A PARALLEL METHOD FOR METADATA CORRELATION IN SEMANTIC WEBS

Iman Rasekh[1] and Davoud Poor Yousefian Barfeh [2]

[1] Institute of computer science, University of Philippines at Los-Banos,
iman.rasekh@gmail.com
[2] College of Computer Studies ,Lyceum of the Philippines University (LPU)
davoodpb@yahoo.com

ABSTRACT

Parallel computing can makes programs run faster because there are more processors running it. It can wildly use in next generation of web which is known as Semantic Web. Semantic Web which is known as a new collaborative movement toward Web3.0 that led by the World Wide Web Consortium; aims at converting the current web of unstructured documents into a "web of data" by using metadata. Metadata facilitates machines to understand the semantics, or meaning, of information on the World Wide Web. In this paper I tried to implement a parallel broadcast – reduction method for implementing metadata in semantic webs.

KEYWORDS

Web 3.0, Semantic Webs; Multi processor systems; Metadata; broadcast; reduction; parallel processing

1. INTRODUCTION

Metadata is a data about data content that is being used in Semantic Web. Current web pages include metadata specifying what language it's written in, what tools were used to create it, and where to go for more on the subject, allowing browsers to automatically improve the experience of users. Metadata in Semantic webs contains more information than current web pages. The machine should know how to analyse data. One of the most important parts of data analysing in semantic webs is to find information about the all anchor nodes that exist in search graph. In the proposed method in this article two different algorithms are used to achieve this goal; a sequential algorithm and a parallel algorithm

The rest of the paper is organized as follows: Section 2 introduces the Semantic web knowledge graph. In Section 3 Communication operations in parallel processing are discussed and in section 4 our metadata is implemented; finally, Conclusions and remarks are explained in Section 5.

2. SEMANTIC WEB KNOWLEDGE GRAPH

The Semantic Web is a major research initiative of the World Wide Web Consortium to create a metadata-rich web of resources that can describe themselves not only by how they should be displayed or syntactically , but also by the meaning of the metadata[1]. The main purpose of the Semantic Web is driving the evolution of the current Web by enabling users to find, share, and combine information more easily. A machines cannot accomplish search tasks without human direction, because web pages are designed to be read by people, not machines. The semantic web is a vision of information that can be readily interpreted by machines, so machines can perform more of the tedious work involved in finding, combining, and acting upon information on the web [2].

2.1 Knowledge Graph

The Knowledge Graph is a knowledge base source which is being used by Google to enhance its search engine's search results with semantic-search [3]. It provides structured and detailed information about the topic in addition to a list of links to other sites. The goal is that users would be able to use this information to resolve their query without having to navigate to other sites and assemble the information themselves [4][5].

In Semantic webs knowledge graph define by the term RDF. RDF is the simplest of the semantic mark-up languages and actually began life independently of and at about the same time as XML, RDF has three object types: first Resources, which is anything being described in an RDF document: second, Properties; which are used to qualify resources and Statements; which are combination of an RDF resource with a specific property RDF statements are also called triples.

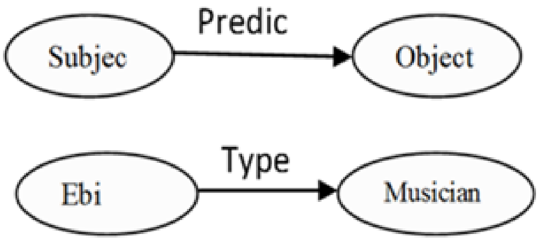


Figure. 1. Examples of RDF graphs

RDFS2 which is defined as a higher level of RDF; provides mechanisms for developing custom RDF vocabularies that are appropriate for a particular field. RDFS allows specific resources to be described as instances of more general classes.

A knowledge graph in semantic web is defined as a group of one or more RDF schemes [6].

A. Dynamic Searching strategy in semantic web

Search Engine Optimization (SEO) is a fundamental concept in Semantic Webs and refers to the collection of techniques and practices that allow a site to get more traffic from search engines. The search engines usually are smart enough to award you that rank by default. Search engines have become more and more popular on the web, nearly anyone trying to get seen on the web can benefit from a little SEO loving [3].Recently, folksonomy and ranking strategy based on links are getting popular in Semantic Web Search Engine Optimization and spreading widely. Folksonomy is one of the components of Web 2.0 which is extended to Semantic Web page ranking is also an important approach in web search strategy.

