IJCSI

International Journal of Computer Science Issues

Volume 6, Issue 2, November 2009
ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

IJCSI proceedings are currently indexed by:

www.IJCSI.org

EDITORIAL

This is the final volume for the year 2009. Volume 6th demonstrates the solidarity, trust and increasing interest of authors and readers towards the International Journal of Computer Science Issues (IJCSI). IJCSI has indeed established itself as a venue for excellence in the area of scientific publications and dissemination of knowledge. With its open access policy, IJCSI will continue to serve the scientific community in its utmost capacity. As mentioned in earlier edition, the quality of a journal is mostly dependent on the quality of papers it received and published. IJCSI has shown its level of quality through articles that it has published so far.

We would like to thank enormously all reviewers of IJCSI who have been servicing the journal so far. IJCSI will always maintain its activity of sending all corresponding authors a free print copy of the journal.

We are pleased to present IJCSI Volume Sixth split in two issues (IJCSI Vol. 6, Issue 2). The statistics for this volume is as follows. This volume attracted a total of 39 papers out of which 13 papers have been accepted with minor revisions, 23 were sent back with major revisions and 3 were full reject. The paper acceptance rate for this volume is 33.3%.

As usual, we hope you would find important ideas, concepts, techniques, or results in this issue that might be important for your research activities.

Happy Reading and a Happy New Year 2010

IJCSI Editorial Board
November 2009
www.IJCSI.org

IJCSI Editorial Board 2009

IJCSI Reviewers Committee 2009

- Mr. AliReza Shahrestani, University of Malaya (UM), National Advanced IPv6 Centre of Excellence (NAv6), Malaysia
- Mr. Blagoj Ristevski, Faculty of Administration and Information Systems Management - Bitola, Republic of Macedonia
- Mr. Mauricio Egidio Cantão, Department of Computer Science / University of São Paulo, Brazil
- Mr. Jules Ruis, Fractal Consultancy, The netherlands
- Mr. Mohammad Iftekhar Husain, University at Buffalo, USA
- Dr. Deepak Laxmi Narasimha, Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
- Dr. Paola Di Maio, DMEM University of Strathclyde, UK
- Dr. Bhanu Pratap Singh, Institute of Instrumentation Engineering, Kurukshetra University Kurukshetra, India
- Mr. Sana Ullah, Inha University, South Korea
- Mr. Cornelis Pieter Pieters, Condast, The Netherlands
- Dr. Amogh Kavimandan, The MathWorks Inc., USA
- Dr. Zhinan Zhou, Samsung Telecommunications America, USA
- Mr. Alberto de Santos Sierra, Universidad Politécnica de Madrid, Spain
- Dr. Md. Atiqur Rahman Ahad, Department of Applied Physics, Electronics & Communication Engineering (APECE), University of Dhaka, Bangladesh
- Dr. Charalampos Bratsas, Lab of Medical Informatics, Medical Faculty, Aristotle University, Thessaloniki, Greece
- Ms. Alexia Dini Kounoudes, Cyprus University of Technology, Cyprus
- Mr. Anthony Gesase, University of Dar es salaam Computing Centre, Tanzania
- Dr. Jorge A. Ruiz-Vanoye, Universidad Juárez Autónoma de Tabasco, Mexico
- Dr. Alejandro Fuentes Penna, Universidad Popular Autónoma del Estado de Puebla, México
- Dr. Ocotlán Díaz-Parra, Universidad Juárez Autónoma de Tabasco, México
- Mrs. Nantia Iakovidou, Aristotle University of Thessaloniki, Greece
- Mr. Vinay Chopra, DAV Institute of Engineering & Technology, Jalandhar
- Ms. Carmen Lastres, Universidad Politécnica de Madrid - Centre for Smart Environments, Spain
- Dr. Sanja Lazarova-Molnar, United Arab Emirates University, UAE
- Mr. Srikrishna Nudurumati, Imaging & Printing Group R&D Hub, Hewlett-Packard, India
- Dr. Olivier Nocent, CReSTIC/SIC, University of Reims, France
- Mr. Burak Cizmeci, Isik University, Turkey
- Dr. Carlos Jaime Barrios Hernandez, LIG (Laboratory Of Informatics of Grenoble), France
- Mr. Md. Rabiul Islam, Rajshahi university of Engineering & Technology (RUET), Bangladesh
- Dr. LAKHOUA Mohamed Najeh, ISSAT - Laboratory of Analysis and Control of Systems, Tunisia
- Dr. Alessandro Lavacchi, Department of Chemistry - University of Firenze, Italy
- Mr. Mungwe, University of Oldenburg, Germany
- Mr. Somnath Tagore, Dr D Y Patil University, India
- Ms. Xueqin Wang, ATCS, USA
- Dr. Fondjo Fotou Franklin, Langston University, USA

- Mr. Supheakmungkol SARIN, Waseda University, Japan
- Mr. Shoukat Ullah, Govt. Post Graduate College Bannu, Pakistan
- Dr. Vivian Augustine, Telecom Zimbabwe, Zimbabwe
- Mrs. Mutalli Vatila, Offshore Business Philipines, Philipines
- Dr. Emanuele Goldoni, University of Pavia, Dept. of Electronics, TLC & Networking Lab, Italy
- Mr. Pankaj Kumar, SAMA, India
- Dr. Himanshu Aggarwal, Punjabi University,Patiala, India
- Dr. Vauvert Guillaume, Europages, France
- Prof Yee Ming Chen, Department of Industrial Engineering and Management, Yuan Ze University, Taiwan
- Dr. Constantino Malagón, Nebrija University, Spain
- Prof Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India
- Mr. Angkoon Phinyomark, Prince of Singkla University, Thailand
- Ms. Nital H. Mistry, Veer Narmad South Gujarat University, Surat, India
- Dr. M.R.Sumalatha, Anna University, India
- Mr. Somesh Kumar Dewangan, Disha Institute of Management and Technology, India
- Mr. Raman Maini, Punjabi University, Patiala(Punjab)-147002, India
- Dr. Abdelkader Outtagarts, Alcatel-Lucent Bell-Labs, France
- Prof Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India
- Mr. Prabu Mohandas, Anna University/Adhiyamaan College of Engineering, india
- Dr. Manish Kumar Jindal, Panjab University Regional Centre, Muktsar, India
- Prof Mydhili K Nair, M S Ramaiah Institute of Technnology, Bangalore, India
- Dr. C. Suresh Gnana Dhas, VelTech MultiTech Dr.Rangarajan Dr.Sagunthala Engineering College,Chennai,Tamilnadu, India
- Prof Akash Rajak, Krishna Institute of Engineering and Technology, Ghaziabad, India
- Mr. Ajay Kumar Shrivastava, Krishna Institute of Engineering & Technology, Ghaziabad, India
- Mr. Deo Prakash, SMVD University, Kakryal(J&K), India
- Dr. Vu Thanh Nguyen, University of Information Technology HoChiMinh City, VietNam
- Prof Deo Prakash, SMVD University (A Technical University open on I.I.T. Pattern) Kakryal (J&K), India
- Dr. Navneet Agrawal, Dept. of ECE, College of Technology & Engineering, MPUAT, Udaipur 313001 Rajasthan, India
- Mr. Sufal Das, Sikkim Manipal Institute of Technology, India
- Mr. Anil Kumar, Sikkim Manipal Institute of Technology, India
- Dr. B. Prasanalakshmi, King Saud University, Saudi Arabia.
- Dr. K D Verma, S.V. (P.G.) College, Aligarh, India
- Mr. Mohd Nazri Ismail, System and Networking Department, University of Kuala Lumpur (UniKL), Malaysia
- Dr. Nguyen Tuan Dang, University of Information Technology, Vietnam National University Ho Chi Minh city, Vietnam
- Dr. Abdul Aziz, University of Central Punjab, Pakistan
- Dr. P. Vasudeva Reddy, Andhra University, India

TABLE OF CONTENTS

1. Using social annotation and web log to enhance search engine - (pg 1-6)
Vu Thanh Nguyen, University of Information Technology, Ho Chi Minh City, VietNam

2. Document Searching System based on Natural Language Query Processing for Vietnam Open Courseware Library – (pg 7-13)
Dang Tuan Nguyen and Ha Quy-Tinh Luong, Faculty of Computer Science, University of Information Technology, Vietnam Nationnal University HCMC, Ho Chi Minh City, Vietnam

3. Gesture Recognition with a Focus on Important Actions by Using a Path Searching Method in Weighted Graph – (pg 14-19)
Kazumoto Tanaka, Dept. of Information and Systems Engineering, Kinki University, Higashi-hiroshima, 739-2116, Japan

4. Lexicographic Multi-objective Geometric Programming Problems – (pg 20-24)
A. K. Ojha, School of Basic Sciences, IIT Bhubaneswar, Orissa, Pin-751013 , India
K. K. Biswal, Department of Mathematics, CTTC Bhubaneswar, B-36, Chandaka Industrial Area, Bhubaneswar, Orissa, Pin-751024, India

5. High Availability Cluster System for Local Disaster Recovery with Markov Modeling Approach – (pg 25-32)
T. T. Lwin and T. Thein, University of Computer Studies, Yangon, Myanmar

6. A Brief History of Context – (pg 33-43)
Kaiyu Wan, Computer Science Department, East China Normal University, Shanghai, 200, China

7. ICT in Universities of the Western Himalayan Region in India: Status, Performance- An Assessment – (pg 44-52)
Dhirendra Sharma, University Institute of Information Technology, Himachal Pradesh University, Shimla, Himachal Pradesh 171 005, India
Vikram Singh, Department of Computer Science and Engg, Ch. Devi Lal University, Sirsa, Haryna 125 055, India

IJCSI International Journal of Computer Science Issues, Vol. 6, No. 2, 2009
ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

Using social annotation and web log to enhance search engine

VU THANH NGUYEN

University of Information Technology
Ho Chi Minh City, VietNam

Abstract

Search services have been developed rapidly in social Internet. It can help web users easily to find their documents. So that, finding a best method search is always an imagine. This paper would like introduce hybrid method of LPageRank algorithm and Social Sim Rank algorithm. LPageRank is the method using link structure to rank priority of page. It doesn't care content of page and content of query. Therefore, we want to use benefit of social annotations to create the latent semantic association between queries and annotations. This model, we use algorithm SocialPageRank and LPageRank to enhance accuracy of search system. To experiment and evaluate the proposed of the new model, we have used this model for Music Machine Website with their web logs.

Keywords: PageRank, LpageRank, ScocialPageRank, TF-IDF

1. Introduction

In the period of Internet, search engine is the popular tool and the necessary tool for the web users. However, they have not satisfied when using them such as: difficult to choose the right result in the huge results; difficult to have a right query; difficult to know which results are similar, so on… Therefore, over the past decade, there are many researching to improve the quality of web search. Most of them try to improve some aspects:

1) Rearrange the web pages according to the query document similarity. In this area, some techniques are anchor text generation, metadata extraction, link analysis, search log mining, profile query extraction, profile web usage,..

2) Ordering the web pages according to their priority. It doesn't care the query of user when ranking the priority of web pages. Some techniques are PageRank, HITS,…

In this paper, we optimize web search using LPageRank model by applying social annotations with two aspects: similar ranking and static ranking.

Similar ranking is used to estimate similarity between a query and a web page by annotations. They provide good information to summary of the corresponding web pages. They are metadata which can be used to calculate the similarity between a query and a web page. However, in some web pages, the annotations may be sparse and incomplete. These thing make gap between the annotations and queries. In this paper, we apply an algorithm SocialSimRank (SSR) to enhance this problem.

Another, static ranking is the algorithm ranking priority of web page by structure link.

It uses relationship of web pages in their site to valuate importance of pages. This method must browse all web page to built structure site.

There are many approaches to get the structure site. One of them is to built the structure site with traffic logs which use the information from the website's logs.

Recently, traffic log is used to enhance web search more and more such as: Yahoo patent, LPageRank model of Brin and Motwani and Winograd, LPageRank of Qing Cui and Alex Dekhtyar. However, they often focus one of approach static ranking or similar ranking.

In this paper, the model LPageRank with web log introduced by Qing Cui and Alex Dekhtyar in 2005 and the social annotation information are applied to our model. We have proposed a procedure that computes a score for a web page according to the number of visits to the page and the traffic pattern on the site. We have this hybrid method to build a local search engine. The search engine has been deployed with the web log of website http://machines.hyperreal.org/. The rest of the paper is organized as follows. In Section 2 we describe LPageRank algorithm. In Section 3 we describe social annotation search. In Section 4 we describe the search engine which we have built. In Section 4 we describe our initial experiments and provide the results.

2. LPageRank: Using logs in local search

2.1 PageRank algorithm

The PageRank algorithm introduced by Brin, Motwani and Winograd is a mechanism in determining the overall importance of a web page. Intuitively, PageRank of a web page is an approximation of a probability to reach this page from some other pages on the web. This computation assigns to each page a PageRank based on the current structure of the website. In the absence of information about human traffic patterns on the web, the PageRank computation assumes that on each page, the user is as likely to follow a specific link, as any other link.

Each link, PageRank assumes that the user is not biased in his/her choice of the link. Thus, the probability to follow a specific link is $\frac{1-\alpha}{m}$. When considering globally, the α leads to the following recursive formula for computing a PageRank (PR) of a page:

$$PR(A) = \alpha + (1-\alpha)\left(\sum_{B \in Parents(A)} \frac{PR(B)}{N(B)}\right)$$ [6]

where, Parents(A) is the set of all web pages which link to A and N(B) is the number of outgoing links to distinct pages found on page B. Typically, PageRank is computed iteratively, starting with a state in which it is uniformly distributed among all web pages, and continuing until a stabilization condition holds.

2.2 Probabilistic Graph of Web Page Collection

The PageRank implicitly models are used the behavior in terms of a probabilistic graph of the web. Indeed, each page in the collection can be thought of as a node in a graph. A link from page B to page A can be modeled as a directed edge (B, A). Finally, the PageRank computation assumes that, given a page B, the probability to follow some outgoing edge (B, A) is $\frac{1-\alpha}{N(A)}$. The triple G=(W, E, P) where N is a set of nodes, $E \subset WxW$ is a set of directed edges and P:E→[0,1], s.t.,

$(\forall B \in N)\sum_{(B,A)\in E} P(B,A) \le 1$ is called a *probabilistic graph* [3].

The probabilistic graph constructed (implicitly) by the PageRank computation assumes a uniform probability distribution for outgoing edges for each node.

2.3 LPageRank

This method is introduced by Qing Cui and Alex Dekhtyar in 2005. Basically, the LPageRank algorithm is a PageRank computation based on a probabilistic graph of a web page collection that reflects traffic patterns obtained from the logs.

Suppose G= (W, E, P) is a probabilistic graph over a collection of web pages W. Then, LPageRank (LPR) of a web page is computed as follows:

$$LPR(A) = \alpha + (1-\alpha) \sum_{B \in Parents(A)} LPR(B)P(B,A)$$ [3]

Note that LPR(A) = PR(A) for graphs G, in which $P(B,A) = \frac{1-\alpha}{N(B)}$ for all edges (B, A).

This model is similar to the Yahoo patent in 2002 as using traffic logs to enhance search engine. However, compared with LPageRank, it has many differences. This method uses web logs to build probabilistic graph and then improves the PageRank algorithm. The URL frequency is not only used to score each URL but also strengthen other URLs which it links to. Moreover, in this model, traffic logs are used to build the structure of website, which is easily than using crawler.

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

3. Search engine with social annotation

3.1. Web page annotators

Web page annotators are web users who use annotation to organize memories and share their favorite online. They provide cleaner data which are usually good summarizations of web pages for user's browsing. Besides, similar or closely related annotations are usually given to the same web pages. Base on this observation, SocialSimRank (SSR) is used to measure the similarity between the query and annotations.

In 2007, Shenghua Bao, Xiaoyuan Wu, Ben Fei, Guirong Xue, Zong Su, Yong Yu had an identify about SSR. They consider that SSR is a measure the popularity of web pages from web page annotator's point of view. In figure 1 the Social annotations indicate a, d and c. The a web page is more population than d and c. Each annotation has a relationship with others. The similar (semantically-related) annotations are usually assigned to similar (semantically-related) web pages by users with common interests. In the social annotation environment, the similarity among annotations in various forms can further be identified by the common web pages they annotated.

