in addition, Web news content is made up of a number of natural paragraphs, each natural paragraph contains several Chinese punctuations, this model uses the corresponding regular expression to eliminate the disturbance of noise object and determine whether the extracted information are Web news contents.

## 4.2. The Semantic Analysis Layer of Web News Content

As a core layer of evaluation model, the semantic analysis layer of Web news content is mainly responsible for the four tasks based on Web news content extracted in the extraction layer of Web news elements. First of all, this model processes Web news content and store segmentation results in the Web news corpus using Chinese word segmentation algorithm based on Lucene [18]. Secondly, this model marks accurately time words of Chinese word segmentation results in Web news corpus based on common temporal lexicon, temporal orientation thesaurus, temporal adverb thesaurus and temporal preposition library, meanwhile, locate time words and its corresponding event further based on the logical relations of time and event. Thirdly, this model forms a series of the <Time, Event> two-tuples semantic information and store it into a Web news corpus through calibrate the relationship of time and event labeled. Finally, this model calculates the semantic similarity and the time difference among the news instances evaluated and other news to get semantic relations using the <Time, Event> two-tuples of Web news corpus, in order to use for the filtering layer of Web news noise records.

### 4.2.1. The Extraction Process of Web News Semantic Information

In the Web news contents, time is an important property that describes the events, it has played an important guide role in obtaining the news' theme and extracting information contained in the news for network users [19]. This model grasps the development skeleton of core events and related sub events through analyzing the occurrence time of core events and extensive events in Web news contents. Therefore, it will be a great significance for evaluating the Web news currency to mark accurately time word of Web news, locate the events described by the time word, and perform effective association.

The Chinese time has own semantic features comparing with the English time, its labeling process has also a certain difficulty degree, which not only uses the segmentation results of the Web news corpus, but also uses various types of time thesaurus. The algorithm of labeling Chinese time words and storing Web semantic information is shown as follows taking the process of Web news instances as an example.

Algorithm1: Tagging\_Chinese\_time

Input: strUrl, WebNewsDB={r<sub>1</sub>, r<sub>2</sub>, ..., r<sub>i</sub>, r<sub>i+1</sub>, ..., r<sub>n</sub>}.

Output: pubTime, webContent, position.

Tagging\_Chinese\_time(strUrl)

//The object of strUrl stores URL of Web news instance.

Begin

setpubTime(selectPubTime(strUrl));

```
//Extract strUrl news instance time of releasing in Web news corpus and store in
pubTime

attribute.

//This attribute value is assessment time of news instance evaluated.

setwebContent(selectContent(strUrl));

//Extract strUrl news instance Chinese word segmentation content and store in
webContent

attributes.

convert_ArrayList(webContent);

//Store strUrl news instance Chinese word segmentation content with word tagging to
object

of list.

for i=0 to list.size()-1 do
{

//Travel the list, mark the position that time word appears, and append it to the
object of

list position.

if (timeWord(list.get(i))) then
{ position.add(i); }
}

event=Locating_event(position,list);

//Locate events and extract Web semantic information, which is described in Algorithm2.

for n=0 to event.size()-1 do
{

keyWord(event.get(n));

//Do Statistic for the occurrence frequency that event keyword appears in <Time,
Event> two-tuples.
```

```
timeevent.add(event.get(n));

//Append <Time, Event> two-tuples to the object of list timeevent, which is able to
represent the relationship between time and events.

}

//Store <Time, Event> two-tuples semantic information of list timeevent object to Web
news

corpus.

for i=0 to timeevent.size()-1 do
{
str=str+(String)timeevent.get(i)+"\n";

DBConnection.update(con,"update newsinfo set keyOfEvent='"+str+"\where

                                                                    url='"+strUrl+
"");
}

End
```

The logical relationship of time and events is divided into the three categories [20]. Firstly, the incident occurred at the marked time point or time period, secondly, the incident occurred before the marked time point or time period, thirdly, the incident occurred after the marked time point or time period. These logical relationships are primarily determined by the composition of time words and time phrase according to time orientation words and time adverbs. This model uses the principle of minimum distance to locate events based on logical relationship of time and event, use events characteristic words to indicate specific events based on the method of statistical characteristic value. The algorithm of positioning Chinese event characteristic words and extracting Web semantic information is described as follows taking the process of Web news instances as an example.