2.1 Page Rank - PR (E)

Page Rank is an algorithm to rank websites in their search engine results. It works by counting the number and quality of links to a page to determine how important the website is. More important websites receive more links from other websites. The numerical weight that it assigns to any given element E is referred to as the Page Rank of E and denoted by PR (E). The Page Rank of a page is the probability of arriving at that page after a large number of clicks. This happens to equal t-1 where t is the expectation of the number of clicks. One main disadvantage of Page Rank is that it favours older pages. A new page, even a very good one, will not have many links unless it is part of an existing site [4].

$$\text{Page Rank} \propto 1/(\text{The Number of clicks}) \tag{1}$$

The Page Rank assume a probability distribution between 0 and 1 as the initial value for each page. The Page Rank value for a page **u** is dependent on the Page Rank values for each page **v** contained in the set $\mathbf{B_u}$ (the set containing all pages linking to page **u**), divided by the number $L(v)$ of links from page **v**. In other words, the Page Rank is equal to the document's own Page Rank score divided by the number of outbound links L.

$$PR(u) = \sum_{v\epsilon B_u} \frac{PR(v)}{L(v)} \tag{2}$$

Damping factor is defined as the probability, at any step, that the person will continue is a damping factor *d*. The damping factor will be set around 0.85[5]. The damping factor is subtracted from 1 and this term is then added to the product of the damping factor and the sum of the incoming Page Rank scores. The result is divided by the number of documents (N) that is:

$$PR(u) = \frac{1-d}{n} + d\sum_{v\epsilon B_u}\frac{PR(v)}{L(v)} \quad (3)$$

To more simplify it we can write the PageRank of the web page i as

$$PR_i = \frac{(1-d)}{n} + d\sum_{j=1}^{n}\frac{PR_j}{L_{ji}} \quad (4)$$

In which L_{ji} is the number of out links from page *j* when page *j* is linked to page *i*.. I noticed that I should consider the historical behaviour of all users. I calculate L_{ji} by divide the number of clicks from page *j* to page *i* when *j* is linked to *i*.

$$c_{ji} = \sum_{u\in U} c_{ji}(u) \quad (5)$$

$$L_{ji} = \frac{\sum_{j=1}^{n} C_{ji}}{C_{ji}} \quad (6)$$

In which $C_{ji}(u)$ is the number of clicks of a user *u* from page *j* to *i*, while *U* is a set of all users.

To consider *User's Preference* to a web page and support personalized search. It is necessary to count the number of a user's clicks in all pages that have linked to page i over the number of a user's clicks in web pages that have linked to all the pages. Equation (7) describes its definition.

$$PE_i(u) = \frac{\sum_{j=1}^{n} C_{ji}(u)}{\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ji}(u)} \quad (7)$$

Finally, Equation (8) is my modified PageRank algorithm for the proposed system [6].

$$PR_i = PR(u_i) = (1-d)PE_i + d\sum_{j=1}^{n}\frac{PR_j}{L_{ji}} \quad (8)$$

2.2 Heuristic Method

Heuristic Method is a Method to improve the above Three step method, in this method we choose M, RDF triples that have a common source and edge labels, M is a factor based on bushiness of anchor nodes or based on a distance Function of node from the anchor node, Finally we choose M Triple with common source. Heuristic Method causes a priority for nodes that are closer to anchor node. The possible problem in heuristic method is when the user's query include two or more anchor nodes, In this case we can find some path between anchor nodes rather than adjacency sub graphs of anchor nodes and then add adjacency sub graphs of this paths to our output and finally RDF triples are sorted based on sources and neighbourhood with anchor node and ultimately based on cost of edges.

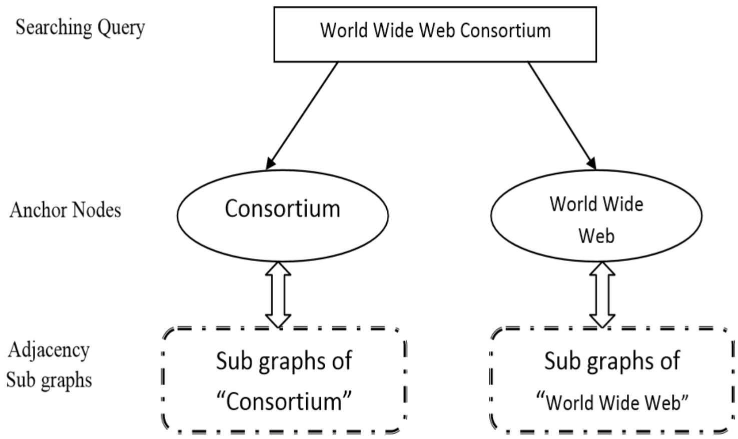


Figure 1. Three steps of Dynamic Searching strategy for the phrase "World Wide Web Consortium"