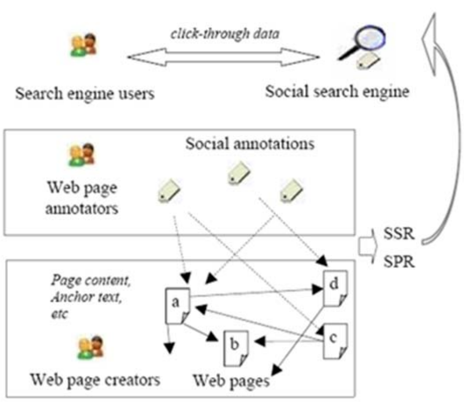

Figure 1 Social annotations

To calculate the similar annotation, we build a relation graph between social annotations and web pages with its edges indicating the user count. Assume that there are *NA* annotations, *NP* web pages and *NU* web users. *MAP* is the *NA*×*NP* association matrix between annotations and pages. *MAP*(*ax*,*py*) denotes the number of users who assign annotation *ax* to page *py*. Letting *SA* be the *NA*×*NA* matrix whose element *SA*(*ai*, *aj*) indicates the similarity score between annotations *ai* and *aj* and *SP* be the *NP*×*NP* matrix each of whose element stores the similarity between two web pages, we propose SocialSimRank(SSR), an iterative algorithm to quantitatively evaluate the similarity between any two annotations.

Step 1: Init:

Let $S_A^0(a_i, a_j) = 1 \; for \; each \; a_i = a_j \; otherwise \; 0$

$S_P^0(p_i, p_j) = 1 \; for \; each \; p_i = p_j \; otherwise \; 0$

Step 2: Do {

For each annotation pair (a_i, a_j) do

$$S_A^{k+1}(a_i, a_j) = \frac{c_A}{|P(a_i)||P(a_j)|}$$

$$\sum_{m=1}^{|P(a_i)|} \sum_{n=1}^{|P(a_j)|} \frac{\min(M_{AP}(a_i, p_m), M_{AP}(a_j, p_n))}{\max(M_{AP}(a_i, p_m), M_{AP}(a_j, p_n))} S_P^k(P_m(a_i), P_n(a_j))$$

For each page pair (p_i, p_j) do

$$S_P^{k+1}(p_i, p_j) = \frac{C_P}{|A(p_i)||A(p_j)|}$$

$$\sum_{m=1}^{|A(p_i)|} \sum_{n=1}^{|A(p_j)|} \frac{\min(M_{AP}(a_m, p_i), M_{AP}(a_n, p_j))}{\max(M_{AP}(a_m, p_i), M_{AP}(a_n, p_j))} S_A^{k+1}(A_m(p_i), A_n(p_j))$$

}

Step 3 Output $S_A(a_i, a_j)$

In this algorithm, *CA* and *CP* denote the damping factors of similarity propagation for annotations and web pages, respectively. *P*(*ai*) is the set of web pages annotated with annotation *ai* and *A*(*pj*) is the set of annotations given to page *pj*. *Pm(ai)* denotes the *m*th page annotated by *ai* and *Am(pi)* denotes the *m*th annotation assigned to page *pi*.

Note that the similarity propagation rate is adjusted according to the number of users between the annotation and web page.

Letting *q*={*q*1,*q*2,…,*qn*} be a query which consists of *n* query terms and *A*(*p*)={*a*1,*a*2,…, *am*} be the annotation set of web page *p*,

Equation (4) shows the similarity calculation method based on the SocialSimRank.

3.2. Dynamic Ranking page with web log and similarity annotation

Given the probabilistic graph in the previous step, the LPageRank computation is performed. In the first iteration, the LPageRank of each page is allocated by the total scores of all forward links. In the next iteration, the LPageRank of all pages are computed as follows:

$$LPR_{i+1}(A) = \alpha + (1-\alpha) \sum_{B \in Parents(A)} LPR_i(B) P(B,A)$$

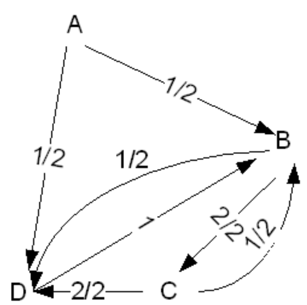

Figure 2 Illustration of quality transition between the users, annotations, and pages in the LPageRank and similarity annotation algorithm

This process is stopped when the difference of LPR_i and LPR_{i+1} is less than small number ε . At the present implementation, ε is set to 0:0.00001. Then we use the score of each link to calculate the similar annotation.

4. Experimental results

Some preliminary experiments have been conducted to test the LPageRank-augmented retrieval. In the following section, the obtained results will be described.

5. Experimental Setup

The search engine described in this paper had been implemented on virtual web services and deployed on the local website. All of files of Music Machine Website are downloaded and a virtual website is built such as a online website. To test its performance, we have considered two alternative local search methods: Google's domain-restricted search with URL http://machines.hyperreal.org/ and local PageRank.

Our search engine indexes around 14002 web pages in the http://machines.hyperreal.org/ domain. Only HTML and text files are indexed. For this experiment, the collection of logs is taken from the http://machines.hyperreal.org/ web servers.

Table 1: Summary of result: Google and model local search

Key words	Google	New model
Arp-Sequencer	6	2
Roland TR-606 Dramatic	2	2
MonoPoly	23	24
Kawai K3	12	12
ragtime piano	1	1
Kawai XD5	10	10
BASS DRUM	77	76
Maplin 3800	5	6
Maplin 5600S	7	7
Hammond Auto-Vari	4	4
Univox Micro-Rhythmer-12	9	8

A test dataset has been selected consisting of 11 simple keyword queries shown in Table 1.

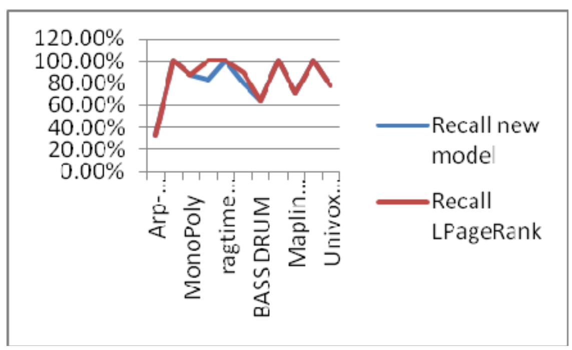

Figure 3: The averaged recall on all query from LPageRank and LPageRank with neural network

6. Results

The conclusion is as follows:

- Overall, Google's domain-restricted search is showed the best performance on the test dataset in both precision and average expected precision.
- Local search site has a new order of pages on which the URLs have a high hit weight that often has high rank.
- Some pages disappear from the result of the local search when the hit weight is too low.
- The result of the local search is sensitive with a hit weight.

7. Discussion

From the above experiment, two positive conclusions can be reached. First, if everything is given equally, the LPageRank outperforms regular PageRank and provides more relevant links and better quality lists of links to the users. Second, the local search's retrieval appears to be more significantly different than the retrieval by the domain-restricted Google's.

On the other hand, it is a notice that the LPageRank combined by TF-IDF is not enough better than Google's domain-restricted search. There are two potential reasons for Google's better performance: First, while the exact matching formulas of Google are proprietary secrets and ever since we know that Google combines PageRank with more than just TF-IDF based on retrieval. Google's engine analyzes the position of keywords on in the HTML, the text of links pointing to the page, and possibly many other factors ignored by TF-IDF. Second, we have observed that Google has indexed potentially a much larger number of pages in the "http://machines.hyperreal.org/" domain. While some documents indexed by Google are not HTML or text files, and while some other documents are no longer present on the http://machines.hyperreal.org/ web server, it is reasonable to assume that Google's list of http://machines.hyperreal.org/ pages is larger. In part, this can be explained by our crawling mechanism - we will never find any web pages that are NOT reachable from the top page of the website, http://machines.hyperreal.org/. At the same time, Google's crawler is global and can detect these pages if they are linked from outside sources. These results are encouraging and showing that the use of the LPageRank with appropriate retrieval methods may improve retrieval results. They also suggest two avenues for further improvement. First, we must subject our search engine to more rigorous testing, involving independent users rather than authors. Second, we must work to improve the IR component of the search engine and make it resemble more the actual methods used in existing web search engines. We also note that the local search can be used as a means of analyzing the website traffic and changes in it over time.

8. Conclusions

In this article, the use of web logs has been studied to argue single website search engines – a type of search engine almost ubiquitously present on all major commerce, academic and interest based on websites. Our preliminary results indicate that our proposed method, local search, outperforms the standard PageRank when coupled with a simple retrieval method (TF-IDF) and clusters of the neural network. An another discovery is that although our local search-based on a search engine shows worse performance than the domain restricted Google search. It helps the users to retrieve significantly different results, and thus shows complimentary to Google behavior. A plan for further studies of the LPageRank and similar annotation has been briefly outlined as well.

In conclusion, the web log resource is used to improve the local search engine, especially the PageRank algorithm and cluster sessions. This model has the high recall but it is still a low process and not a good precision. Although the result is not good, the benefits of web logs in local search engine are discovered, for instance, lower costs in building the structure graph and high reflections of web usages in the result.

References

[1] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, (2008), Introduction to Information Retrieval, Cambridge University Press., pp(346-368).

[2] Jin Zhou, Chen Ding, Dimitrios Androutsos (2006), Improving Web site using web server logs.

[3] Qing Cui, Alex Dekhtyar (2005), On Improving Local Website Search Using Web Server Traffic Logs: A Preliminary Report.

[4] Olfa Nasraoui, Mrudula Pavuluria (2004), Complete this Puzzle: A connectionist approach to accurate web recommendations based on a committee of predictors.

[5] Olfa Nasraoui, Mrudula Pavuluria (2004), Accurate web recommendations based on profile specific URL-predictor Neural networks.

[6] The PageRank citation ranking: bringing order to the web (1998), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.1768

[7] Arkady Borkovsky, Douglas M. Cook, Jean-Marc Langlois, Tomi Poutanen, and Hongyuan Zha (2002), Using network traffic logs for search enhancement.

[8] Nguyen Đinh Thuc, Artificial intelligence, neural network and their applications. Education Publishing House, 2000

Author Biographies – Vu Thanh Nguyen

The author born in 1969 in Da Nang, VietNam. He graduated University of Odessa (USSR), in 1992, specialized in Information Technology. He postgraduated on doctoral thesis in 1996 at the Academy of Science of Russia, specialized in IT. Now he is the Dean of Software Engineering of University of Information Technology, VietNam National University HoChiMinh City.

Research: Knowledge Engineering, Information Systems and software Engineering.

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

Document Searching System based on Natural Language Query Processing for Vietnam Open Courseware Library

Dang Tuan NGUYEN and Ha Quy-Tinh LUONG

Faculty of Computer Science

University of Information Technology, Vietnam Nationnal University HCMC

Ho Chi Minh City, Vietnam

Abstract

The necessary of buiding the searching system being able to support users expressing their searching by natural language queries is very important and opens the researching direction with many potential. It combines the traditional methods of information retrieval and the researching of Question Answering (QA). In this paper, we introduce a searching system built by us for searching courses on the Vietnam OpenCourseWare Program (VOCW). It can be considered as the first tool to be able to perform the user's Vietnamese questions. The experiment results are rather good when we evaluate this system on the precision and the run-time of answering the Vietnamese questions.

Keywords: *Natural Language Processing, Document Retrieval, Search, Question Answering, Knowledge Base*.

1. Introduction

The necessary of buiding the searching system being able to support users expressing their searching by natural language queries is very important and opens the researching direction with many potential. It combines the traditional methods of information retrieval and the researching of Question Answering (QA).

Several searching engineering systems with supporting English language in e-library based on the natural language query processing were built by our previous publications [1], [2], [3], [4], [5], [6], [7], [8]. In the continued researches, we aimed to developing a document searching system base on answering to the user's Vietnamese questions.

In this paper, we introduce a searching system built by us for searching courses on the Vietnam OpenCourseWare Program[1] (VOCW). It can be considered as the first tool be able to perform the user's Vietnamese questions. The experiment results are rather good when we evaluate this system on the precision and the run-time of answering the Vietnamese questions.

[1] http://vocw.edu.vn/

2. System Architecture

The system model is represented in the Fig. 1, including the main components as follow:

- The extracting component: extracting the data in the metadata pages of VOCW.
- The ontology-based knowledge (VOCW ontology) for storing VOCW data.
- The parsing component: analyzing the syntactic of Vietnamese question and returning the parse tree and generating tree. The generating tree will be the input of the generating query. It operates on the set of the defined syntactic rules for analyzing Vietnamese question.
- The generating component: generating SPARQL[17] query for the querying component.
- The querying component: inserting, updateting and querying them to the VOCW ontology.

On this system model, the system processes as follow:

- When the user inputs the Vietnamese question, the parsing component will analyze its syntax. We defined the set of EBNF syntactic rules before. If user's question syntactic isn't belong to this set, the system returns that this question isn't analyzed (no syntactic exactly) and then stop program. By contrast, it returns the parser tree and the generating tree. The generating component will use the generating tree for creating SPARQL query. And the querying component gets the data on VOCW ontology.
- The extracting component extracts the data in the metadata pages of VOCW. And the querying component inserts the extracted data to VOCW ontology.

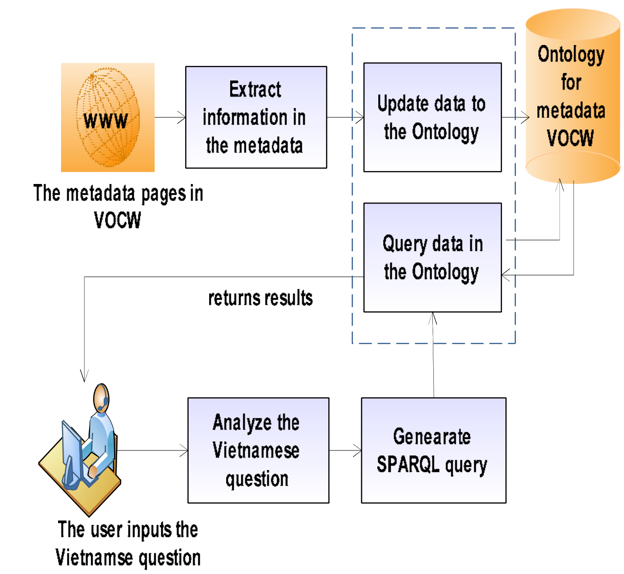

Fig. 1 System architecture

For example, we consider the Vietnamese question:

- *"Ai đã viết sách Toan?"*
  ("Who wrote Toan book?")

This question will be syntactically parsed and then will output the parse tree and generating tree as follow:

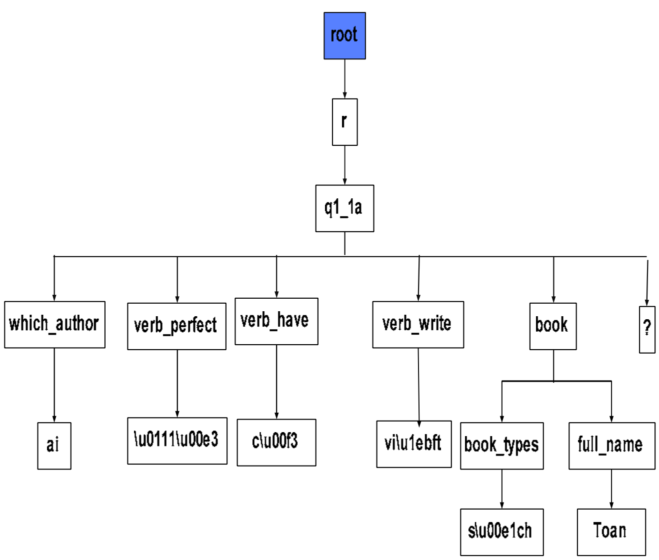

Fig. 2 Syntactic tree of a Vietnamese query (base on antlrworks generated parse tree)

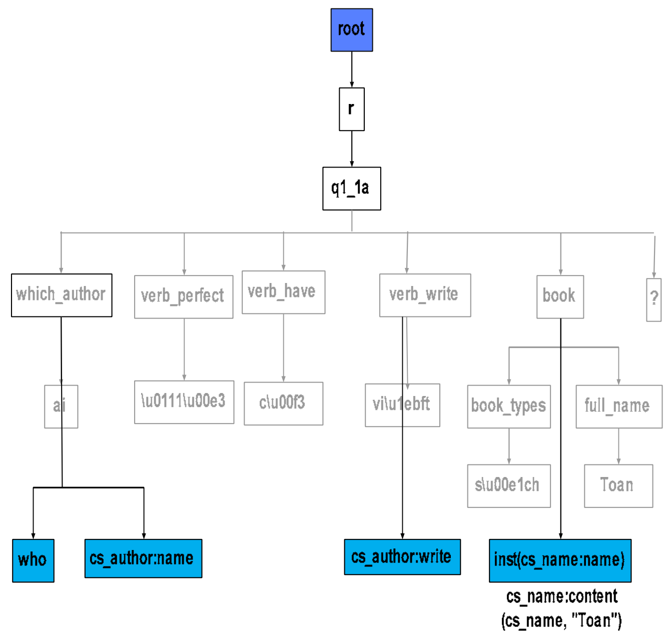

Fig. 3 The generating tree

The SPARQL query is created from the generating tree:

```
SELECT DISTINCT ?authorname
FROM <http://localhost/owl_test/vocw_full.owl>
WHERE {
   {?author cs_author:content ?authorname}.
   {?author cs_author:write ?course}.
   {?year cs_created:content ?yearname
   FILTER regex(?yearname , "^2009$", "i" )}.
   {?course cs_name:isWrittenIn ?year}.
   {?course cs_name:content ?coursename
   FILTER regex(?coursename , "^cấu trúc dữ liệu$", "i" ).
   }
}
```

3. Syntactic Rules

The parsing component is built for analyzing Vietnamese questions regarding to the data such as name, language, summary, authors, copyright holders, maintainers keywords, version and affiliations.