Algorithm2: Locating\_event

Input: position, list.

Output: event

Locating\_event(position, list)

Begin

```
for j=0 to position.size()-1 do
```

```
//Travel time position labeled and locate events that time corresponds.

{

    int k=(Integer)position.get(j);

    String time_word=(String)list.get(k);

    //Extract time word labeled.

    boolean is=false;

    while (!((String)list.get(k+1)).endsWith("/t") && k<list.size()-2) do

        //Search the corresponding event of precursor time between adjacent words.

        {

            if (((String)list.get(k+1)).endsWith("/n") &&
testChinese((String)list.get(k+1)))

                then

                    //Append event words searched that precursor time corresponds to the object of

                    event_word.

                    {

                        event_word.append((String)list.get(k+1)+" ");

                        is=true;

                    }

                }

            if (!is) then

                { continue; }

            else

                {

                    String final_event_word=event_word.toString();

                    final_event_word=final_event_word.replace("/n", "");

                    event.add(time_word+" "+ final_event_word);

                }

        }

    }

}
```

```
//Append two-tuples to the object of list event through coordinating time and  
corresponding event.  
  
}  
  
}  
  
End
```

After extracting the keywords that represent events, this model combines the marked time and located event, the content of Web news instance is simplified and represented by a series of <Time, Event> two-tuples semantic information, in order to compute Web news semantic similarity and time difference.

#### 4.2.2. The Calculation of Web News Semantic Similarity and Time Difference

According to the number of keywords that represent events, this model adopts the Vector Space Model or string matching method to calculate the semantic similarity of Web news using the <Time, Event> two-tuples set extracted. If the number of keywords that represent events is more than ten, then VSM is used to calculate semantic similarity of Web news, if the number of keywords that represent events is not more than ten, this model cannot provide sufficient dimension to distinguish the relationship among the events, the accuracy of Web news semantic similarity using VSM calculation method is also poor, therefore, this model uses string matching method to calculate the semantic similarity of Web news.

The time difference of Web news is a value among the releasing time of Web news instance evaluated and other related Web news [21]. This model uses the following formula to calculate the time difference of Web news because of the difference in time partition granularity.

$$T = \frac{t_i - \min(t_j)}{\max(t_j) - \min(t_j)} \quad (\text{Fomular 1})$$

As shown in formula 1,  $t_i$  shows the releasing time of Web news instance evaluated,  $\min(t_j)$  shows the minimum of all Web news releasing time related to Web news instances evaluated,  $\max(t_j)$  shows the maximum of all Web news releasing time related to Web news instances evaluated,  $T$  shows the time difference of Web news.

#### 4.3. The Filtering Layer of Web News Noise Records

The filtering layer of Web news noise records is mainly responsible for filtering noise records of lower semantic similarity and larger time difference of Web news based on Web news semantic similarity and time difference calculated, in order to reduce interference of noise data for Web news currency evaluation.

This model uses Fisher judgment method to filter the noise data, this process is mainly divided into the two steps, first of all, this model uses training sample set to get the discriminant, secondly, this model uses it to classify the data and filter noise data sets obtained. Web news records filtered are stored into corpus utilized by the evaluation layer of Web news currency.

#### 4.4. The Evaluation Layer of Web News Currency

The evaluation layer of Web news currency is mainly responsible for three tasks based on Web news records filtered. First of all, this model uses Web news semantic similarity and time difference to calculate semantic distance of Web news instances, secondly, this model infers the currency interval of the Web news instances and Web news evaluated using semantic distance of Web news calculated, thirdly, this model infers the currency evaluation result of Web news instances and Web news evaluated based on the currency interval inferred automatically, which can establish foundation for evaluating currency of network information automatically. This model uses the Euclid Distance formula to calculate semantic distance of Web news instances shown in formula 2 based on the records filtered.

$$D = \sqrt{(1 - S)^2 + \omega T^2} \quad (\text{Formula 2})$$

S shows the Web news semantic similarity, T shows the time difference of Web news,  $\omega$  shows weight coefficient, D shows the semantic distance of Web news instances. It can be seen from the formula that the value of D is smaller, the semantic distance of Web news instance is smaller, and the degree of association among Web news instances is larger.

This model can infer the shortest currency interval of Web news after obtaining the relationship of semantic distance distribution among Web news evaluated and other Web news instances. The number of Web news instances contained in this interval make up eighty percent of total news, and the sum of similarity among news that are in currency interval is up to the largest.