2.3Metadata

Metadata (Meta content) defines as "data about data" or "data about the containers of data" or "data about data content" or "content about content". The National Information Standards Organization accepted all the above definitions. Metadata was the contents and context of data files. Current web pages include metadata specifying what language it's written in, what tools were used to create it, and where to go for more on the subject, allowing browsers to automatically improve the experience of users. Metadata in Semantic webs contains more information than current web pages. The machine should know how to analyse data. One of the most important parts of data analysing in semantic webs is to find information about the all anchor nodes that exist in search graph for these reason each anchor node in semantic should guess the adjacency matrix of the whole search graph . In section four I try to use a parallel algorithm to find the adjacency matrix.

3. COMPUTATIONAL OPERATIONS IN PARALLEL PROCESSING

True concurrency happens in Multi-processor systems. In these systems several tasks can be executed at the same time by using a Conventional architecture. Parallelism addresses each of its components in significant ways. Different applications utilize different aspects of parallelism. Data intensive applications utilize high aggregate throughput, server applications utilize high aggregate network bandwidth, and scientific applications typically utilize high processing and memory system performance. Many interactions in practical parallel programs occur in well-defined patterns involving groups of processors. The programmer should define efficient implementations of these operations in order to improve performance, reduce development effort and cost, and improve software quality. There are several methods for broadcasting and reducing information between processors [11]. Here I'm introducing two important methods that I used in my proposed method; One-to all broadcast and All-to-One Reduction.

3.1. One-to-All Broadcast and All-to-One Reduction

In one-to-all broadcast One processor has a piece of data (of size m) it needs to send to everyone. The dual of one-to-all broadcast is all-to-one reduction. In all-to-one reduction, each processor has m units of data. These data items must be combined piece-wise (using some associative operator, such as addition or min), and the result made available at a target processor. Figure 2 shows the general form of reduction and broadcast.

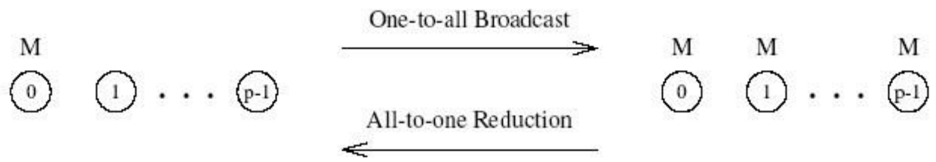


Figure 2. The general form of reduction and broadcast

3.2 Performance Metrics in multiprocessors

The most important Performance Metrics in parallel processing are speedup, Efficiency and cost that define as follows; Speedup (S) is the ratio of the time taken to solve a problem on a single Efficiency is a measure of the fraction of time for which a processing element is usefully employed mathematically, it is given by

$$E = \frac{S}{P} \tag{9}$$

Cost is the product of parallel runtime and the number of processing elements used ($p. T_P$). Cost reflects the sum of the time that each processing element spends solving the problem [11].

3. METADATA IMPLEMENTATION

Our goal is to declare a parallel method for metadata correlation in Semantic Webs; in fact we are trying to find information about all the nodes that exist in search graph then get the default Page Rank of each node and send the information to all other nodes so the real Page Rank of each page can be calculated easily and replace the default page rank.

To achieve this goal each we need to find the Adjacency matrix of each unknown knowledge graph of semantic web, the Page rank of each anchor node should be known. The adjacency matrix of a finite graph G on n vertices is the n × n matrix where the no diagonal entry aij is the number of edges from vertex i to vertex j, and the diagonal entry aii, depending on the convention. There exists a unique adjacency matrix for each isomorphism class of graphs (up to permuting rows and columns), and it is not the adjacency matrix of any other isomorphism class of graphs. But our specific adjacency matrix for Semantic Webs knowledge graph is infinite we and at first the graph should be discovered and that's our biggest challenge!

3.1 Sequential Algorithm

Algorithm 1 declares a pseudo-code for a sequential algorithm for discovering the adjacency matrix of an infinite graph. The input of the algorithm is a vector of known predefined Page Ranks and the function NN as a predefined function can find the neighbor nodes of each node. In this algorithm we defined InVector as the vector of known vertices with predefined Page Ranks and Transferred as vector of vertices that have been checked and already exists in the Adjacency Matrix. This algorithm for each node in input sequence find the neighbour nodes and put them in a temporary array and then it checks all the nodes in the temporary array if any of them does not exist in transferred array then the node will add to it and placed in the correct position the page rank will be calculated based on equation (8).