For examples, some different questions as follow:

- *Ai đã viết cuốn sách B vào năm 2000?*
  (Who wrote book B in 2000?)
- *Nhà xuất bản nào đã phát hành cuốn B trong năm 2008?*
  (Which publisher published book B in 2008?)
- *Sách B được tác giả A viết vào năm nào?*
  (What year did writer A write book B?)

Syntax of Vietnamese questions can be built by EBNF notation (Extended Backus–Naur Form). We defined 40 syntactic rules for Vietnamese questions regarding the above data.

In this part, we use ANTLR tool for generating lexer and parser code in order to analyzing and checking syntax of Vietnamese questions (set of the defined EBNF syntactic rules). And we design this component for adding new rules easily but no modifying code much in future.

TABLE 1. SYNTACTIC RULES

No	Syntactic rules
1	<Q1.1a> = <what_author> [<vperfect>] [<interrogative1>] <verb_write> <book> {[<conjunction>] <book>} [<time_phrase>] “?”
2	<Q1.1b> = [<time_phrase>] [“,”] <what_author> [<vperfect>] [<interrogative1>] <verb_write> <book> {[<conjunction>] <book>} “?”
3	<Q1.1c> = <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <what_author> <verb_write> [<time_phrase>] “?”
4	<Q1.1d> = [<time_phrase>] [“,”] <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <what_author> <verb_write> “?”
5	<Q1.2a> = [<interrogative3>] <creator> [<possessive>] <book> {[<conjunction>] <book>} <verb_be> <author> [<interrogative2>] “?”
6	<Q1.2b> = [<interrogative3>] <author> <verb_be> <creator> [<possessive>] <book> {[<conjunction>] <book>} [<interrogative2>] “?”
7	<Q1.2c> = <author> [<interrogative3>] <verb_be> <creator> [<possessive>] <book> {[<conjunction>] <book>} [<interrogative2>] “?”
8	<Q1.3a> = [<interrogative3>] <author> [<vperfect>] [<interrogative1>] <verb_write> <book> {[<conjunction>] <book>} [<time_phrase>] [<interrogative2>] “?”
9	<Q1.3b> = [<time_phrase>] [“,”] [<interrogative3>] <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <author> <verb_write> [<interrrogative2>]“?”
10	<Q1.4a> = <author> [<vperfect>] [<interrogative1>] <verb_write> <book> {[<conjunction>] <book>} [<prep_time>] < what_time > “?”
11	<Q1.4b> ::= <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <author> <verb_write> [<prep_time>] <what_time> “?”
12	<Q2.1a> = <what_publisher> [<vperfect>] [<interrogative1>] <verb_publish> <book> {[<conjunction>] <book>} [<time_phrase>] “?”
13	<Q2.1b> = [<time_phrase>] [“,”] <what_publisher> [<vperfect>] [<interrogative1>] <verb_publish> <book> {[<conjunction>] <book>} “?”
14	<Q2.1c> = <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <what_publisher> <verb_publish> [<time_phrase>] “?”
15	<Q2.1d> = [<time_phrase>] [“,”] <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <what_publisher> <verb_publish> “?”
16	<Q2.2a> = [<interrogative3>] <publisher> [<vperfect>] [<interrogative1>] <verb_publish> <book> {[<conjunction>] <book>} [<time_phrase>] [<interrogative2>] “?”
17	<Q2.2b> = [<time_phrase>] [“,”] [<interrogative3>] <publisher> [<vperfect>] [<interrogative1>] <verb_publish> <book> {[<conjunction>] <book>} [<interrogative2>] “?”
18	<Q2.2c> = [<interrogative3>] <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <publisher> <verb_publish> [<time_phrase>] [<interrogative2>] “?”
19	<Q2.2d> = [<time_phrase>] [“,”] [<interrogative3>] <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <publisher> <verb_publish> [<interrogative2>] “?”
20	<Q2.3a> = <publisher> [<vperfect>] [<interrogative1>] <verb_publish> <book> {[<conjunction>] <book>} [<prep_time>] <what_time> “?”
21	<Q2.3b> = [<prep_time>] <what_time> <publisher> [<vperfect>] [<interrogative1>] <verb_publish> <book> {[<conjunction>] <book>} “?”
22	<Q2.3c> = <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <publisher> <verb_publish> [<prep_time>] <what_time> “?”
23	<Q2.3d> = [<prep_time>] <what_time> <book> {[<conjunction>] <book>} [<vperfect>] <vpassive> <publisher> <verb_publish> “?”
24	<Q3.1a> = <book> [<of_author>][<by_publisher>][<time_phrase>] <is_of> <what_subject> ?
25	<Q3.1b> = [<time_phrase>] [,] <book> [<of_author>] [<by_publisher>] <is_of> <what_subject> ?
26	<Q3.1c> = <field> <possessive> <book> [<of_author>] [<by_publisher>] [<time_phrase>] <interrogative4> ?
27	<Q3.1d> = [<time_phrase>] [,] <field> <possessive> <book> [<of_author>] [<by_publisher>] <interrogative4> ?
28	<Q3.2a> = <book> [<of_author>] [<by_publisher>] [<time_phrase>] [<interrogative1>] <is_of> <subject> [<interrogative2>] ?
29	<Q3.2b> = [<time_phrase>] [,] <book> [<of_author>] [<by_publisher>] [<interrogative1>] <is_of> <subject> [<interrogative2>] ?
30	<Q3.2c> = <book> [<of_author>] [<by_publisher>] [<time_phrase>] [<interrogative3>] <verb_be> <book_type> <is_of> <subject> [<interrogative2>] ?
31	<Q3.2d> = [<time_phrase>] [,] <book> [<of_author>] [<by_publisher>] [<interrogative3>] <verb_be> <book_type> <is_of> <subject> [<interrogative2>] ?
32	<Q3.3a> = [<time_phrase>] [,] <author> [<vperfect>] [interrogative1] <verb_write> [<plural>] <book_type> <verb_have> <what_subject> ?
33	<Q3.3b> = <author> [<vperfect>] [interrogative1] <verb_write> [<plural>] <book_type> <verb_have> <what_subject> [<time_phrase>]?
34	<Q3.3c> = [<time_phrase>] [,] <author> [<vperfect>] [interrogative1] <verb_write> [<plural>] <book_type> <is_of> <what_subject> ?
35	<Q3.3d> = <author> [<vperfect>] [interrogative1] <verb_write> [<plural>] <book_type> <is_of> <what_subject>

	[<time_phrase>] ?
36	<Q3.4a> = <publisher> [<vperfect>] [<interrogative1>] <verb_publish> [<plural>] <verb_have> <what_subject> [<time_phrase>] ?
37	<Q3.4b> = [<time_phrase>] <publisher> [<vperfect>] [<interrogative1>] <verb_publish> [<plural>] <verb_have> <what_subject> ?
38	<Q3.4c> = <publisher> [<vperfect>] [<interrogative1>] <verb_publish> [<plural>] <is_of> <what_subject> [<time_phrase>] ?
39	<Q3.4d> = [<time_phrase>] <publisher> [<vperfect>] [<interrogative1>] <verb_publish> [<plural>] <is_of> <what_subject> ?
40	<Q4.1a> = [plural] [book_type] [<verb_have> <subject>] [<by_author>] [<time_phrase>] <interrogative4> ?
41	<Q4.1b> = [<time_phrase>] [,] [plural][book_type] [<verb_have><subject>] [<by_author>] [interrogative4] ?
42	<Q4.1c> = [plural][book_type] [<is_of><subject>] [<by_author>] [<time_phrase>]<interrogative4> ?
43	<Q4.1d> = [<time_phrase>] [,] [plural][book_type] [<is_of><subject>] [<by_author>] <interrogative4> ?
44	<Q4.2a> = [plural] <book_type> [<verb_have> <subject>] <by_publisher> [<time_phrase>] <interrogative4> ?
45	<Q4.2b> = [<time_phrase>][,][plural]<book_type> [<verb_have> <subject>] <by_publisher> <interrogative4> ?
46	<Q4.2c> = [plural]<book_type> [<is_of><subject>] <by_publisher> [<time_phrase>] <interrogative4> ?
47	<Q4.2d> = [<time_phrase>] [,] [plural] <book_type> <is_of> <subject> <by_publisher> <interrogative4> ?
48	<Q5.1a> = <book> [<vperfect>] <vpassive> [<publisher>] <verb_publish> <what_place> [<time_phrase>] "?"
49	<Q5.1b> = [<time_phrase>] [","] <book> [<vperfect>] <vpassive> <verb_publish> <what_place> "?"
50	<Q5.2> = <publisher><verb_locate><what_place> "?"
51	<Q6.1a> = [<verb_buy>] <book> <verb_cost> "?"
52	<Q6.1b> = <price> [<possessive>] <book> [<what_price>] "?"
53	<Q7.1> = <how_many> <book> <in_elib> "?"
54	<Q7.2a> = <author> [<vperfect>] [<interrogative1>] <verb_write> <how_many> <book> [<time_phrase>] "?"
55	<Q7.2b> = [<time_phrase>] [","] <author> [<vperfect>] [<interrogative1>] <verb_write> <how_many> <book> "?"
56	<Q7.3a> = <publisher> [<vperfect>] [<interrogative1>] <verb_publish> <how_many> <book> [<time_phrase>] "?"
57	<Q7.3b> = [<time_phrase>] [","] <publisher> [<vperfect>] [<interrogative1>] <verb_publish> <how_many> <book> "?"

4. Ontology Model

Ontology is built from the common terminologies containing concepts (terms), properties, definitions of its and relationships of its, including constraints. It supplies the possibility of the semantic representation and reasoning support

A statement is a triple (resource-property-value) for defining the property of the resource. "Resource" is considered as an object, a thing we want to speak. Example, Resource: authors, books, publishers, places, people, hotels…, URL, URI. "Property" is a kind of the particular resource, describing relations between resouces, for example, "written by", "age", "title"… And "value" can be a reaource or literal

Example:

- *Tác giả A viết sách B*
  (Author A writes book B)

is represented as follow:

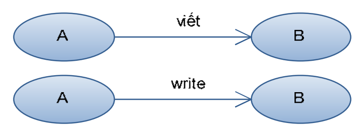

Fig. 4 Triple of write(cs_author,cs_name)

First, we build the semantic model between the object classes. Then, each statement is a triple, where resource and value are instances of classes.

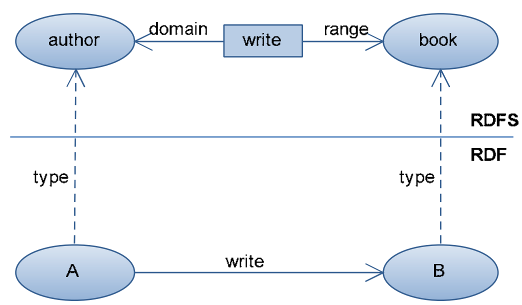

Fig. 5 The RDF/RDFS layers of write(cs_author,cs_name)

- *Resource: "Tác giả A" (Writer A)*
- *Property: "viết bởi" (written by)*
- *Value: : "sách B" (book B)*

In case of many statements: the example of the semantic net representing relations between "Writer", "Book", "Year" and "Publisher":

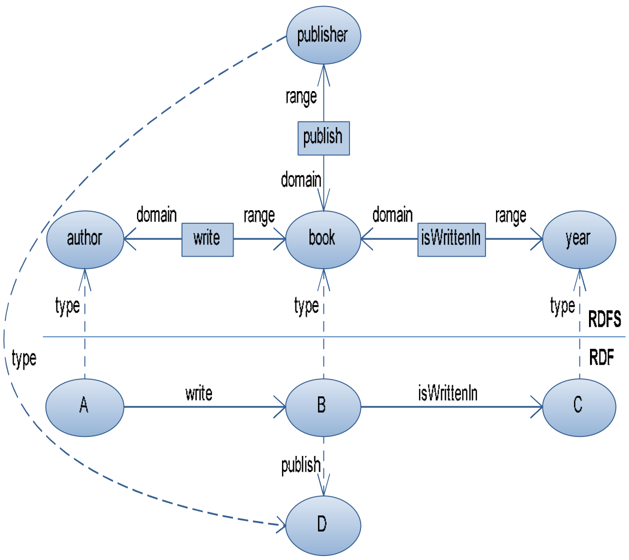

Fig. 6 The RDF/RDFS layers for "A write B book" and "B book is written by A" sentences

Terminology (resource, property, value) is organized, built and defined in hierarchy favourable to defining semantics exactly.

In the model above, we could add new properties from the existed properties (SWRL[14,15])

write(author,book) → *isWrittenBy(book,author)*

The vocwqa component carries out adding new data by SPARQL query language. Then, it checks the consitence by Pellet [16] và does inferences by Jess [15]

5. Generating Ontology Query

We consider single Vietnamese questions. It could contain one or many conjunctions "và" (and) or "hoặc" (or).

- In case of word "hoặc" (or): SPARQL supports keyword UNION for querying one possibility or more.
- In case of word "và" (and):

Example:

- *Ai đã có viết sách "Toan" và sách "Van" trong năm 2009?*

This sentence is understood in other way as follow:

{Ai đã có viết sách "Toan" trong năm 2009 mà
{người này cũng viết sách "Van" trong năm 2009}
?}

The example form:

Who Author Name[Toan,Van] ?

We have:

- the resources and property: Author, Name.
- the question word: who.
- the value: Toan, Van.

This case is resolved with the nested SPARQL form. Because each node of the mapping tree could be root node of other subtree, the mapping syntactic tree could represent nested query in order to participate in SPARQL generating process.

We see that this case has the general form as follow:

- Set the resources, property and question word: A_i.
- Set the value: x_i.

The general form:

$$A_1\ A_2 \ldots A_i[x_{i1}, x_{i2}, \ldots, x_{im}] \ldots A_k[x_{k1}] \ldots A_n?$$

Each question has only a component of word (terminology) questioned with a lot of the specific data by the linking-word "và" (and) or "hoặc" (or). Other components could be questioned at most a specific datum

Return example:

{Ai đã có viết sách "Toan" mà (1)
{người này cũng viết sách "Van" } (2)
?}

Its form:

Who Author Name[Toan]
[Who Author Name[Van]] ?

We see that sub-sentence (2) differ only sentence (1) in book "Van".

Let consider:

$$A_1\ A_2 \ldots A_i[x_{i1}, x_{i2}, \ldots, x_{im}] \ldots A_k[x_{kl}] \ldots A_n?$$

We could analyze into m sentences. Sentences differ each other in component i.

$A_1\ A_2 \ldots Axi_1 \ldots A_i[xk_1] \ldots A_n?$
$A_1\ A_2 \ldots Axi_1 \ldots A_i[xk_2] \ldots A_n?$
...
$A_1\ A_2 \ldots Axi_1 \ldots A_i[xk_m] \ldots A_n?$

On the syntactic tree, each node $A_i[xk_j]$ contains subtree. This subtree has root node $A_i[x_{kj}]$ including all subnode of parent-tree exception nodes $A_i[x_{kl}]_{(l=1..m)}$.