According to the currency interval of Web news inferred, network users not only determine whether Web news instances reflect the current public hot spots, but also define the attention degree of Web news instances from subtler granularity. This model will divide the Web news currency into several levels in the future through determining the position of Web news instance releasing time evaluated in currency interval.

#### 5. The Experiments of Model

The author designs the experimental form of model shown in Figure 2 based on Web news currency evaluation model's design and description of function in each layer. This form use the Matisse Form Class of MyEclipse platform as the top container including several modules [22], the first module has the function of obtaining, selecting, storing, analyzing, filtering and evaluating Web news group instances, the second module has the function of crawling Web news instance evaluated source data, title, releasing time and content, the third module has the function of performing word segmentation for Web news instance content evaluated and word frequency statistic, the fourth module has the function of extracting <Time, Event> two-tuples set and storing Web news instance results evaluated that have been crawled, calculated and analyzed into Web news corpus, the fifth module has the function of generating currency interval and conclusions of Web news instances evaluated.

The screenshot shows a web application interface with a tabbed menu at the top. The selected tab is 'Web News Evaluation'. Below the menu, there are three sub-tabs: 'Web News', 'Web News Instance' (which is active), and 'Web News KeyWord'. The main content area contains several input fields and a button. The fields are labeled 'Web News URL:', 'Web News Source:', 'Web News Title:', 'Web News Time:', and 'Web News Content:'. A 'Web News Crawling' button is positioned between the first and second 'Web News URL:' fields.

**Figure 2. The Experimental Form of Network Information Currency Evaluation Model**

In the process of model experimental design and implementation, the author uses object-oriented programming language and design mode of Model, View, Controller [23]. In the Model, the core functions of each layer are implemented in evaluation model, in the View, the core functions results of each layer are shown using components that can be added to form in evaluation model, the Controller is mainly responsible for converting the processing requests of View to Model. The MVC design mode adopted can logically divide the whole process of model experiment clearly, which can enhance the function expansibility of model experiment, improve the efficiency of management and maintenance for experimental platform, moreover, reflect effective separation of business logic, presentation logic and control logic in model experimental process.

The author expatiates on the process of model experiment and the final evaluation result taking <http://news.163.com/13/0602/11/90C4AT4J00014J34.html> as an object of Web news instance evaluated.

In the crawling form of Web news instance content, the users can input the Web news instance URL that is shown in Figure 3 and click Web News Crawling button. The experimental controller converts the request of crawling Web news instance content received to the processing model, the crawling form of Web news instance content efficiently and accurately displays the result of Web news source data, title, releasing time and content crawled, which is shown in Figure 3.



Figure 3. The Crawling Form of Web News Instance Content

In the analyzing form of Web news instance content, the users can click Web News Divided Words button, the experimental controller converts the request of word segmentation received to the processing model according to Web news text content that has already been crawled in the crawling form of Web news instance content, the analyzing form of Web news instance content efficiently and accurately displays the results of word segmentation. The users can click Web News Words Frequences button, the experimental controller converts the request of word frequency statistics received to the processing model according to the results of word segmentation, the analyzing form of Web news instance content efficiently and accurately displays the results of word frequency statistics, which is shown in Figure 4.



Figure 4. The Analyzing Form of Web News Instance Content

In the extracting form of Web news instance content, the users can click Web News Time And Event button. The experimental controller converts the request of extracting <Time, Event> received to the processing model according to the results of experimental step described, the extracting <Time, Event> form of Web news instance content efficiently and accurately displays the result of <Time, Event> extraction, which is shown in Figure 5.