Algorithm 1 Sequential Algorithm

```
Input (The input vertices) :
 InVector= {u1, u2, u3}  // vector of known vertices with
 predefined  Page Ranks
Main Sequential code:
i=0;        // for U indices
j=0;       // indices for temp
k=0;
t=0;      // indices for transferred
Transferred [];  //A vector of vertices that have been
checked
Temp [];  //A vector for temporary vertices

Step 1-Form the  Adjacency Matrix:
Adj[][] ;
Adj[0][0]=0 ;

            For each vertex in InVector do
            {
             Transferred [t] = InVector [i];
            Call the function:
            Temp[j] =NN (InVector [i]);
             For each j in temp {
             If (temp[j] exists in Transferred) then  Remove
             Temp[j];
             K= length (InVector)+1;
             InVector [k] =Temp[j];
             }
Assign values to Adjacency matrix:
                        For each j {
                        Adj[j][k]=1;
                        }
                        t++;
            }//End for
Step 2- calculate the page rank based equation (8)
```

Figure 3. The sequential algorithm

3.2 Time complexity of sequential algorithm

The main challenge of Sequential algorithm is its time complexity; we aim to calculate the Page Rank as faster as possible .in this algorithm the external loop repeats *n* times, where n is the maximum possible number of nodes, each of internal loops repeats m times where m is the maximum degree of nodes and K is the time complexity of NN function so the total time is:

$$T\ (n)\ =\ n\ (k\ +\ m\ +\ m)\ =\ n\ (k\ +\ 2m) \tag{10}$$

So our time complexity is

$$T\ (n)\ =\ O\ (mnk) \tag{11}$$

If $(\mathrm{k} + 2\mathrm{m}) << n$ then

$$T(n)\ =\ O(n) \tag{12}$$

So with a sequential algorithm we can do the algorithm in polynomial time.

3.3 Parallel Algorithm

A parallel algorithm can help to achieve a better time complexity. In parallel model we have almost the same algorithm; the only difference is that *function* NN is call in parallel in each call. The parallel system implemented with Send & Received method ; in which each element of the $InVector$ will send to separate processors then NN will be done in parallel and at last the main processor will collect the elements of the Adjacency Matrix and calculate the page ranks.

Algorithm 2 : Parallel algorithm

Input (The input vertices) :
InVector= {u1, u2, u3}
Main paralell code:
i=0; **// for U indices**
j=0; **// indices for temp**
k=0;
t=0; **// indices for transferred**
Transferred []; **//A vector of vertices that have been checked**
Temp []; **//A vector for temporary vertices**

Step 1-Form the Adjacency Matrix:
Adj[][] ;
Adj[0][0]=0 ;

For each InVector[i] in (length (InVector)/p) step p
//p is the number of processors

Send Procedure:
Send InVector[i] to the processor $P_{n;}$
Transferred[t] = InVector[i];

In parallel:
Temp[j] =NN (InVector [i]);
Receive Procedure:
Get data (temp arrays) from each secondary processor
Loop1(For each temp from each of processors):

```
        For each temp
         processors thus p temp arrays exist
        Loop2:
                For each element j  {
                If (temp[j] exists in Transferred)
                Remove Temp[j];
                K= length (InVector) +1;
                InVector [k] =Temp[j];
                }
        Loop3(Assign values to Adjacency matrix):
                For each j {
                Adj[j][k]=1; }
}

Step 2- calculate the page rank based on equation (8)
```

Figure 4. The parallel algorithm

3.4 Time complexity of the parallel algorithm

In the main processor the external loop repeats $\frac{n}{p}$ times, where n is the maximum possible number of nodes.

The *send procedure* repeats *log p times* (*a broadcast algorithm*). The time complexity of *Get data* is also $log\, p$ (a reduction algorithm). The time complexity of loop1 is P and for loop2 and loop3 is m and K is the time complexity of NN function so the total time complexity is:

$$T(n) = \frac{n}{p}(\log p + logp + p(m+m) + k) = \frac{n}{p}(2\log p \ + 2pm + k) \quad (13)$$

We know that $p << n$ and also p << m so

$$T(n) \ = \ 2n \ + \ 2mn \ + \ nk \quad (14)$$

$$T(n) \ = \ O(n(m \ + \ k)) \quad (15)$$

As you can see the time complexity is better than the sequential method. so the Page Rank can be calculated faster than the sequential algorithm.