The above example will be generated to SPARQL query:

```
SELECT DISTINCT ?authorname
FROM <http://localhost/owl_test/vocw_full.owl>
WHERE {
   {?author cs_author:content ?authorname}.
   {?author cs_author:write ?course}.
   {?course cs_name:content ?coursename
   FILTER regex(?coursename , "^Toan$", "i" ).
   }
   .{
      SELECT DISTINCT ?authorname
      FROM <http://localhost/owl_test/vocw_full.owl>
      WHERE {
      {?author cs_author:content ?authorname}.
      {?author cs_author:write ?course}.
      {?course cs_name:content ?coursename
      FILTER regex(?coursename , "^Van$", "i" ).
      }}
   }
}
```

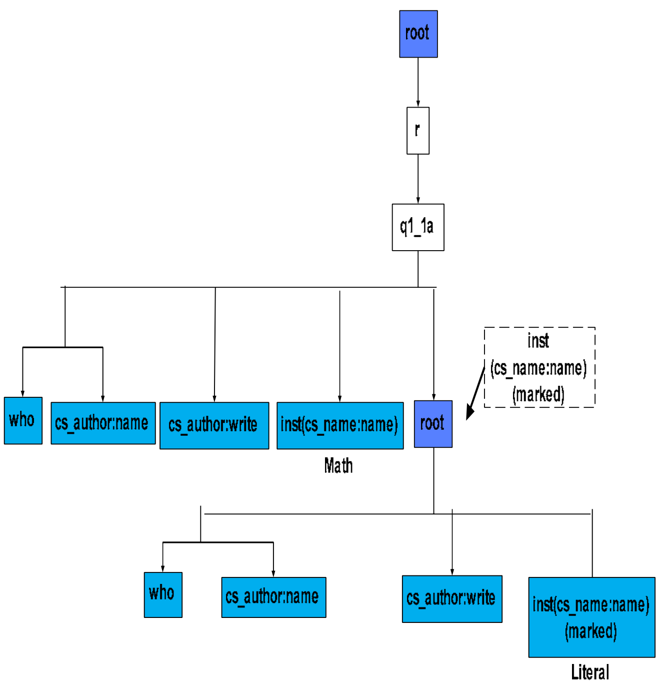

Fig. 7 The generating tree of "Who wrote Math and Literal book?"

6. Experiments

Base on the model system in Fig. 1, we developed the searching system with the specific characteristics for VOCW. It is evaluated according to the testing method that the track QA@INEX (Question Answering of INEX forum) used. We created the list of experiment questions and evaluated them manually by checking the answer appropriating to the question or not. According to this method, we didn't consider the satisfaction of the answer for the question. Thus, we only consider the "acceptability" of the answer for the question.

We developed the standard test suite including 40 Vietnamese questions corresponding with 40 syntactic rules, respectively. In addition, we collected more 91 questions (the random test collection) from users. We ran them on the computer with the configuration:

- Intel(R) Core(TM)2 Duo CPU, T5750 @ 2.00GHz, 997MHz, 0.99GB of RAM.
- WindowXP SP2

And we have the results as follow:

- The standard test suite:
  Percent test passed: 40/40 (100%)
- The random test collection:
  Percent test passed: 77/91 (84.62% test passed)
- The total run time: 2.54m /91
- The average run time of one question : 1.65s

With the above results and evaluation methods, the system is rather well.

7. Conclusions

We built the model, deployed the system for VOCW in reality. It is experimented and evaluated. And the results are rather good as the above observation. However, there are the first results in this application domain. The processing on the complex Vietnamese is a big challenging, so no real system in nowadays can solve it.

We concentrate on the research on the stronger syntactic parser for Vietnamese. The most important we need to extend the semantic model for the query and the methods and technologies for processing the semantic model in order to understanding by the computer.

The experience on the built system is useful for us to develop the other similar searching systems processing the user's Vietnamese question

References

[1] Dang Tuan Nguyen, Tuyen Thi-Thanh Do, "E-Library Searching by Natural Language Question-Answering System", Proceedings of the Fifth International Conference on Information Technology in Education and Training (IT@EDU2008), pages: 71-76, Ho Chi Minh and Vung Tau, Vietnam, December 15-16, 2008.

[2] Dang Tuan Nguyen, Tuyen Thi-Thanh Do, "e-Document Retrieval by Question Answering System", International Conference on Communication Technology, Penang, Malaysia, February 25-27,

2009. Proceedings of World Academy of Science, Engineering and Technology, Volume 38, 2009, pages: 395-398, ISSN: 2070-3740.

[3] Dang Tuan Nguyen, Tuyen Thi-Thanh Do, "Natural Language Question Answering Model Applied To Document Retrieval System", International Conference on Computer Science and Technology, Hongkong, China, March 23-25, 2009. Proceedings of World Academy of Science, Engineering and Technology, Volume 39, 2009, pages: 36-39, ISBN: 2070-3740.

[4] Dang Tuan Nguyen, Tuyen Thi-Thanh Do, "Document Retrieval Based on Question Answering System", Proceedings of the Second International Conference on Information and Computing Science, pages: 183-186, Manchester, UK, May 21-22, 2009. ISBN: 978-0-7695-3634-7. Editions IEEE.

[5] Dang Tuan Nguyen, Tuyen Thi-Thanh Do, Quoc Tan Phan, "A Document Retrieval Model Based-on Natural Language Queries Processing", Proceedings of the International Conference on Artificial Intelligence and Pattern Recognition (AIPR), pages: 216-220, Orlando, FL, USA, July 13-16, 2009. ISBN: 978-1-60651-007-0. Editions ISRST.

[6] Dang Tuan Nguyen, "Interactive Document Retrieval System Based-on Natural Language Query Processing", Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, pages: 2233-2237, Baoding, Hebei, China, July 12-15 2009. ISBN: 978-1-4244-3703-0. Editions IEEE.

[7] Dang Tuan Nguyen, Tuyen Thi-Thanh Do, Quoc Tan Phan, "Integrating Natural Language Query Processing and Database Search Engine", Proceedings of the 2009 International Conference on Artificialal Intelligence - ICAI'09, Volume 1, pages: 137-141, Las Vegas, Nevada, USA, July 13-16, 2009. ISBN: 1-60132-107-4, 1-60132-108-2 (1-60132-109-0). CSREA Press.

[8] Dang Tuan Nguyen, Tuyen Thi-Thanh Do, Quoc Tan Phan, "Natural Language Interaction-Based Document Retrieval", *The 2nd IEEE International Conference on Computer Science and Information Technology 2009* (ICCSIT 2009), Volume 4, pages: 544-548. Beijing, China, August 8-11, 2009. ISBN: 978-1-4244-4520-2. Editions IEEE.

[9] Dragan Gasˇevic´, Dragan Djuric´, Vladan Devedzˇic' (2006), Model Driven Architecture and Ontology Development, page 58-68, 91-108, p.107-108.

[10] Grigoris Antoniou và Frank van Harmelen (2004), A Semantic Web Primer, The MIT Press Cambridge, Massachusett, London, England, p.31-33.

[11] Natalya F. Noy và Deborah L. McGuinness (2001), Ontology Development 101: A Guide to Creating Your First Ontology

[12] Terence Parr, The Pragmatic Bookshelf, Raleigh, North Carolina Dallas, Texas (2007), The Definitive ANTLR Reference, Building Domain-Specific Languages, page 98-99

[13] Franz Baader, Deborah L. McGuinness, Daniele Nardi, Peter F. Patel-Schneider (), THE DESCRIPTION LOGIC HANDBOOK: Theory, implementation, and applications, page 47-100, page 495-505

[14] Protégé (http://protege.stanford.edu/)

[15] Jess (http://www.jessrules.com/)

[16] Pellet (http://clarkparsia.com/pellet/)

[17] Jena (http://jena.sourceforge.net/)

Gesture Recognition with a Focus on Important Actions by Using a Path Searching Method in Weighted Graph

Kazumoto TANAKA

Dept. of Information and Systems Engineering, Kinki University
Higashi-hiroshima, 739-2116, Japan

Abstract

This paper proposes a method of gesture recognition with a focus on important actions for distinguishing similar gestures. The method generates a partial action sequence by using optical flow images, expresses the sequence in the eigenspace, and checks the feature vector sequence by applying an optimum path-searching method of weighted graph to focus the important actions. Also presented are the results of an experiment on the recognition of similar sign language words.

Keywords: *Gesture Recognition, Sign Language, Optical Flow, Appearance-based, Eeigenspace, Weighted Graph, Path Searching.*

1. Introduction

There have been a number of studies on automatic gesture recognition method. One of the main purposes of the gesture recognition is to realize sign language recognition. In order for ordinary persons to communicate with deaf people, the method is needed to translate sign language into natural language.

It is said that there are similar gestures among sign language words, and some part of the gestures (hereafter, they are called "partial actions") play an important role in differentiating those words. For example, in Japanese sign language words, an action of thumb for the word "lend" is very significant for distinguishing the word from the word "return". However, little attention has been given to the gesture recognition that focuses on such partial actions. Thus, this study has developed a recognition method for focusing on important partial actions. The method promises to improve recognition rate of the gestures that have important meanings in their local actions.

Most of studies on gesture recognition have been conducted by utilizing Hidden Markov Models [1][2][3], Dynamic Time Warping [4][5] or Neural Networks [6][7], while the novel recognition method that utilizes a string-matching method based on a path-searching method has been proposed [8]. The target of the recognition method was relatively large motions as a wide hand-swing, not local actions often used in sign language words. However,

by modifying the string-matching method, this study has realized a matching method for focusing on important partial actions to recognize sign language words successfully.

In this paper, a generation method of time-series feature vectors of gestures is described first. The feature vectors are employed in the "modified" string-matching. Next, an optimum path-searching method in weighted graph where the cost of the graph edges corresponding to important partial actions is reduced is proposed for realizing our matching method. Also presented are the results of an experiment on the recognition of similar sign language words.

2. Gesture recognition by using a path searching method in weighted graph

Since the interest in this study rests on the effects of the matching method by which feature vectors of important partial actions are focused to distinguish similar gestures, we will proceed to discuss single gestures as the study subject. Thus, the author will have the gesture be in a stationary state at the start and end of the action and will not address any special word spotting.

In general, gesture recognition methods utilize time-series feature vectors for pattern matching in a gesture dictionary. The way of feature vector composition can be divided broadly into two categories: one utilizes geometric characteristics in images (model-based method) and another utilizes appearance characteristics of images (appearance-based method) [9][10]. The appearance-based method has merits in representing objects that are difficult to express geometric characteristics. In the study, feature vectors of complex optical flow images from sign language gestures are composed by using the appearance-based method.

As known well, appearance based methods often utilize the eigenspace method for reducing dimensions of the feature vector. The feature vector generation method by using the appearance-based method with the eigenspace is described in 2.1 and 2.2. The method of gesture representation for constructing gesture dictionary and the matching method by modifying the string-matching method are described in 2.3 and 2.4 respectively.

2.1 Generation of partial action sequence

The method generates several partial action images from the optical flow images which are calculated from motion images by using a local correlation operation. This study established the criteria for the generation as the position of flow vectors and the directional change. The generation procedure is given in the following:

Step 1: A raster scan is performed to search already labeled flow vectors on the optical flow image, Image_i (the suffix i is the frame number; the initial value is 1). If such a flow vector is found, labeling by 8-neighbours will be conducted until unlabeled vector is not found within the connected component of the flow. However, if the angle between the labeled vector and its neighbor vector exceeds a given threshold, no labeling is conducted. Once this raster scan is completed, the next raster scan will be performed on the Image_i to search unlabeled flow vectors. If an unlabeled flow vector is found in the scan, new label will be attached and, in a similar manner as the above, labeling by 8-neighbours will be conducted.

Step 2: If the angle between the flow vector at the coordinate (m, n) of Image_i v:=(vx, vy) and the flow vector u at the coordinate (m+vx*dt, n+vy*dt) of Image_i+1 is less than the threshold, the label same as that of v will be attached to u. Here, dt is a sampling time span. This is performed to all flow vectors of Image_i.

Step 3: When i:=i+1 and i is not over the last number of the image series, one will return to Step 1.

Step 4: Each group of flow vectors that have the same label is extracted and then is put in a new frame as a new optical flow image (see Fig. 1). This is performed to all labels at each optical flow image.

Step 5: For all the images created in Step 4, if the number of the images that have the same labeled flow vectors is less than a given threshold, the images will be removed for noise canceling.

Step 6: The images that have the same labeled flow vectors are superimposed on the first image among them in time order at each label. As a result, the images are combined into one image at each label (see Fig. 1). The combined image is called partial action image in the paper.

Step 7: The partial action sequence is generated by arranging the partial action images in label order.

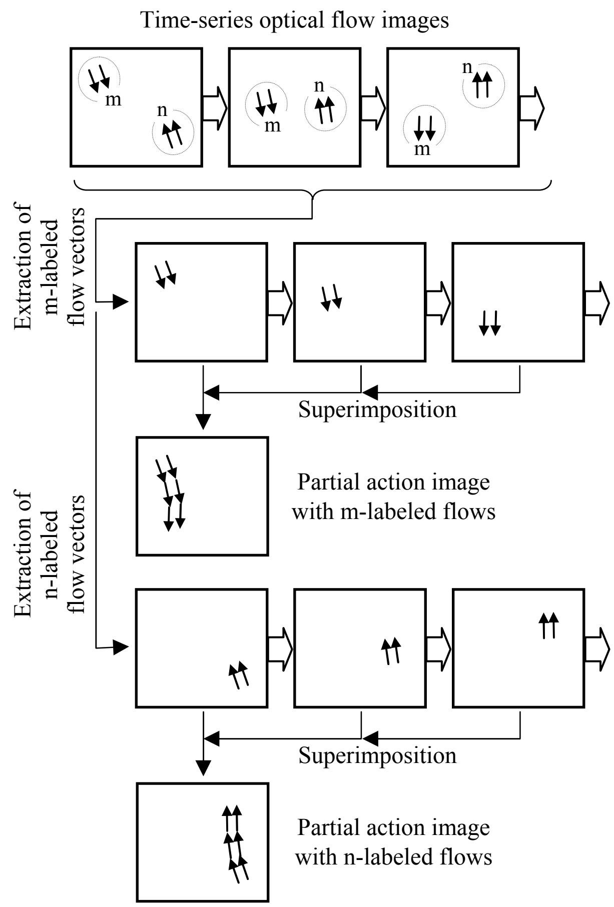

Fig. 1: Generation of partial action image

2.2 Generation of feature vector sequence by the eigenspace method

Because the eigenspace method is sensitive to any position shift of the subject for recognition, it was based on the premise that the gesture would be performed in a nearly identical position inside the window set in the image.

First, the operation described in 2.1 is performed for all dynamic images of each gesture that are prepared for constructing the gesture dictionary. After the operation, for each partial action image, the method generates a vector v by arranging flow vector components in raster scanning order on a partial action image and obtains a set of vector v. The components of the flow vector at each pixel are laid out from component x to component y in the vector.

Next, from all the vector v obtained, the matrix V:= [v_1, v_2, …, v_N] is obtained. The eigenspace based on the eigenvectors that correspond to the upper-ranked number k of the eigenvalue is determined, by solving the characteristic equation (1) of the covariance matrix U of the matrix which is obtained by subtracting the mean vector of the column vectors of the matrix V from each column vector.

$$\lambda_i \cdot e_i = U \cdot e_i \tag{1}$$

From this, each vector v of the partial action will be projected by Formula (2) to calculate the feature vector u on the eigenspace.

$$u = [e_1, e_2, \cdots, e_k]^T (v - \bar{v}) \tag{2}$$

Lastly, the feature vector sequence on the eigenspace <u_1, u_2, …, u_N> can be obtained from the partial action sequence.