<b>Web News Content:</b>
【环球网报道】据《印度时报》6月1日报道，印度内阁安全委员会日前表示可能会很快批准印度陆军的建议，将在印度东北部同中国的交界地带集结一支超过4万军人，专用于进攻的山地打击军。消息人士透露，增加这支打击军预计要在第十二个五年计划(2012-2017年)期间耗资约110亿美元。印度国防部预计，财政部不会就这个雄心勃勃的计划进一步提出反对，该计划将在印度东北部同中国的交界地带集结一支专用于进攻的部队。据报道，印度陆军建议在印度东北部同中国的交界地带，增加一个山地打击军、两个独立的步兵旅和两个独立的装甲旅，以填补整个印中实际控制线上的作战漏洞，并具备攻击能力。这支山地打击军将拥有超过4万军人，其总部设在西孟加拉邦的巴纳格爾。它将首次赋予印度在发起袭击的情况下对中国西藏自治区启动攻击行动的能力。该打击军将拥有两个驻扎在高海拔地区的师，以进行快速反应。报道称，在过去十年中，印度开始在军事能力和基础设施能力上追赶中国。现在印正努力同中国保持均势。印度一名高级官员说，这支山地打击军将是赶上中国军力的重要一步。印度还在中国边界扩展其导弹和战斗机能力。急于追赶中国发展迅速的军事能力和发达的基础设施。
<b>Web News Time And Event</b>
6月1日内阁委员会 日前批准山地计划建议 过去能力基础设施军事追赶 现在能力均势

Figure 5. The Extracting <Time, Event> Form of Web News Instance Content

In the evaluation form of Web news instance content, the users can click Web News Time Range and Effect button. The experimental controller converts the request of evaluation received to the processing model according to the results of experimental step described, the evaluation form of Web news instance content efficiently and accurately displays the evaluation result of Web news currency, which is shown in Figure 6.

<b>Web News Time:</b>
2013-06-02 11:38:13
<b>Web News Content:</b>
【环球网报道】据《印度时报》6月1日报道，印度内阁安全委员会日前表示可能会很快批准印度陆军的建议，将在印度东北部同中国的交界地带集结一支超过4万军人，专用于进攻的山地打击军。消息人士透露，增加这支打击军预计要在第十二个五年计划(2012-2017年)期间耗资约110亿美元。印度国防部预计，财政部不会就这个雄心勃勃的计划进一步提出反对，该计划将在印度东北部同中国的交界地带集结一支专用于进攻的部队。据报道，印度陆军建议在印度东北部同中国的交界地带，增加一个山地打击军、两个独立的步兵旅和两个独立的装甲旅，以填补整个印中实际控制线上的作战漏洞，并具备攻击能力。这支山地打击军将拥有超过4万军人，其总部设在西孟加拉邦的巴纳格爾。它将首次赋予印度在发起袭击的情况下对中国西藏自治区启动攻击行动的能力。该打击军将拥有两个驻扎在高海拔地区的师，以进行快速反应。报道称，在过去十年中，印度开始在军事能力和基础设施能力上追赶中国。现在印正努力同中国保持均势。印度一名高级官员说，这支山地打击军将是赶上中国军力的重要一步。印度还在中国边界扩展其导弹和战斗机能力。急于追赶中国发展迅速的军事能力和发达的基础设施。
<b>Web News Time Range and Effect</b>
2013-05-21---2013-06-02
Good!

Figure 6. The Evaluation Form of Web News Instance Content

## 6. Conclusion

This paper completes a research on model of network information currency evaluation based on Web semantic extraction method, which takes Web news as a research object, from the research point of innovation, takes Web news semantic extraction method as a research core and executes the process of Web news data acquirement, data analysis, data filtering and currency evaluation. The model experiment does key contributions for the feasibility of network information currency evaluation request, improving the efficiency of retrieving network information for users, enhancing the availability of websites, building scientifically and improving service functions of websites, and improving business operational efficiency and clicking rate of website. In a word, the process of design and research has certain practical application value.

## References

- [1] P. Suraj and N. Surya, "Cloud Computing and Scientific Applications-Big Data", Scalable Analytics, and Beyond. Future Generation Computer Systems, vol. 29, no. 7, (2013), pp. 1774-1775.
- [2] Z. Xiao and L. Fang, "Mining aspects and opinions from microblog events", Journal of Computational Information Systems, vol. 9, no. 6, (2013), pp. 2399-2400.
- [3] Y. Qiang, W. Lianren and Z. Lan, "Social network based microblog user behavior analysis", Physica A-Statistical Mechanics and its Applications, vol. 392, no. 7, (2013), pp. 1712-1713.
- [4] L. Zhunchen, T. Jintao and W. Ting, "Improving keyphrase extraction from web news by exploiting comments information", Lecture Notes in Computer Science, vol. 7808, (2013), pp. 141-142.