4. CONCLUSIONS

In this research we use two algorithms to find the exact connection structure of the searching graph in order to calculate page rank faster than before but in semantic webs our graph is not finite we should discover the graph and that’s our biggest challenge to solve this problem. In this paper we used two algorithm sequential algorithm and a parallel algorithm for find the adjacency matrix of an unknown knowledge graph of the semantic web. We calculated the time complexity; Page Rank and all performance metrics including speedup (Equation 16); efficiency (Equation 17) and cost(Equation 18) for both parallel and serial algorithms. Table 1 show performance metrics factors for parallel method based on the number of nodes and processors.

$$S = \frac{T_1}{T_p} = \frac{mnk}{n(m+k)} = \frac{mk}{(m+k)} \quad (16)$$

$$E = \frac{S}{P} = \frac{mk}{p(m+k)} \quad (17)$$

$$C = p.T_p = p.n.(m + k) \quad (18)$$

Table I. Efficiency of parallel method

N	S	E	C
1000	227.2727	0.018182	1833333.3
2000	454.5455	0.018182	7333333.3
3000	681.8182	0.018182	16500000
4000	909.0909	0.018182	29333333
5000	1136.364	0.018182	45833333
6000	1363.636	0.018182	66000000
7000	1590.909	0.018182	89833333
8000	1818.182	0.018182	117333333
9000	2045.455	0.018182	148500000
10000	2272.727	0.018182	183333333
11000	2500	0.018182	221833333
12000	2727.273	0.018182	264000000
13000	2954.545	0.018182	309833333

As you can see we can achieve a better speedup and decrease the cost by using parallel algorithm. Table 1 shows the advantages of using parallel method, in this table we considered k=n/5, m=n/6 and p=5.

REFERENCES

[1] [1] F. Manola and E. Miller (eds), “RDF Primer.” W3C Working Draft 23 January 2003.

[2] Berners-Lee, Tim; James Hendler; Ora Lassila (May 17, 2001). "The Semantic Web". Scientific American Magazine. Retrieved March 26, 2008.

[3] Singhal, Amit (May 16, 2012). "Introducing the Knowledge Graph: Things, Not Strings". Official Blog (of Google). Retrieved May 18, 2012.

[4] Waters, Richard (May 16, 2012). "Google To Unveil Search Results Overhaul". Financial Times. Retrieved May 16, 2012.

[5] Staff (December 4, 2012). "Get smarter answers from the

Knowledge Graph from Português to 日本語 to русский".

Google. Retrieved December 4, 2012.

[6] D. Brickley and R. V. Guha (eds), “RDF Vocabulary Description

Language 1.0: RDF Schema.” W3C Working Draft 23 January 2003.

[7] Gibbons, Kevin. "Do, Know, Go: How to Create Content at Each Stage of the Buying Cycle". Search Engine Watch. Retrieved 24 May 2014.

[8] Moore, Ross. "Connectivity servers". Cambridge University Press. Retrieved 24 May 2014

[9] Gyöngyi, Zoltán; Berkhin, Pavel; Garcia-Molina, Hector; Pedersen, Jan (2006), "Link spam detection based on mass estimation", Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB '06, Seoul, Korea), pp. 439–450.

[10] Page, Larry, "PageRank: Bringing Order to the Web" at the Wayback Machine (archived May 6, 2002), Stanford Digital

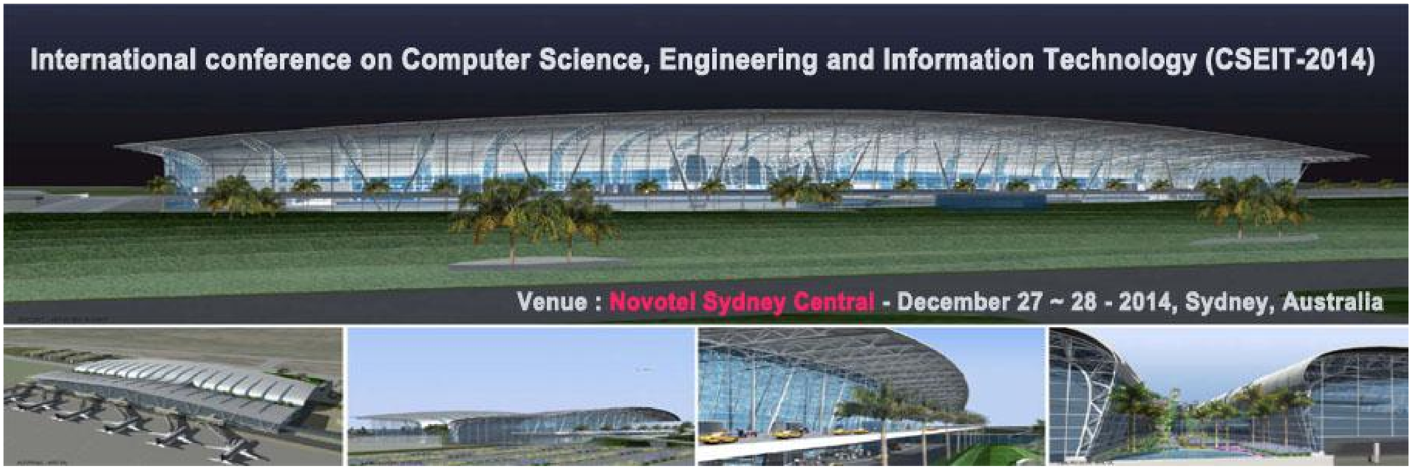
Library Project, talk. August 18, 1997 (archived 2002)