2.3 Gesture representation for constructing gesture dictionary

Each gesture in the gesture dictionary is represented by a cluster sequence on the eigenspace. Each cluster is formed by plural patterns of a feature vector that are obtained from a variation (discrepancy) of a partial action to cope with the problem of the variation. Each cluster is approximated by the k-dimensional normal distribution by Formula (3).

$$f(x;\mu,\Sigma) = \frac{1}{\sqrt{2\pi}^{k}\sqrt{|\Sigma|}} exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right] \tag{3}$$

Here, μ is the mean vector of the cluster, Σ is the covariance matrix of the distribution of the cluster. Thus, each gesture in the dictionary will be represented as a sequence of the cluster c represented by the k-dimensional normal distribution as seen in Fig. 2.

$$\{c_1, c_2, \cdots\} \equiv \{\overbrace{f(x;\mu^1,\Sigma^1)}^{\text{Normal distribution of partial action } P_1},\ \overbrace{f(x;\mu^2,\Sigma^2)}^{\text{Normal distribution of partial action } P_2},\ \cdots\}$$

Fig. 2: Gesture representation in the dictionary

2.4 Matching method by a path-searching method in weighted graph

First, Fig. 3 shows the outline of the matching method [8] of one-dimensional feature sequences by using the string-matching method [11] based on a path-searching method. The graph in Fig. 3 have X{A, B, C, A, E, F, G} of the one-dimensional sequence on the horizontal axis and Y{A, H, C, I, F, J} on the vertical axis, and a white circle is attached to the intersection where sequential elements match each other. Added from the intersection is an edge in the lower left diagonal direction. For each edge, a cost is established; the cost is 0 for the added edge in the diagonal direction and the rest are all 1. At this time, obtaining the minimum cost path from the node V_{00} to V_{76}, {A, C, F}, which corresponds to the white circle on the edge in the diagonal direction included on that path, is LCS (Longest Common Subsequence) of X and Y, thus the number of elements of LCS *length*(*LCS*(X, Y)) is 3. From this and Formula (4), the degree of similarity *Sim*(X, Y) can be obtained. Here, the method uses Dijkstra's method to obtain LCS.

$$Sim(X,Y) := \frac{length(LCS(X,Y))}{max(length(X), length(Y))} \tag{4}$$

Next, the matching method for enabling focusing on important partial actions is described in the followings:

Regarding the sequence of the cluster in the dictionary and the feature vector sequence of recognition subjects, individual elements (corresponding to partial actions) are laid out composing a graph G_{PQ} as in Fig. 3. If the cluster sequence P and the feature vector sequence Q are

$$P := \{c_1, c_2, \cdots, c_P\} \tag{5}$$

$$Q := \{u_1, u_2, \cdots, u_Q\} \tag{6}$$

respectively, which elements will match each other are determined by the thresholding Mahalanobis distance to be defined in Formula (7).

$$d \equiv \sqrt{\left(u_i - \mu^{c_j}\right)^T {\Sigma^{c_j}}^{-1}\left(u_i - \mu^{c_j}\right)} \tag{7}$$

Here, μ^c is the mean vector of the cluster c, Σ^c is a covariance matrix of the distribution of the cluster. From this matching result, the edge in the diagonal direction is added to G_{PQ}, the cost of 1 or 0 is allocated for each edge according to the rules.

The focus on important partial action will be conducted as follows: Let a cluster c_k in the sequence P correspond to an important partial action. If P is laid down on the horizontal axis to G_{PQ}, the cost of the edge will be changed by Formula (8).

$$\text{For all } j : j \in \{0,1,\cdots,n\} \text{ do} \quad cost\big((V_{k-1\,j}, V_{k\,j}) \big) := m + n \tag{8}$$

Here, m and n are the numbers of elements of P and Q respectively, V_{ij} is any node in the graph. From this, if Q has an element that is included in c_k, restrictions can be established so that the edge in the diagonal direction toward that matching node can be selected as a path. In the example in Fig. 3, if the fourth element of Sequence X, Element A, corresponds to an important partial action, LCS is {A, F} and *length*(*LCS*(X, Y)) will be 2.

From the graph composed in the above, one can obtain *LCS*(P, Q) using the Dijkstra method and search in the dictionary based on *Sim*(P, Q). In the study, programming was done by the creator of the dictionary designating the important partial actions and attaching a flag to that data.

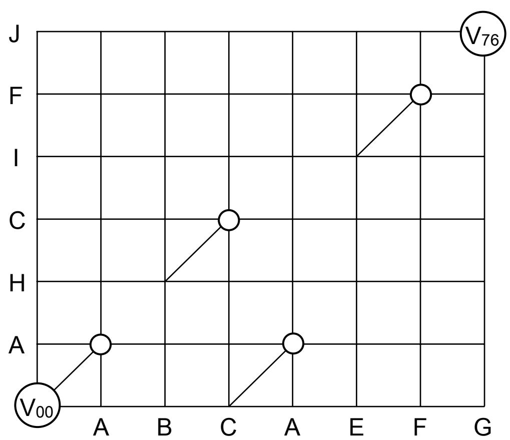

Fig. 3: A Graph for the matching method of 1D feature sequences

3. Experiment

To verify the efficacy of this method, the study conducted a recognition test on Japanese sign language words. The words used in this experiment included "say," "order," "return (home)," "lend." Of them, "say and order" and "return (home) and lend" are very similar in action respectively. Two subjects made the action 30 times for each word and the optical flow images were obtained to create a gesture dictionary for basic 16 words including the four words. The eigenspace was decided as 4-dimensional space to represent each feature vector of the words.

The equipment used for image processing was a personal computer (CPU: Intel(R) Core(TM)2 Duo CPU, 2.20GHz; system memory 3.07GB; OS: Windows XP). The library of OpenCV was used for image processing. The resolution for image processing was 320 x 240. The entire upper body of the subject was imaged in front of a blackout curtain. The video rate was 60fps.

Fig. 4 shows an example of the optical flow image when performing a gesture of "lend". An extracted optical flow image of the thumb and that of the hand without the thumb are also shown in the figure. The partial action of the thumb is an important feature to distinguish "lend" from "return (home)." An example of the partial action image of the thumb is shown in Fig. 5. The partial action image was generated by combining the same labeled optical flow images of the thumb.

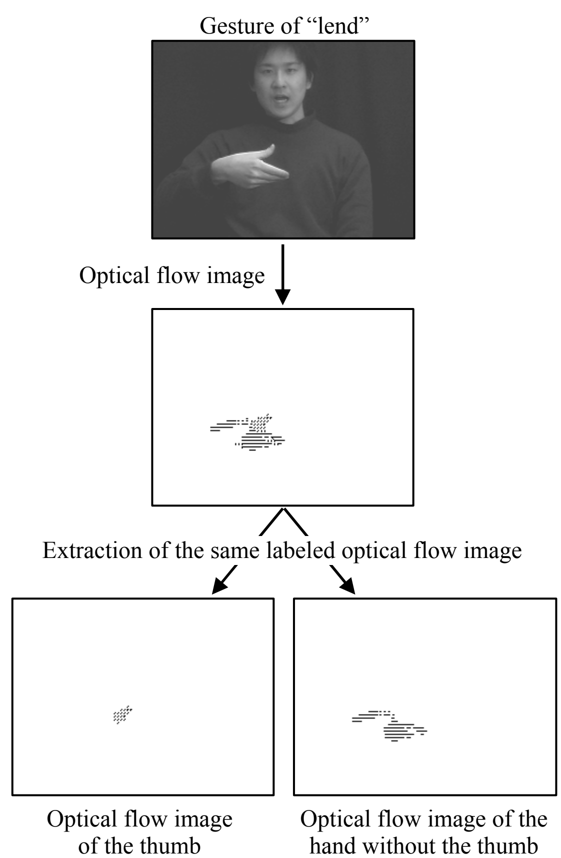

Fig. 4: Optical flow images when performing a sign language word

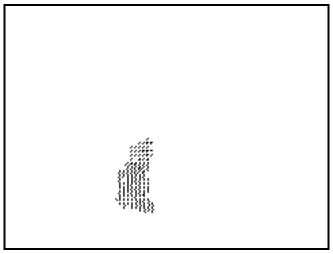

Fig. 5: Partial action image of the thumb

For the recognition experiment, a total number of four people, including two subjects at the time of creating the gesture dictionary, performed a gesture 20 times per word. Table 1 shows recognition rates of the four words and the average recognition rate of the basic 16 words actions by using the proposed method. Table 2 shows recognition rates when important features were not focused. A and B in the tables denote the subject group and the other group, respectively, at the time of creating the dictionary.

Table 1: Recognition rates with the proposed method

	A	B
"say"	69.0%	63.5%
"order"	70.5%	65.0%
"return"	77.0%	70.5%
"lend"	76.0%	67.5%
Average	79.5%	68.5%

Table 2: Recognition rates without important feature focusing

	A	B
"say"	58.0%	51.0%
"order"	60.5%	49.0%
"return"	62.5%	52.5%
"lend"	64.5%	57.0%
Average	76.5%	65.5%

4. Discussion

The results of the experiment show that the recognition rate with focused important features was greater than that without focusing. Thus, the efficacy of this method was confirmed. Furthermore, the following points were clarified:

The recognition rate of "say" and "order" was lower than that of other words. The difference between the two words was the difference in the action near the starting location. However, since the variance of the starting location was great, the matching of this important partial action often failed. Thus, it is needed to improve the robustness against the location shift for the future work.

The reason for the recognition rate of Group B being lower than that of Group A was the difference in individual "habits." For this reason, it will be necessary to obtain gesture patterns from a large number of people and improve the model accuracy of the partial action cluster. For the model, the distribution expression of "habits" by Gaussian Mixture Model can be used.

Overall recognition rates are lower than those in the previous studies of sign language word recognition. This is because the study did not use shapes of the hands and fingers, which are important pieces of information. However, the purpose of this study was to make it possible to recognize gestures that have important local actions. Therefore, in this respect, the author believes that the purpose has been achieved.

5. Conclusion

For the identification of similar gestures, the paper proposed the gesture recognition method with a focus on important local actions. The method generates the partial action sequence by using optical flow images, expresses the sequence in the eigenspace, and checks the feature vector sequence by applying the optimum path-searching method of weighted graph. The paper also showed the efficacy of this method by conducting a recognition test on similar sign language words. For the future, in addition to the improvement on the robustness against the location shift, the author plan to tackle the issue of employing this method for sign language and gestures that include body parts other than hands (shaking of the head, etc). Future work will involve further verification of the method with various sign language words.

Acknowledgments

This work was partially supported by a Grant-in-Aid for Scientific Research (c) 20500856 from the Japan Society for the Promotion of Science.

References

[1] S. Yang and R. Sarkar, "Gesture Recognition Using Hidden Markov Models from Fragmented Observations", in IEEE

Computer Society Conference on Computer Vision and Pattern Recognition, 2006, Vol. 1, pp. 766-773.

[2] C. Vogler and D. Metaxas, “Handshapes and Movements: Multiple-Channel American Sign Language Recognition”, Lecture Notes in Computer Science, Vol. 2915, 2004, pp. 247-258.

[3] T. Starner and A. Pentland, “Visual Recognition of American Sign Language Using Hidden Markov Models”, in International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 189-194.

[4] J. F. Lichtenauer, E. A. Hendriks and M. J. T. Reinders, “Sign Language Recognition by Combining Statistical DTW and Independent Classification”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 11, 2008, pp. 2040-2046.

[5] J. Alon, V. Athitsos and et al., “Simultaneous Localization and Recognition of Dynamic Hand Gestures”, in IEEE Workshop on Motion and Video Computing, 2005, Vol. 2, p.254-260.

[6] Q. Munib, M. Habeeb and et al. “American Sign Language (ASL) Recognition Based on Hough Transform and Neural Networks”, Expert Systems with Applications Vol. 32, 2007, pp.24-37.

[7] M. V. Lamar, M. S. Bhuiyan and A. Iwata, “Hand Gesture Recognition Using T-CombNET: A New Neural Network Model”, IEICE Transactions, Vol.E83-D, No.11, 2000, pp.1986-1995.

[8] K. Koara, A. Nishikawa and et al. “Gesture Recognition Based on 1-Dimensional Encoding of Motion Changes”, in the 10th International Conference on Advanced Robotics, 2001, pp.639-644.

[9] T. B. Moeslund and E. Granum, “A Survey of Computer Vision-Based Human Motion Capture”, Computer Vision and Image Understanding, Vol.81, 2001, pp. 231–268.

[10] A. Bobick and J. Davis, “An Appearance-Based Representation of Action”, in the 13th International Conference on Pattern Recognition, 1996, vol. 1, pp.307.

[11] E. W. Myers, “An O(ND) Difference Algorithm and Its Variations”, Algorithmica Vol.1, Vol. 2, 1986, pp.251-266.

Kazumoto T. is an associate professor of Faculty of Engineering, Kinki University, Japan. He received B.S. degree from Chiba University in 1981, and Ph.D. degree from Tokushima University in 2002. In 1981 he joined Mazda Motor Corporation, he was engaged in research and development of robot vision technology. His current interest is mainly gaming technology. He is a member of Japanese Society for Information and Systems in Education and Institute of Electronics, Information and Communication Engineers.

Lexicographic Multi-objective Geometric Programming Problems

Dr.A. K. Ojha[1] and K. K. Biswal[2]

[1]School of Basic Sciences, IIT Bhubaneswar, Orissa, Pin-751013, India

[2]Department of Mathematics, CTTC Bhubaneswar, B-36, Chandaka Industrial Area, Bhubaneswar, Orissa, Pin-751024, India

Abstract

A Geometric programming (GP) is a type of mathematical problem characterized by objective and constraint functions that have a special form. Many methods have been developed to solve large scale engineering design GP problems. In this paper GP technique has been used to solve multi-objective GP problem as a vector optimization problem. The duality theory for lexicographic geometric programming has been developed to solve the problems with posynomial in objectives and constraints.

Keywords: *Lexicographic minimization, Geometric programming, duality theory, vector minimization and vector maximization.*

1. Introduction

In many real-life optimization problems, multiple objectives have been taken into account, which may be related to the economical, social and environmental aspects of optimization problems. The multiple objectives are usually incommensurate and in conflict with one another. In general, a multiple objective optimization problem does not have a single solution that could optimize all objectives simultaneously. It never search for optimal solution but for efficient solution that can best suited compromise solution to all multiple objectives. Though the geometric programming technique due to Duffin et al.[3] helps to solve various types of nonlinear single objective posynomial problems but there are very few work have been made in this direction to solve multiple objective GP problems. Lexicographic optimization approach is one such technique to handle multiple objective GP problem. Several mathematical and game theoretic applications of nonlinear lexicographic optimizations are reported by Behringer[1]. Impressed upon the work of Behringer, Nijkamp[6] applied the lexicographic optimization technique in a land use problem for industrial activities in a newly created industrial area in a Rhine-delta region near Rotterdam. Application of linear lexicographic due to Isermann[4] and Turnovec[8] strengthen to handle scalar valued optimization problems. Biswal[2] used fuzzy programming[10] to solve multi-objective geometric problem where as lexicographic order and duality has been studied by Martinez[5]. In this paper we have applied lexicographic geometric programming technique to solve special type of multi-objective optimization problem.

The organization of the paper is as follows: Following introduction the definition of multi-objective geometric programming and lexicographic optimization have been discussed in Section-2 and 3 respectively. Definition of lexicographic geometric programming has been discussed in Section 4 and the numerical examples have been incorporated in Section 5. Finally the conclusion has been presented in Section 6.

2. Multi-objective geometric programming

A multi-objective geometric programming problem can be defined as:

Find $x = (x_1, x_2, \dots, x_n)^T$ so as to

$$\min : g_{k0}(x) = \sum_{t=1}^{T_{ko}} C_{k0t} \prod_{j=1}^{n} x_j^{a_{k0tj}} \ , k = 1,2,\dots, p \tag{2.1}$$

subject to

$$g_i(x) = \sum_{t=1}^{T_i} C_{it} \prod_{j=1}^{n} x_j^{d_{itj}} \leq 1 \,, i = 1,2,...,m \quad (2.2)$$

$$x_j > 0, j = 1,2,...,n \qquad (2.3)$$

where C_{k0t} for all k and t are positive real numbers and d_{itj} and a_{k0tj} are real numbers for all i, k, t, j.
T_{k0}=number of terms present in the k^{th} objective function.
T_i =number of terms present in the i^{th} constraint.
In the above multi-objective geometric programming problem there are p number of minimization type objective functions, m number of inequality type constraints and n number of strictly positive decision variables.
Let us define $F(x)=\{g_{10}(x), g_{20}(x),...., g_{p0}(x)\}$
and $G(x)=\{g_1(x)-1, g_2(x)-1,...., g_m(x)-1\}$
Now the above optimization problem can be rewritten as:

$$lex \min: \{F(x): x \in R^n, G(x) \leq 0^m\} \quad (2.4)$$

which is called lexicographic geometric programming (LGP) problem.