- [5] G. Yike, "Big data, big science, big collaboration: Delivering connected R&D for better value", *Scientific Computing*, vol. 30, no. 2, (2013), pp. 5-6.
- [6] Y. Janghyeok, "Detecting weak signals for long-term business opportunities using text mining of Web news", *Expert Systems with Applications*, vol. 39, no. 16, (2012), pp. 12544-12545.
- [7] Z. Xudong, "Research on Key Issues of Deep Web Semantic Annotation Based on Ontology Learning", Su Zhou: Su Zhou University, (2012).
- [8] Z. Liang and Z. Yujing, "A Design of Personality Information Recommendation Based on Web Semantic Mining", *Computer Knowledge and Technology*, vol. 07, no. 08, (2011), pp. 1731-1732.
- [9] C. Duygu and E. Atilla, "A broker-based semantic agent for discovering Semantic Web services through process similarity matching and equivalence considering quality of service", *Science China Information Sciences*, vol. 56, (2013), pp. 2-3.
- [10] L. Kun and J. Lili, "Research of Semantic Web Service Discovery Algorithm Based on Constraint Extraction and Structure Analysis", *Computer Engineering & Science*, vol. 35, no. 08, (2013), pp. 144-146.
- [11] H. Hui, C. Xueguang and W. Zhiwu, "Survey of Automatic Model Composition Based on Semantic Web Service", *Computer Science*, vol. 40, no. 07, (2013), pp. 9-11.
- [12] Y. Min, H. Zhiqiu and M. Xiaoxiao, "Period-effective Index Mechanism for Initiative and Stateful Web Service", *Application Research of Computers*, vol. 24, no. 07, (2004), pp. 224-225.
- [13] S. Yunfei, S. Guoqiang, J. Lihua and Q. Zheng, "Web Pageviews for Personalized Recommendation Model Based on Time-validity", *Computer Engineering*, vol. 32, no. 13, (2006), pp. 80-81.
- [14] Y. Chaojun, "Improvement of Web Usability with UI Navigation Based on an Information Auto-Linking Method", *Journal of Modern Information*, vol. 29, no. 09, (2009), pp. 55-57.
- [15] M. Cuichang and C. Shujin, "Information Behavior Method Based on Usability Evaluation of Information System", *Information Studies: Theory & Application*, vol. 36, no. 05, (2013), pp. 98-100.
- [16] B. Zhibin, Y. Da and L. Juan, "Design and Implementation of Automated Usability Evaluation System for Web System", *Computer Engineering and Design*, vol. 34, no. 10, (2013), pp. 3649-3652.
- [17] S.-H. Li, S.-M. Huang, D. C. Yen and J.-C. Sun, "Semantic-based transaction model for web service", *Information Systems Frontiers*, vol. 15, no. 2, (2013), pp. 394-396.
- [18] X. Chi, "An analysis of key issues in Chinese word segmentation", *Journal of Computational Information Systems*, vol. 9, no. 3, (2013), pp. 891-893.
- [19] L. Mohan, L. Jianzhong and G., "Evaluation of Data Currency", *Chinese Journal of Computers*, vol. 35, no. 11, (2012), pp. 2349-2350.
- [20] B. Shuotian, N. Yue, Y. Sha and Z. Tingshao, "Predicting reader's emotion on Chinese web news articles", *Lecture Notes in Computer Science*, vol. 7719, (2013), pp. 18-19.
- [21] J. Bo, L. Zhiyua, "A new algorithm for semantic web service matching", *Journal of Software*, vol. 8, no. 2, (2013), pp. 352-353.
- [22] H. Liancai, "Application of MVC platform in bank E-CRM", *International Journal of Service, Science and Technology*, vol. 6, no. 2, (2013), pp. 34-36.
- [23] L. Yong-Fei and C. Zhen-Guo, "Design and implement of news publishing system based on MVC design pattern", *Advances in Intelligent Systems and Computing*, vol. 181, (2013), pp. 756-757.

## Author



Mo Chen received the M.S. degree from China University of Geosciences (Beijing) Information Engineering College in Computer Application Technology. He is a lecturer in the Department of E-commerce at Business College of Beijing Union University. He Engaged in Data Structure, Object-Oriented Programming, Database Theory and Application, Computer Network Technology and Applications, and other courses teaching and research work. His main research interests are in Web Semantic Analysis, Computer Application System Construction and Algorithm Research. He has published papers in the core journals, participated in scientific and teaching research projects.  
Tel: 15011477138  
E-mail: mo.chen@buu.edu.cn