[11] Computer Chronicles: Parallel Processing (1986) .

Authors

Iman Rasekh a PhD student of computer science (Minor in Mathematics) in University of Philippines at los-Banos (UPLB) which is an English based university in the Philippines, He graduated from Master's program in Computer Engineering on July 2010, with thesis on "Design of a Farsi text to speech system using Morphs Database and Neural Networks". He published 4 papers in different journals out of his thesis He also published paper about Ecotourism in Persian. Additionally, Right now UPLB he is working on his dissertation in Semantic Webs (Search Engine optimization for semantic webs- that will be implemented on a bioinformatics database) and he has already published three papers out of his dissertation; He is expected to finish it before May 2015

Home Contact



Home Paper Submission Program Committee Accepted papers Contact us

Accepted Papers

Home

Paper Submission

Program Committee

Accepted papers

Contact us

Use of Eigenvalues and Eigenvectors to Analyze Bipartivity of Network Graphs

Natarajan Meghanathan, Jackson State University, United States

ABSTRACT

This paper presents the applications of Eigenvalues and Eigenvectors (as part of spectral decomposition) to analyze the bipartivity index of graphs as well as to predict the set of vertices that will constitute the two partitions of graphs that are truly bipartite and those that are close to being bipartite. Though the largest eigenvalue and the corresponding eigenvector (called the principal eigenvalue and principal eigenvector) are typically used in the spectral analysis of network graphs, we show that the smallest eigenvalue and the smallest eigenvector (called the bipartite eigenvalue and the bipartite eigenvector) could be used to predict the bipartite partitions of network graphs. For each of the predictions, we hypothesize an expected partition for the input graph and compare that with the predicted partitions. We also analyze the impact of the number of frustrated edges (edges connecting the vertices within a partition) and their location across the two partitions on the bipartivity index. We observe that for a given number of frustrated edges, if the frustrated edges are located in the larger of the two partitions of the bipartite graph (rather than the smaller of the two partitions or equally distributed across the two partitions), the bipartivity index is likely to be relatively larger.

Toward Object Interaction Mining By Starting With Object Extraction Based On Pattern Learning Method

Mujiono Sadikin[1]and Ito Wasito[2], [1]Universitas Mercu Buana, Indonesia and [2]Universitas Indonesia, Indonesia

ABSTRACT

This paper presents a global framework of object interaction mining of large corpora. Motivated by the fact that objects contained in unstructured format document interact each other semantically, this study proposes a method to extract those objects and their interactions. This global framework is started with an initial step approach to extract a single object in the unstructured format document. In this study, this initial step is a pattern learning method which is applied to drug-label document to extract its drug-name. The pattern learning is performed by utilizing an existing external knowledge, identified regular expression surrounding the object target, and the probabilities of those regular expressions. The results of this study show that the best performance of pattern learning achieved is 0,78 f-score. By adjusting some parameters and or improving the approaching methods, the performance has a potentiality to be improved.

Clustering of Multi Script Documents Using k-Means Algorithm

Neeru Garg[1] and Punjab Munish Kumar[2], [1]Lovely Professional University, India and [2]Punjab University Rural Centre, India.

ABSTRACT

This paper aims at the script identification problem of handwritten text document which facilitates the clustering of data according to their type of script. In this paper, collection of different types of handwritten text document i.e. Devanagari, Gurumukhi and Roman is taken as input and then cluster all these documents according to script type whether i.e. Gurumukhi, Devanagari or Roman. Clustering of handwritten multi-script document scheme proposed in this paper is divided into two phases. First phase used to extract the features of given text images. We have extracted four types of features, namely, circular curvature feature, horizontal stroke density feature, pixel density feature value and zoning based feature. In the second phase, features extracted in the previous phase, used for clustering with k-Means algorithm. In this study, we have considered 4850 samples of isolated characters of Devanagari, Gurumukhi and Roman script.