3. Lexicographic optimization problem

Let us denote O^n an n-dimensional zero vector and $O^{m\times n}$ an m×n zero matrix. An inequality of the type $x \geq O^n$ means $x \geq O^n$, but $x \neq O^n$
A vector $x \in R^n$ is said to be lexicographically non negative if either x = O^n or its first non-zero component is positive and we denote it by $lex\, x \geq O^n$
Similarly a non zero vector $x \in R^n$ is said to be lexicographically non positive if its first non zero component is negative and it is denoted by $lex\, x < O^n$
An m×n matrix A is called lexicographically non negative if all its columns are lexicographically non negative and we denote it as

$$lex\, A \geq O^{m\times n}$$

In the similar manner we can define lexicographically non positive and lexico-graphically negative vector and its corresponding matrix.
Let F be an p-dimensional vector valued function defined on R^n and $X \subset R^n$.
A vector $x^* \in X$ is said to be a lexicographically minimal point of F with respect to X if for any $x \in X$ such that $lex\, F(x^*) \leq F(x)$

The problem of finding lexicographically minimum point of F with respect to X is called lexicographically minimizing problem which is denoted as :

$$lex \min: \{F(x): x \in X\} \quad (3.1)$$

Similarly a point $\bar{x} \in X$ is said to be lexicographically maximum point of F with respect to X if $lex\, F(\bar{x}) \geq F(x)$ and this problem is called lexicographically maximizing problem given by

$$lex \max: \{F(x): x \in X\} \quad (3.2)$$

If we assume

$$X = \{x \in R^n : G(x) \leq O^m\} \quad (3.3)$$

then the lexicographically minimizing problem can be defined as,

$$lex \min: \{F(x): G(x) \leq O^m\} \qquad (3.4)$$

A lexicographically minimum point of F with respect to (3.3) is called an optimal solution to the problem given by (3.4).
The problem given by (3.4) is called convex optimization if all the components of F and G are convex functions.

4. Lexicographic geometric Programming

The lexicographic multi-objective geometric programming defined by

$$lex \min: \{F(x): x \in R^n, G(x) \leq O^m\}$$

where the functions are defined by (2.1), (2.2) and (2.3).
Now we will prove the following theorem for the existence of unique optimal solution of the lexicographic optimization GP problem.

Theorem:- If the primal problem of the geometric programming is consistent and its dual program has a maximizing point with strictly positive components then the primal problem of geometric programming has a unique optimal solution if and only if the rank of its exponent matrix is equal to the number of columns.

Proof :-From the duality theorem due to Duffin[3], we know that for each optimization point x^* for the primal problem there exist a maximizing point w^* of dual program which is given by the following equations.

$$C_{k0t} \prod_{j=1}^{n} x_j^{a_{k0tj}} = w^*_{k0t} v(w^*), k = 1,2,...p, t = 1,2,....T_{k0}. \quad (4.1)$$

$$C_{it}\prod_{j=1}^{n} x_j^{d_{itj}} = \frac{w_{it}^*}{\lambda_i(w^*)}, i=1,2,...m, t=1,2,....T_i. \quad (4.2)$$

Taking logarithm on both sides of equation (4.1) and (4.2) we have

$$\ln(x_1^{a_{k0t1}} x_2^{a_{k0t2}} ... x_n^{a_{k0tn}}) = \ln\frac{w_{k0t}v(w^*)}{C_{k0t}}, k=1,2,....p, t=1,2,...T_{k0}$$

$$\ln(x_1^{d_{it1}} x_2^{a_{it2}} ... x_n^{a_{itn}}) = \ln\frac{w_{it}^*}{C_{it}\lambda_i(w^*)}, i=1,2,...m, t=1,2,...T_i.$$

which can be expressed as

$$\mathrm{a}_{k0t1}\ln \mathrm{x}_1 + \mathrm{a}_{k0t2}\ln \mathrm{x}_2 + \ldots + \mathrm{a}_{k0tn}\ln \mathrm{x}_n = \ln(\beta_{k0t}); \quad \mathrm{k}=1,2,....\mathrm{p};\ \mathrm{t}=1,2,...,\mathrm{T}_{k0} \quad (4.3)$$

And

$$\mathrm{d}_{it1}\ln \mathrm{x}_1 + \mathrm{d}_{it2}\ln \mathrm{x}_2 + ... + \mathrm{d}_{itn}\ln \mathrm{x}_n = \ln(\beta_{it}); \quad \mathrm{i}=1,2,....\mathrm{m};\ \mathrm{t}=1,2,...\mathrm{T}_i \quad (4.4)$$

where $\beta_{k0t} = \dfrac{w_{kot}^* v(w^*)}{C_{k0t}},\ t=1,2,....T_{k0}$

and

$$\beta_{it} = \frac{w_{it}^*}{C_{it}\lambda_i(w^*)},\ t=1,2,....T_i,\ i=1,2,....,m.$$

Now by substituting
$z_j = \ln x_j$; $j = 1, 2, ..., n$

$$\ln(\beta_{k0t}) = \gamma_{k0t}, k=1,2,....p,\ t=1,2,.....,T_{k0}$$

and $\ln\beta_{it} = \gamma_{it}, i=1,2,....m,\ t=1,2,....,T_i$

the equations (4.3) and (4.4) reduces to

$$a_{k0t1}z_1 + a_{k0t2}z_2 + ... + a_{k0tn}z_n = \gamma_{k0t},\ t=1,2,...,T_{k0} \quad (4.5)$$

and

$$d_{it1}z_1 + d_{it2}z_2 + ... + d_{itn}z_n = \gamma_{it},\ i=1,2,...,m,\ t=1,2,...,T_i \quad (4.6)$$

If we assume $T = \sum_{i=1}^{m} T_i$ then we will have T number of equations and n variables, which is exactly the dimension of the exponent matrix.

Writing the equations in the matrix form we have

$$Az = \gamma \quad (4.7)$$

where A is the exponent matrix of dimension T×n and T ≥ n. From the basic knowledge of linear algebra the system of equations (4.7) has a unique solution if and only if the rank of the matrix A is equal to the number of its columns. With the substitution

$$e^{z_j} = x_j \text{ or } z_j = \ln x_j,\ j=1,2,...,n$$

Our lexicographic optimization problem can be expressed as.

$$lex\ \min : \{F(z) : G(z) \le 1_m\} \quad (4.8)$$

where F(z) is an p-dimensional vector valued function with

$$g_{k0}(z) = \sum_{t=1}^{T_{k0}} C_{k0t} e^{\sum_{j=1}^{n} a_{k0tj} z_j}, \text{ k=1,2,….,p}$$

and $g_i(z) = \sum_{t=1}^{T_i} C_{it} e^{\sum_{j=1}^{n} d_{itj} z_j}$, i=1,2,….,m

Due to the monotonicity of logarithm function the problem (4.8) can be expressed as:

$$lex\ \min : \{\ln F(z) : \ln G(z) \le O^m\} \quad (4.9)$$

with

$$\ln F(z) = \{\ln g_{10}(z), \ln g_{20}(z), ..., \ln g_{p0}(z)\}$$

and $\ln G(z) = \{\ln g_1(z), \ln g_2(z), ..., \ln g_m(z)\}$

Introducing new variables

$$w_{kot} = \frac{C_{k0t} e^{\sum_{j=1}^{n} a_{k0tj} z_j}}{\sum_{t=1}^{T_{k0}} C_{k0t} e^{\sum_{j=1}^{n} a_{k0tj} z_j}}, \text{ k=1,2,…,p} \quad (4.10)$$

And $w_{kit} = u_{ki} \dfrac{C_{it} e^{\sum_{j=1}^{n} d_{itj} z_j}}{\sum_{t=1}^{T_i} C_{it} e^{\sum_{j=1}^{n} d_{itj} z_j}}$, k=1,2,….p, i= 1,2,….,m (4.11)

After suitable transformation [3] the dual problem associated with the lexicographic geometric programming problem can be obtained.

$$lex\ \max : V(w)$$

such that $\sum_{t=1}^{T_{k0}} w_{k0t} = 1$, k=1,2,….,p (4.12)

$$\sum_{t=1}^{T_{k0}} w_{k0t} a_{k0tj} + \sum_{i=1}^{m}\sum_{t=1}^{T_i} w_{kit} d_{itj} = 0,\ k=1,2,....,p,\ j=1,2,....,n, i=1,2,....,m$$

(4.13)

$$u_{ki} = \sum_{i=1}^{T_i} w_{kit}, \quad k=1,2,\ldots,p, \quad i=1,2,\ldots,m \qquad (4.14)$$

where $V_k(w)$ is the p-dimensional vector valued function of the form

$$V_k(w) = \sum_{t=1}^{T_{k0}} \left(\frac{C_{k0t}}{w_{kot}}\right)^{k0t} \prod_{i=1}^{m}\prod_{t=1}^{T_i}\left(\frac{d_{it}}{w_{kit}}\right)^{w_{kit}} \prod_{i=1}^{m} u_{ki}^{\;u_{ki}}, \quad k=1,2,\ldots,p \qquad (4.15)$$

According to the theory of geometric programming problem (4.12) is called the normality conditions, (4.13) is called orthogonality condition and (4.15) is the dual function.

The number $d = \sum_{k=1}^{p} T_{k0} + \sum_{k=1}^{m} T_i - n - 1$ is the degree of difficulty.

As the problem with zero degree of difficulty is easily solvable then the dual problem can be solved to get the maximizing vector w^*. Since the vector w^* is the unique solution to the dual constraints, it is also the maximizing vector for the dual problem.

Using this dual optimizing vector the optimal solution x^* to the primal problem can be determined by using following relationships.

$$C_{k0t} x_1^{a_{k0t1}} x_2^{a_{k0t2}} \ldots x_n^{a_{k0tn}} = w_{k0t}^* V_k(w_{k0t}^*), \quad k=1,2,\ldots,p \qquad (4.16)$$

$$C_{it} x_1^{d_{it1}} x_2^{d_{it2}} \ldots x_n^{d_{itn}} = \frac{w_{it}^*}{u_{ki}^*}, \quad k=1,2,\ldots,p;\ i=1,2,\ldots,m \qquad (4.17)$$

where $u_{ki}^* = \sum_{i=1}^{T_i} w_{kit}^*$, $k=1,2,\ldots,p$ (4.18)

5. Numerical Example

Example: Let us consider a numerical example

$$lex\min F(x) = \left\{ g_{10}(x) = x_1^{-1}x_2^{-1}x_3^{-2},\ g_{20}(x) = x_1^{-1}x_2^{-3}x_3^{-5} + x_1^{-1}x_2^{-1} \right\}$$

subject to $x_1 x_2 x_3^2 + x_2 x_3 \le 10$

$x_1 x_3 \le 2$

$x_1, x_2, x_3 > 0$

Solution Procedure of LGP problems.

Step 1:
At first the objective functions of the multi-objective GP problem are ranked according their priority. Let us assume that the first objective function is in priority one i.e. P_1 and the second objective function in priority 2 i.e. P_2 and so on, and p^{th} objective function in priority p i.e P_p.

Step 2:
Then first objective function $g_{10}(x)$ is minimized subject to all the original constraints. Let the minimum of the first objective be $g_{10}^{(1)}$ at $x^{(1)}$. Then we move to step 3.

Step 3:
Then the second objective function $g_{20}(x)$ is minimized subject to the original constraints with one additional constraint, i.e. $g_{10}(x) \ge g_{10}^{(1)}$

Let the minimum value of the second objective function be $g_{20}^{(2)}$ at $x^{(2)}$. Then we move to next step.

Step 4:
In step 4 third objective function $g_{30}(x)$ is minimized subject to the original constraint with two additional constraints, i.e. $g_{10}(x) \ge g_{10}^{(1)}$

$g_{20}(x) \ge g_{20}^{(2)}$

Same procedure is repeated for all the objective functions.

Step 5:
Finally, the last objective function is minimized subject to all the original constraints plus additional p-1 constraints.

$g_{10}(x) \ge g_{10}^{(1)}$, $g_{20}(x) \ge g_{20}^{(2)}$, $\ldots g_{p-1,0}(x) \ge g_{p-1,0}^{(p-1)}$

Let A be the exponent matrix.

$$A = \begin{pmatrix} -1 & -1 & -2 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

The rank of the matrix is 2 which is less than the number of columns.

The dual program of the function $g_{10}(x)$ is as

$$\max : V(w) = \left(\frac{1}{w_{101}}\right)^{w_{101}} \left(\frac{1/10}{w_{111}}\right)^{w_{111}} \left(\frac{1/10}{w_{112}}\right)^{w_{112}} \left(\frac{0.5}{w_{211}}\right)^{w_{211}} \left(w_{111} + w_{112}\right)^{w_{111}+w_{112}}$$

Subject to: $w_{101} = 1$

$$-w_{101} + w_{111} + w_{211} = 0$$

$$-w_{101} + w_{111} + w_{112} = 0$$

$$-w_{101} + 2w_{111} + w_{112} + w_{211} = 0$$

The solution of the dual program gives $w_{101} = 1, w_{111} = 0.666...,$ $w_{112} = 0.3333334, w_{211} = 0.3333334$ and the mean value of dual $v(w^*) = 0.15$. Using the primal dual relationship we have the optimal solution of $g_{10}(x)$ are $x_1^* = 0.9086967$, $x_2^* = 1.514494$, $x_3^* = 2.200954$ and its primal optimal solution is 0.15.
Similarly the dual program of $g_{20}(x)$ is defined as,

$$\max : V(w) = \left(\frac{1}{w_{201}}\right)^{w_{201}} \left(\frac{1}{w_{202}}\right)^{w_{202}} \left(\frac{1/10}{w_{211}}\right)^{w_{211}} \left(\frac{1/10}{w_{212}}\right)^{w_{212}} \left(\frac{0.5}{w_{211}}\right)^{w_{211}} \left(w_{211} + w_{212}\right)^{w_{211}+w_{212}}$$

Subject to: $w_{201} + w_{202} = 1$

$$-w_{201} - w_{202} + w_{211} + w_{221} = 0$$

$$-3w_{201} - w_{202} + w_{211} + w_{221} = 0$$

$$-5w_{201} + 2w_{211} + w_{212} + w_{221} = 0$$

The solution of the problem gives its optimal value $V_2(w^*) = 0.4316470 \times 10^{-1}$ with $w_{201}^* =$ 0.6666667, $w_{202}^* =$ 0.3333…, $w_{211}^* =$ 1, $w_{212}^* =$ 1.3333…, $w_{221}^* = 0$
Using primal dual relationship we have $x_1 = 3.020273$, $x_2 = 23.01163$, $x_3 = 0.2483217$

6. Conclusions

In this chapter solution procedure of lexicographic GP has been presented. Unless the objective functions ranked properly, lexicographic solution may not be acceptable to a design engineer. If there are p number of objective functions one may formulate p! no of ways priority, which is a very difficult task for a design engineer. Also sometimes more than one objective functions remain in a priority. Unless the priority is proper solution of a real life problem gives some abnormal result. To set the proper ranking, method of Analytic Hierarchy Process (AHP) may be adopted. Two popular method of AHP by Saaty[7] namely row-column Addpotion method and Eigen-value method is used to find the proper ranking (weights) of the objective function. If the weights of the objective function can be estimated, then using the weights multi-objective GP problem can be converted to a single objective GP problem and solved.

References

[1] Behringer,F.A., Lexicographic quasiconcave multiobjective program-ming, Zeitschrift fier operation research 21, 103-116, 1977.
[2] Biswal,M.P., Fuzzy Programming technique to solve multi-objective geometric programming problem, Fuzzy sets and systems 51,67-71 ,1992.
[3] Duffin R.J.,Peterson E.L.and Zener C.M., Programming Theory and Application, Wiley, New York. ,1967
[4] Isermann,H., Linear lexicographic optimization, OR-spektrum 4,223-228. ,1982
[5] Martinez-Legaz,J.E., Lexicographic order and duality in multiobjectve programming, European Journal of operational research 33, 342-348, 1988.
[6] Nijkamp,P. Environmental policy analysis in operational methods and models, Wiley, New York. ,1980
[7] Saaty T.L., A Scaling Method for Priorities in Hierarchical Structures, Journal of Mathematical pshychology, 15, 234-281, 1977
[8] Turnovec F., Lexicographic optimization problems in production scheduling in optimization, Theory, methods, Applications, Vol-I Dum Tech- niky CSVTS Praha, Prague, 295-306, 1985
[9] Zimmermann,H.J. ,Fuzzy programming and linear programming with several objective functions. Fuzzy sets and systems 1, 46-55. ,1978
[10] Zimmermann,H.J. Fuzzy set theory and its applications, 2nd ed Kluwer Academic ,publishers, Dordrecht-Boston, 1990.

High Availability Cluster System for Local Disaster Recovery with Markov Modeling Approach

T.T.Lwin and T.Thein

University of Computer Studies
Yangon, Myanmar

Abstract

The need for high availability (HA) and disaster recovery (DR) in IT environment is more stringent than most of the other sectors of enterprises. Many businesses require the availability of business-critical applications 24 hours a day, seven days a week, and can afford no data loss in the event of a disaster. It is vital that the IT infrastructure is resilient with regard to disruption, even site failures, and that business operations can continue without significant impact. As a result, DR has gained great importance in IT. Clustering of multiple industries standard servers together to allow workload sharing and fail-over capabilities is a low cost approach. In this paper, we present the availability model through Semi-Markov Process (SMP) and also analyze the difference in downtime of the SMP model and the approximate Continuous Time Markov Chain (CTMC) model. To acquire system availability, we perform numerical analysis and SHARPE tool evaluation.

Keywords: *availability, cluster system, local disaster recovery, markov modeling*

1. Introduction

High availability clusters (also known as HA Clusters or failover Clusters) are computer clusters implemented to provide high availability of services. They operate by having redundant computers or nodes which are used to provide service when a system component fails.

A cluster is a collection of computer nodes -- independent, self-contained computer systems working together – to provide a more reliable and powerful system than a single node alone [8]. Clustering has proven to be a very effective method for scaling to larger systems for added performance, as well as providing higher levels of availability and lower management costs. For this reason, software packages such as IBM's RS/6000 Cluster Technology [8] (i.e., Phoenix) and Microsoft's Cluster Services [5] (i.e., Wolf pack) are being used to build high availability systems.

Disaster recovery solutions have gained popularity in the past few years because of their ability to tolerate disasters and to achieve the reliability and availability.

2. Related Work

Hunter [5] described some system characteristics that benefit from clustering and presented a two-node Microsoft Cluster Service (MSCS) cluster configuration and also presented an availability model of that system using Markov modeling techniques.

In [1] they discussed high availability and disaster recovery solutions, and described how HA and DR solutions differ from one another and how they can be combined to provide the highest levels of resiliency for IT infrastructures.

Trivedi et. al [10] described an availability model for a high availability platform using a multi-level hierarchical composition approach that mixes reliability block diagrams and Markov chains, so as to allow detailed behavior to be captured while avoiding state space explosion.

Song et al [9] provided novel solutions with three –key components, availability modeling, model evaluation and data analysis and examined numerical solutions for Markov models on the uniformization method. This paper also presents a monitoring and data analysis framework, which is responsible for failure analysis and availability reconfiguration.

The semi-Markov decision model is a powerful tool in analyzing sequential decision process with random decision epochs [2]. They presented the application of Markov decision process algorithm, a joint optimization of inspection rate and its corresponding maintenance policy are also presented.

3. System Architecture

The architecture is based on an active-passive high availability solution. Each service under high availability needs at least two identical servers: a primary host, on which the service run, one or more secondary hosts, able to recover the application. As a result of failure detection, the active-passive roles are switched. A heartbeat keepalive system is used to monitor the health of the nodes in the cluster. A disaster recovery solution is typically composed of two nodes, one active and one passive. The active node is usually called master or production node, and the passive node is called secondary or standby node. During normal operation, the only working node is the master node; in the event of a node failover or switchover, the standby node takes over the production role, by taking its IP number, and completely replacing the master one.

To maintain the standby node for failover, the standby node contains homogenous installations and applications: data and configurations must also be constantly synchronized with the master node.

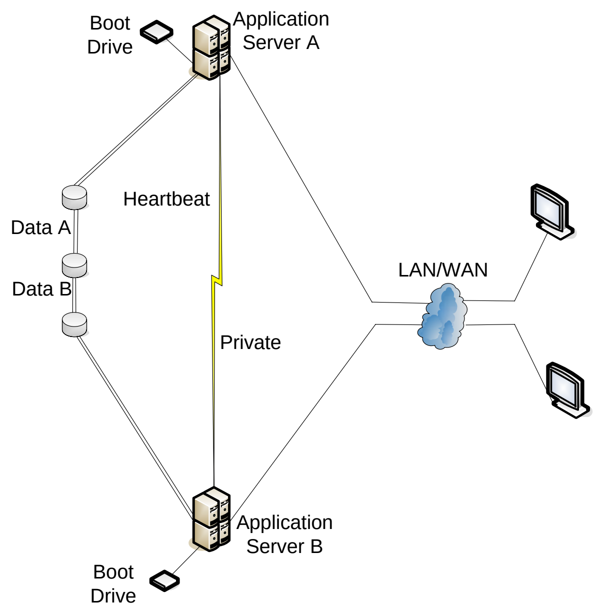

Figure1: System Framework

If a crash occurs and if the data is not restored, it can have devastating consequences for a business. So it is imperative for companies to effectively backup and recover data and protect them from huge losses in productivity and downtime.

In this way, hardware exposure is mitigated through physical hardware redundancy. Clustering provides high availability by protecting against a node failure. However, it does not prevent against storage failures. Given the size of typical cluster environments, multiple hard disks are used to build large storage arrays. In Network and System Administration, when large numbers of any one device are used, failure is expected. When a hard disk fails, application disruption is unavoidable, as all the nodes in the cluster could be using that one particular disk as shared storage which contains all files.

With the widespread use of computers, data is becoming more and more important in human life. But all kinds of accidents and disasters occur frequently. Data corruption and data loss by various disasters have become more dominant, accounting for over 60% [1] of data loss .Recent high-profile data loss has raised awareness of the need to plan for recovery of continuity. Many data disaster tolerance technologies have been employed to increase the availability of data and to reduce the data damage caused by disasters [2].

A true disaster recovery solution is the ability to restore full systems quickly on available computing resources which may be local but may also be remote if the situation dictates and must allow recovery from site-wide disasters. The primary site may be completely down, a secondary site located in a non-affected area would be used to restore services until the primary site comes back online.

4. Modeling and Analysis

We propose the two-component system, one component is considered as active and the other as a standby (spare) unit. The failure rates of the active unit and the standby unit are different, and also the effect of failure of the standby unit is different from that of the active unit. Assuming that, the time to restoration and reboot are exponentially distributed with rate μ and β respectively.

We consider a routine diagnostic that is run every T time units, intended to detect the latent fault of the standby unit. While units' failure and restoration times are exponentially distributed, the routine diagnostic time interval is not a continuous time Markov chain. The model for the system with the diagnostic routine is called a semi-Markov chain. To solve this model, we could crudely approximate the time to the next diagnostic to be exponentially distributed with mean $\frac{T}{2}$ .Descriptions of the state are shown in table (1).

Table (1): State Description for Transitions model

State	Descriptions
1	Both active and spare units are working
2	Protection switch fails to cover the failure of the active unit
3	When active unit fails, protection switch successfully restores service by the standby unit
4	The failure of the standby unit while the active unit is still working is detected immediately
5	The failure of the standby unit is not detected
6	The system is in failure state

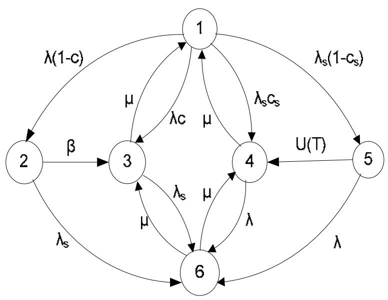

Figure2 : State Transition Model

λ=failure rate of an active unit
λ_s=failure rate of a standby unit
μ=restoration rate of a failed unit
c =coverage probability of an active unit
c_s=coverage probability of a standby unit
T =time units to detect the latent fault of the standby unit

We may compute the steady-state probabilities by first writing down the steady-state balance equations of figure 2 are as follows:

$$\mu P_3 + \mu P_4 = \lambda(1-c)P_1 + \lambda c P_1 + \lambda_s c_s P_1 + \lambda_s(1-c_s)P_1 \quad (1)$$

$$\lambda(1-c)P_1 = (\beta + \lambda_s)P_2 \quad (2)$$

$$\lambda c P_1 + \beta P_2 + \mu P_6 = \mu P_3 + \lambda_s P_3 \quad (3)$$

$$\lambda_s c_s P_1 + \frac{2}{T}P_5 + \mu P_6 = (\mu + \lambda)P_4 \quad (4)$$

$$\lambda_s(1-c_s)P_1 = \left(\frac{2}{T} + \lambda\right)P_5 \quad (5)$$

$$\lambda_s P_2 + \lambda_s P_3 + \lambda P_4 + \lambda P_5 = 2\mu P_6 \quad (6)$$

The conservation equation of figure 2 is obtained by summing the probabilities of all states in the system and the sum of the equation is 1.

$$\sum_{i=1}^{n} P_i = 1 \quad (7)$$

Combining the above-mentioned balance equations with the conservation equations, and solving these simultaneous equations, we acquire the closed-form solution for the system.

$$P_1 = \left[1 + \frac{\lambda(1-c)}{\lambda_s + \beta} - \frac{1}{\mu + \lambda_s}\left(\lambda_s c_s + \frac{\lambda_s(1-c_s)\frac{2}{T}}{\frac{2}{T}+\lambda} - \lambda c - \frac{\lambda(1-c)\beta}{\lambda_s + \beta} \right) + \frac{\lambda_s(1-c_s)}{\frac{2}{T}+\lambda} - \frac{\lambda_s c_s}{\mu} - \frac{\lambda_s(1-c_s)\frac{2}{T}}{\mu\left(\frac{2}{T}+\lambda\right)} + \left(\frac{\mu+\lambda}{\mu+\lambda_s} + 1 + \frac{\mu+\lambda}{\mu} \right) \left(\frac{\lambda(1-c) + \lambda c + \lambda_s c_s + \lambda_s(1-c_s) + \frac{\mu}{\mu+\lambda_s}\left(\lambda_s c_s + \frac{\lambda_s(1-c_s)\frac{2}{T}}{\frac{2}{T}+\lambda} - \lambda c - \frac{\lambda(1-c)\beta}{\lambda_s+\beta} \right)}{\frac{\mu(\mu+\lambda)}{\mu+\lambda_s} + \mu} \right) \right]^{-1} \quad (8)$$

$$P_2 = \left(\frac{\lambda(1-c)}{\lambda_s + \beta} \right) P_1 \quad (9)$$

$$P_3 = \left(\begin{array}{l} \frac{\mu+\lambda}{\mu+\lambda_s} \times \\ \lambda(1-c)+\lambda c+\lambda_s c_s+\lambda_s(1-c_s)+ \\ \frac{\mu}{\mu+\lambda_s} \\ \frac{\left(\lambda_s c_s + \frac{\lambda_s(1-c_s)\frac{2}{T}}{\frac{2}{T}+\lambda} - \lambda c - \frac{\lambda(1-c)\beta}{\lambda_s+\beta} \right)}{\frac{\mu(\mu+\lambda)}{\mu+\lambda_s}+\mu} \\ -\frac{1}{\mu+\lambda_s} \\ \left(\lambda_s c_s + \frac{\lambda_s(1-c_s)\frac{2}{T}}{\frac{2}{T}+\lambda} - \lambda c - \frac{\lambda(1-c)\beta}{\lambda_s+\beta} \right) \end{array} \right) P_1 \tag{10}$$

$$P_4 = \left(\begin{array}{l} \lambda(1-c)+\lambda c+\lambda_s c_s+\lambda_s(1-c_s)+ \\ \frac{\mu}{\mu+\lambda_s} \\ \frac{\left(\lambda_s c_s + \frac{\lambda_s(1-c_s)\frac{2}{T}}{\frac{2}{T}+\lambda} - \lambda c - \frac{\lambda(1-c)\beta}{\lambda_s+\beta} \right)}{\frac{\mu(\mu+\lambda)}{\mu+\lambda_s}+\mu} \end{array} \right) P_1 \tag{11}$$

$$P_5 = \left(\frac{\lambda_s(1-c_s)}{\frac{2}{T}+\lambda} \right) P_1 \tag{12}$$

$$P_6 = \left(\begin{array}{l} \frac{\mu+\lambda}{\mu} \times \\ \lambda(1-c)+\lambda c+\lambda_s c_s+\lambda_s(1-c_s)+ \\ \frac{\mu}{\mu+\lambda_s} \\ \frac{\left(\lambda_s c_s + \frac{\lambda_s(1-c_s)\frac{2}{T}}{\frac{2}{T}+\lambda} - \lambda c - \frac{\lambda(1-c)\beta}{\lambda_s+\beta} \right)}{\frac{\mu(\mu+\lambda)}{\mu+\lambda_s}+\mu} \\ -\frac{\lambda_s c_s}{\mu} - \frac{\lambda_s(1-c_s)\frac{2}{T}}{\mu\left(\frac{2}{T}+\lambda\right)} \end{array} \right) P_1 \tag{13}$$

4.1Semi-Markov Model Analysis

A better approach would be to take the time the next diagnostic to be uniformly distributed over [0, T], resulting in a semi-Markov chain. This is indicated in fig: 2 the transition labeled U (0, T). As occurring in two stages of transitions, the SMP is described by a transition probability matrix P and the vector of sojourn time distributions, **H** (t).

$$H_1 = 1 - e^{-(\lambda+\lambda_s)t} \tag{14}$$

$$H_2 = 1 - e^{-(\beta+\lambda_s)t} \tag{15}$$

$$H_3 = 1 - e^{-(\lambda_s+\mu)t} \tag{16}$$

$$H_4 = 1 - e^{-(\lambda+\mu)t} \tag{17}$$

$$H_5 = \begin{cases} 1-\left(1-\frac{t}{T}\right)e^{-\lambda t}, t < T, \\ 1, t \geq T, \end{cases} \tag{18}$$

$$H_6 = 1 - e^{2\mu t} \tag{19}$$

Let X~EXP (λ) and Y~U (0, T) random variables

$$P(X>Y)=\int_0^T P(X>t)f_Y(t)dt$$

$$=\int_0^T e^{-\lambda t}\frac{1}{T}dt=\frac{1}{\lambda T}\left(1-e^{-\lambda T}\right) \qquad (20)$$

The one-step transition probability matrix P of the DTMC embedded at the time of transitions and the state probabilities of the embedded DTMC are given by the following equations respectively.

$$P=\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}\begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0 & \frac{\lambda(1-c)}{\lambda+\lambda_s} & \frac{\lambda c}{\lambda+\lambda_s} & \frac{\lambda_s c_s}{\lambda+\lambda_s} & \frac{\lambda_s(1-c_s)}{\lambda+\lambda_s} & 0 \\ 0 & 0 & \frac{\beta}{\beta+\lambda_s} & 0 & 0 & \frac{\lambda_s}{\beta+\lambda_s} \\ \frac{\mu}{\lambda_s+\mu} & 0 & 0 & 0 & 0 & \frac{\lambda_s}{\lambda_s+\mu} \\ \frac{\mu}{\lambda+\mu} & 0 & 0 & 0 & 0 & \frac{\lambda}{\lambda+\mu} \\ 0 & 0 & 0 & \frac{1}{\lambda T\left(1-e^{-\lambda T}\right)} & 0 & 1-\frac{1}{\lambda T}\left(1-e^{-\lambda T}\right) \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix} \end{array} \qquad (21)$$

$$v=\left[v_{1,1},v_{1C},v_{0,1},v_{1D},v_{0,0}\right] \qquad (22)$$

To obtain the steady state probabilities, solve the equation

$$\mathbf{v=vP} \qquad (23)$$

This yield

$$v_2=\frac{\lambda(1-c)}{\lambda+\lambda_s}v_1 \qquad (24)$$

$$v_3=\frac{\lambda_s+\mu}{\mu}\left[1-\frac{\mu(\lambda_s+\mu)}{\lambda\mu+\lambda_s\mu+2\mu^2}\left(1+\frac{1}{\lambda+\lambda_s}\left(\lambda_s c_s+\frac{1}{\lambda T}\left(1-e^{-\lambda T}\right)\lambda_s(1-c_s)-\lambda c-\frac{\lambda(1-c)\beta}{\beta+\lambda_s}\right)\right)\right]v_1 \qquad (25)$$

$$v_4=\left[\frac{(\lambda_s+\mu)(\lambda+\mu)}{\lambda\mu+\lambda_s\mu+2\mu^2}\left(1+\frac{1}{\lambda+\lambda_s}\left(\lambda_s c_s+\frac{1}{\lambda T}\left(1-e^{-\lambda T}\right)\lambda_s(1-c_s)-\lambda c-\frac{\lambda(1-c)\beta}{\beta+\lambda_s}\right)\right)\right]v_1 \qquad (26)$$

$$v_5=\frac{\lambda_s(1-c_s)}{\lambda+\lambda_s}v_1 \qquad (27)$$

$$v_6=2\left[\frac{\lambda_s+\mu}{\mu}-\frac{(\lambda_s+\mu)^2}{\lambda\mu+\lambda_s\mu+2\mu^2}\left(1+\frac{1}{\lambda+\lambda_s}\left(\lambda_s c_s+\frac{1}{\lambda T}\left(1-e^{-\lambda T}\right)\lambda_s(1-c_s)-\lambda c-\frac{\lambda(1-c)\beta}{\beta+\lambda_s}\right)\right)-\frac{\lambda c}{\lambda+\lambda_s}-\frac{\lambda(1-c)\beta}{(\lambda+\lambda_s)(\beta+\lambda_s)}\right]v_1 \qquad (28)$$