Differentiation and Integration using MAXIMA

Savaş Tuylu, Nigerian Turkish Nile University, Nigeria. Narsee Monjee Institute of Management Studies,India

ABSTRACT

The methods of calculus lie at the heart of the physical sciences and engineering. Maxima can help you make faster progress, if you are just learning calculus. The examples in this research paper will offer an opportunity to see some Maxima tools in the context of simple examples, but you will likely be thinking about much harder problems you want to solve as you see these tools used here.

This research paper includes Differiantiation and Integration. After examples of using each topic some exercises have been formulated.

Short-term Forecast of Blast Furnace Gas Production Amount Based on Grey RBF Neural Network

Zhimin Lv, Zhao Wang, Ziyang Wang and Juhua An, University of Sci. and Tech, Beijing.

ABSTRACT

Dynamic optimization scheduling of the gas in iron and steel enterprises has great significance to reduce gas emission and the short-term forecast is the premise to realize the energy dynamic scheduling. Based on the characteristics that the influencing factors of blast furnace gas amount are complex and difficult to collect, a grey radial basis function (RBF) neural network forecast model is proposed to predict the gas amount for blast furnace in this paper. Combining grey theory, which is used to preprocess the historical data and obtain abundant information, with RBF neural network makes the effective trend forecast in the next 30 minutes come true. The model using in this paper is proved more accurate by doing contract experiment against the grey BP neural network.

New Run Time Adaptations For Cloud-Based Applications

Usman Hamza, College of Science and Technology, Nigeria

ABSTRACT

Cloud-based development methodology embedded on virtual machine offered by cloud utility providers are becoming increasingly powerful, and becoming more the ecosystem of cloud services. Virtual computing service tool that allow run time adaptation and elasticity which is part of cloud computing, that enable customers to acquire and release computing resources. Run time adaptation is one of the main aims of new cloud-based application Development methodology in this research work that allow users to make changes at run time and help developers to reuse software components so that they can plug-in third party components into their application. This research concentrates on this type of adaptation but in the context of Cloud-based services. Cloud-based services are becoming increasingly popular. Web Services often fit the requirements of being a tool or component, and can be reused in a very similar manner. Hence there is a requirement for adaptation of Web Services just as there is the need for adaptation of new cloud-based application development methodology. There are now quite a few adaptation techniques, but few of them have been identified as adaptation techniques for cloud-based services.This approach to adaptation allows for the changes of web services from one web service to another. The design approach to adaptation uses a tool to navigate within different types of web service using URLs and massage request. These messages are commonly in XML format, hence WSDL can be used to modify them.

Developing a Mobile Survey Application on a Tablet PC for a Reasearch Project

Anandh Parathnandh[1], Jayraj Ramota[2], Shailendra Singh[3], [1]University of Kwa-Zulu Natal, South Africa, [2]School of Health Sciences, South Africa and [3]Nelson R School of Medicine, South Africa.

ABSTRACT

Studies have shown that wireless technology has played a vital role in the attitude towards, and achievement of, learning (Swan, K, et al, 2005). These studies, however, focus on situations where the PC is used as an interactive tool rather than one of revision, such as the ZOOM project at the University of KwaZulu-Natal. The Classroom presenter is one such example (http://classroompresenter.cs.washington.edu).The ZOOM project seeks to provide students with the opportunities to keep in contact with their teachers/lecturers, as well as to view missed lessons/lectures. It serves as a repository of information and a revision tool for students. Aimed primarily at Health Sciences students, who are regularly involved in offcampus community service programmes and need to be able to contact the university and receive educational material, the project falls under UKZN's policy of being a full contact institution. The University of KwaZulu-Natal is believed to be the first tertiary education institution in South Africa to undertake the venture of using tablet PCs explicitly for educational purposes. UKZN has engaged the services of two companies: Kanthea (a South African based service provider of visual communication), and Disruptive Vision (a UK based service provider). The two companies have collectively created ZOOM, a low cost, software-only based solution that is capable of streaming live video on a large scale. One draw back to this technology is that Internet access is required, as lecture recordings cannot be downloaded due to copyright issues. There still remain questions that need to be asked and answered. Which departments are using ZOOM and which are not, and why or why not? Of the 1100 students, how many use it regularly? How

many do not, and how can problems preventing ubiquitous usage be overcome?

Target-Oriented Generic Fingerprint-Based Molecular Representation

Petr Skoda and David Hoksza, Charles University, Czech Republic.

ABSTRACT

The screening of chemical libraries is an important step in the drug discovery process. The existing chemical libraries contain up to millions of compounds. As the screening at such scale is expensive, the virtual screening is often utilized. There exist several variants of virtual screening and ligand-based virtual screening is one of them. It utilizes the similarity of screened chemical compounds to known compounds. Besides the employed similarity measure, another aspect greatly influencing the performance of ligand-based virtual screening is the chosen chemical compound representation. In this paper, we introduce a fragment-based representation of chemical compounds. Our representation utilizes fragments to represent a compound where each fragment is represented by its physico-chemical descriptors. The representation is highly parametrizable, especially in the area of physico-chemical descriptors selection and application. In order to test the performance of our method, we utilized an existing framework for virtual screening benchmarking. The results show that our method is comparable to the best existing approaches and on some data sets it outperforms them

Extracting of Machining Deformation Influencing Factors of the Flexible Material based on Analytical Hierarchy Process Method

Yaohua Deng[1], Jiayuan Chen[1], XiaLi Liu[1], Qiwen Lu[2], Jiandong Shi[1], Liming Wu[1], [1]Guangdong University of Technology, China and [2]HSBC Software Development (Guangdong) Limited, China.

ABSTRACT

and the overlapping and interrelated of the distortion factors of flexible material ,making processing deformation uncertainty, thus, how to evaluate influence degree of distortion factors during flexible material processing is not only the key to solve the problem that how to compensate of the flexible materials processing deformation but also the key to solve the advanced problem like subsequent deformation compensation prediction model input dimension. Thinking about the uncertainty of the flexible material processing deformation factors, this paper starting from the nature of evaluate problem and understanding of elements of evaluators, putting out evaluation and extraction method of flexible material processing deformation factors which based on analytic hierarchy process, regarding processing deformation factors as extract attribute P and relative importance of any two influence factors as extraction evidence and constructing flexible material processing deformation factors analytic hierarchy process model. The model regard change degree of flexible material processing trajectory influence degree Wp conducted by extract attribute P as evaluation index, the greater range of Wp the greater trajectory processing influence of extract attribute P will be, to achieve the extraction of processing deformation influence factors. Experimental tests show that the precision of processing deformation decision streamlined set extracted by analytic hierarchy process is higher than Pawlak reduction and comentropy reduction with the the prediction error of offset compensation decreased 57.69% and 67.83% on the X direction, 56.37% and 66.88% on the Y direction separately, and faster than the Pawlak reduction whit the training time of Neural network decreased 31.5%. Key words: Information Extracting;Machining Deformation Influencing Factors;Analytical Hierarchy Process

Improved method of correcting AES Keys obtained from coldboot attack

Surya Prakash Mishra[1] and Enna Sachdeva[2], [1]Defence Research And Development Organization, India and [2]YMCA University of Science and Technology, India

ABSTRACT

Extraction of cryptographic keys, passwords and other sensitive information from the memory, has been made possible by the data remanence property of DRAM. According to it,DRAM can retain its data for several seconds to minutes without power[2]. The method of extracting probable cryptographic keys from DRAM ,using cold boot attack, a type of side channel attack, from the observed sets of decayed round sub keys stored at contiguous memory locations, has already been proposed[2] .However, these keys are distorted and need to be corrected before being used for decrypting the encrypted files. Various methods for correcting these distorted keys have also been proposed [2,9]. However, it has not been reported much in literature regarding efficacy of these methods, on distorted data obtained from cold boot attack.This paper contains results and observations of extensive experiments carried out by varying timings of cold rebooting the PC that varies the % of distortion of data and led to the observation that the proposed methods are theoretical in nature and practically not effective (for coldboot attack) as they could correct upto 2% of errors in the round key schedule of 128 and 256 bits AES. In this paper,an improved algorithm has been proposed,for correcting upto 15% of errors in distorted round key schedule,obtained from cold boot attack, as well as on intentionally generated random erroneous round key schedules and the algorithm has been implemented successfully to mount the volumes encrypted by popular disk encryption system 'Truecrypt'.

A new Competitive Intelligence-based strategy for Web Page Search

Iman Rasekh, University of Philippines, Philippines

ABSTRACT

Semantic Web is known as next generation of web it is known as a new collaborative movement toward Web3.0 that led by the World Wide Web Consortium (W3C) .it aims at converting the current web of unstructured