The mean sojourn time h_i in state i is

$$h_i = \int_0^{\infty} \left(1 - H_i(t)\right) dt \tag{29}$$

$$h_1 = \frac{1}{\lambda + \lambda_s} \tag{30}$$

$$h_2 = \frac{1}{\beta + \lambda_s} \tag{31}$$

$$h_3 = \frac{1}{\lambda_s + \mu} \tag{32}$$

$$h_4 = \frac{1}{\lambda + \mu} \tag{33}$$

$$h_5 = \frac{1}{\lambda} - \frac{1}{T\lambda^2}\left(1 - e^{-\lambda T}\right) \tag{34}$$

$$h_6 = \frac{1}{2\mu} \tag{35}$$

The state probabilities of the semi-Markov chain are

$$\pi_i = \frac{v_i h_i}{\sum_j v_j h_j} \tag{36}$$

, where $i, j \in \{(1,1), 1C, (0,1), (1,0), 1D, (0,0)\}$

$$\pi_1 = \frac{h_1}{\begin{array}{l} h_1 + \frac{\lambda(1-c)}{\lambda+\lambda_s} h_2 + \\ \left[\begin{array}{l} \frac{\lambda_s+\mu}{\mu} - \frac{(\lambda_s+\mu)^2}{\lambda\mu+\lambda_s\mu+2\mu^2} \\ \left(1 - \frac{1}{\lambda+\lambda_s}\left(\begin{array}{l}\lambda_s c_s + \frac{1}{\lambda T}\left(1-e^{-\lambda T}\right)\lambda_s(1-c_s) \\ -\lambda c - \frac{\lambda(1-c)\beta}{\beta+\lambda_s}\end{array}\right)\right)\end{array}\right] h_3 + \\ \left[\frac{(\lambda_s+\mu)(\lambda+\mu)}{\lambda\mu+\lambda_s\mu+2\mu^2}\left(\begin{array}{l} 1 + \frac{1}{\lambda+\lambda_s} \\ \left(\begin{array}{l}\lambda_s c_s + \\ \frac{1}{\lambda T}\left(1-e^{-\lambda T}\right)\lambda_s(1-c_s) \\ -\lambda c - \frac{\lambda(1-c)\beta}{\beta+\lambda_s}\end{array}\right)\end{array}\right)\right] h_4 + \\ \frac{\lambda_s(1-c_s)}{\lambda+\lambda_s} h_5 + \\ \left[\begin{array}{l} \frac{2(\lambda_s+\mu)}{\mu} - \\ \frac{2(\lambda_s+\mu)^2}{\lambda\mu+\lambda_s\mu+2\mu^2}\left(\begin{array}{l} 1 + \frac{1}{\lambda+\lambda_s} \\ \left(\begin{array}{l}\lambda_s c_s + \frac{1}{\lambda T}\left(1-e^{-\lambda T}\right)\lambda_s(1-c_s) \\ -\lambda c - \frac{\lambda(1-c)\beta}{\beta+\lambda_s}\end{array}\right)\end{array}\right) \\ - \frac{2\lambda c}{\lambda+\lambda_s} - \frac{2\lambda(1-c)\beta}{(\lambda+\lambda_s)(\beta+\lambda_s)}\end{array}\right] h_6 \end{array}} \tag{37}$$

$$\pi_2 = \frac{\lambda(1-c)}{\lambda+\lambda_s} h_2 \times \frac{P_1}{h_1} \tag{38}$$

$$\pi_3 = \left[\begin{array}{l} \frac{\lambda_s+\mu}{\mu} - \frac{(\lambda_s+\mu)^2}{\lambda\mu+\lambda_s\mu+2\mu^2} \\ \left(\begin{array}{l} 1 - \frac{1}{\lambda+\lambda_s} \\ \left(\begin{array}{l}\lambda_s c_s + \frac{1}{\lambda T}\left(1-e^{-\lambda T}\right)\lambda_s(1-c_s) \\ -\lambda c - \frac{\lambda(1-c)\beta}{\beta+\lambda_s}\end{array}\right)\end{array}\right)\end{array}\right] h_3 \times \frac{\pi_1}{h_1} \tag{39}$$

$$\pi_4 = \left[\frac{(\lambda_s+\mu)(\lambda+\mu)}{\lambda\mu+\lambda_s\mu+2\mu^2} \left(\begin{array}{l} 1+\frac{1}{\lambda+\lambda_s} \\ \left(\begin{array}{l} \lambda_s c_s + \frac{1}{\lambda T} \\ (1-e^{-\lambda T}) \\ \lambda_s(1-c_s) - \\ \lambda c - \frac{\lambda(1-c)\beta}{\beta+\lambda_s} \end{array} \right) \end{array} \right) \right] \quad (40)$$

$$\pi_5 = \frac{\lambda_s(1-c)}{\lambda+\lambda_s} h_5 \times \frac{P_1}{h_1} \quad (41)$$

$$\pi_6 = \left[\begin{array}{l} \frac{\lambda_s+\mu}{\mu} - \frac{(\lambda_s+\mu)^2}{\lambda\mu+\lambda_s\mu+2\mu^2} \\ \left(\begin{array}{l} 1-\frac{1}{\lambda+\lambda_s} \\ \left(\begin{array}{l} \lambda_s c_s + \frac{1}{\lambda T}(1-e^{-\lambda T})\lambda_s(1-c_s) \\ -\lambda c - \frac{\lambda(1-c)\beta}{\beta+\lambda_s} \end{array} \right) \end{array} \right) \\ -\frac{2\lambda c}{\lambda+\lambda_s} - \frac{2\lambda(1-c)\beta}{(\lambda+\lambda_s)(\beta+\lambda_s)} \end{array} \right] h_6 \times \frac{P_1}{h_1} \quad (42)$$

5. Experimental Results

The exact model parameter values for the model are not known, however, a good estimate value for a range of model parameter is assumed. Fig: 3 plots the difference between downtime (minutes per year) estimates obtained using the SMP model and that obtained by approximating the U (0, T) distribution by an exponential distribution with mean T/2. We take the values c=0.9, c_s=0.9, μ=1per hour, β=12 per hour, and $\lambda_s = \lambda/4$. We see that the higher the μ/λ ratio, the lower the downtime computed by the two models.

Availability models capture failure and repair behavior of systems and their components. States of the underlying Markov chain will be classified as up states or down states. The system is not available in the state 2 and state 6. The system availability in the steady-state is defined as follows:

$$\text{Availability} = 1 - \text{Unavailability} = 1-(\pi_2+\pi_6) \quad (43)$$

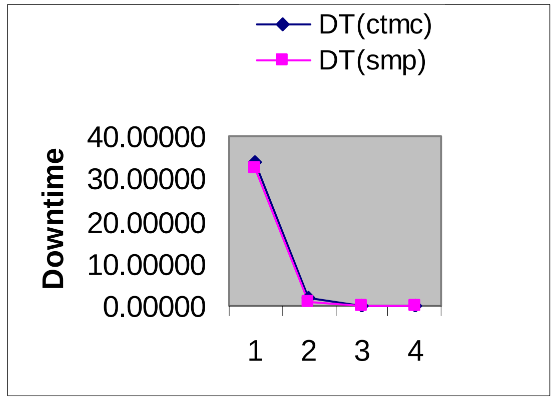

Figure3:Difference in downtime of the SMP model and the approximate CTMC model

5.1 Validation of Closed-form Results

To verify the validity of our formula derivations, we compare the results obtained from the closed-form solution and the results obtained from the numerical solution by SHARPE. We found that our results are same.

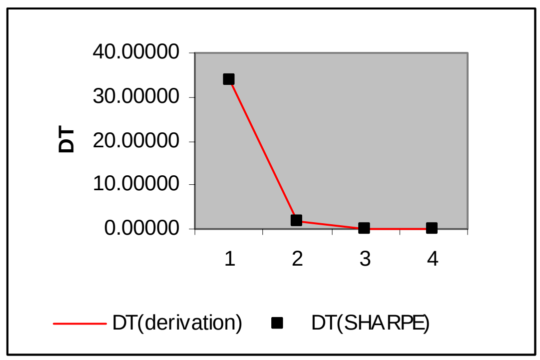

Figure4: Downtime of the CTMC model

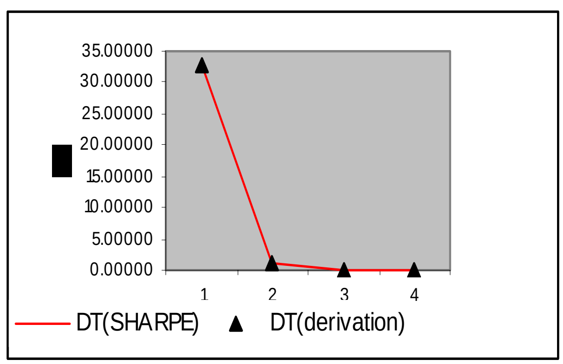

Figure5: Downtime of the SMP model

6. Conclusion

Organizations today face a tough challenge in choosing an appropriate high availability solution that meets their business requirements and IT budgets. To implement this requirement, organizations must give high availability and disaster recovery. High availability systems require fewer failures and faster repair. In this paper we presented high availability cluster and failover availability for disaster events. . We present a Markov model and express availability and downtime in terms of the parameters in the model. We evaluate the feasibility of our clustering model using SHARPE tools.

References

[1] D. Clitherow, M. Brookbanks, N. Clayton, and G. Spear, ‘‘Combining High Availability and Disaster Recovery Solutions for Critical IT Environments,’’ IBM Systems Journal 47, No. 4, 563–575 (2008)

[2] D.Chen, K.S.Trivedi,”Optimization for condition-based maintenance with semi-Markov decision process” Available online at www.sciencedirect.com

[3] R. Gamache, R Short, and Mike Massa, "Windows NT Clustering Service," IEEE Computer, October 1998, pp.55-61.

[4] C.Hirel, A. Robin, Sahner, X.Zang, K.S.Trivedi: “Reliability and performing modeling using SHARPE 2000”. Computer Performance Evaluation/TOOLS 2000. In Lecture Notes in Computer Science; Vol.1786, Springer-Verlag, 2000, pp.345-349.

[5] S. W. Hunter and W. E. Smith, “Availability Modeling and Analysis of a Two Node Cluster,” Proceedings of the 5th International Conference on Information Systems, Analysis and Synthesis, Orlando, FL, October 1999.

[6] Th. Lumpp, J. Schneider, J. Holtz, M. Mueller, N. Lenz, A. Biazetti, and D. Petersen, ‘‘From High Availability and Disaster Recovery to Business Continuity Solutions,’’ IBM Systems Journal 47, No. 4, 605–619

[7] M.Malhotra, A.Reibman:”Selecting and Implementing Phase Approximations for Semi-Markov Models”, Volume 9, Issue 4, 1993, Pages 473-506.

[8] G.F. Pfister, In Search of Clusters: The Coming Battle in Lowly Parallel Computing, Prentice Hall, Englewood Cliffs, NJ, 1998.

[9] H.Song, C.Leangsuksun, R.Nassar, “Availability Modeling and Evaluation on High Performance Cluster Computing Systems,” Journal of Research and Practice in Information Technology, Vol.38, No.4, November 2006.

[10] K. S. Trivedi, R. Vasireddy, D. Trindade, S. Nathan, and R. Castro. Modeling high availability systems. In Proc. Pacific Rim Dependability Conference, 2006.

[11] K.S.Trivedi:” Probability and Statistics with Reliability, Queuing, and Computer Science Applications”, John Wiley and Sons, 2002.

[12] M.Wiboonrat, “Transformation of System Failure Life Cycle,” International Journal of Management Science and Engineering Management, Vol.4 (2008) No.2, pp.143-152.

A Brief History of Context

Kaiyu Wan

Computer Science Department, East China Normal University
Shanghai, 200, China

Abstract
Context is a rich concept and is an elusive concept to define. The concept of context has been studied by philosophers, linguists, psychologists, and recently by computer scientists. Within each research community the term context was interpreted in a certain way that is well-suited for their goals, however no attempt was made to define context. In many areas of research in computer science, notably on web-based services, human-computer interaction (HCI), ubiquitous computing applications, and context-aware systems there is a need to provide a formal operational definition of context. In this brief survey an account of the early work on context, as well as the recent work on many working definitions of context, context modeling, and a formalization of context are given. An attempt is made to unify the different context models within the formalization. A brief commentary on the usefulness of the formalization in the development of context-aware and dependable systems is included.

Keywords: *Context, Context Theory, Context-Awareness.*

1. Introduction

According to the Oxford English Dictionary (OED), context denotes "the circumstances that form the setting for an event". To emphasize a common social usage of the word context OED includes the quotation [12] "I wish honorable gentlemen would have the fairness of what I did say, and not pick out detached words". Although the word context has been used for a long time in many scientific descriptions, literary essays, and in philosophical discourses, its meaning was always left to the reader's understanding. In one of the earlier papers, Clark and Carlson [11] state that Context has become a favorite word in the vocabulary of cognitive psychologists and that it has appeared in the titles of a vast number of articles. They then complain that the denotation of the word has become murkier as its uses have been extended in many directions and deliver the now widespread opinion that context has become some sort of "conceptual garbage can". That context has changed now. The importance of context in information retrieval, knowledge representation, reasoning in AI, and analysis of computer programs have been recognized and there is a serious effort to make a precise technical working definition of the notion of context. More recently, the importance of context was picked up by researchers in many areas of computer science, most importantly those working in Human-Computer Interaction (HCI), semantic web, and trustworthy systems. This intense interest has produced many operational definitions of context, but almost all of them are either informal or use ad hoc notation. We review in this paper the different types of notations and interpretations used for context. The review is classified into *Context in Logic*, *Context in Languages*, and *Context in Systems*. This classification and review are not exhaustive. It is used mainly to trace the historical progression of the systematic study of context in different, but related, areas.

1.1 Structure and Interpretation

The word "context" is derived from the Latin words *con* (meaning "together") and *texere* (meaning "to weave"). The raw meaning of it is therefore "weaving together". A circumstance is a weaving together of many types of entities. Thus, in describing a context we must define a finite set of entities, a finite set of properties for each entity, and the inter-weaving of the properties. As an example, the setting for a "seminar event" is the weaving together of the entities *speaker*, *topic*, *audience*, *time*, *location* and their properties such as *name* and *affiliation* for the speaker, *title* and *abstract* for the topic, *size* and *status* for audience, *clocktime* for time, and *building-address* and *room-number* for locality. We need to associate with each property a value from its domain and bind each entity with the instantiated properties in order to describe the context of seminar. The choice of entities, the choice of properties, and the notation used for binding them are crucial for system development. This choice for context definition has the effect of narrowing down the possible interpretations of declared policies and constraints for system development. Context description also eliminates ambiguities. It should be possible to define contexts in programming languages independent of how it should be used. For example, if a context is defined by *locality* and *time*, then many events may happen in a specific context, and each event may produce a different experience in one context. Therefore, the structural definition of a context is only part of its specification, its other part being the semantics of the *world* associated with the context. The world may be defined by a set of states in programming or by a set of logical formulas in a formal specification. Motivated by a need to specify context as a first class citizen in languages and systems, a formal representation of it was developed in [45].

2. Context in Logic

In this section we review the study of context in logic as a formal object and reasoning. We review intensional logic and some variations of propositional and predicate logic in which context has been embedded as first class citizens.

2.1 Intensional Logic

Intensional Logic [14, 42], a family of mathematical formal systems that permits expressions whose value depends on *hidden context*, came into being from research in natural language understanding. According to Carnap [9], the real meaning of a natural language expression whose truth-value depends on the context in which it is uttered is its *intension*. The *extension* of that expression is its actual truth-value in the different possible contexts of utterance, where this expression can be evaluated. Basically, intensional logics add *dimensions* to logical expressions, and non-intensional logics can be viewed as *constant* in all possible dimensions, i.e. their valuation does not vary according to their context of utterance. *Intensional operators* are defined to *navigate* in the context space. In order to navigate, some dimension *tags* (or indexes) are required to provide placeholders along dimensions. These dimension tags, along with the dimension names they belong to, are used to define the context for evaluating intensional expressions.

Example 1 E: Beijing is now the capital of China.

This expression is intensional because the truth value of this expression depends on the context in which it is evaluated. The intensional natural language operator in this expression is now, which refers to the time dimension. Today it is certainly true, but there existed time points in the past when China had a different capital. For example, before 1949, the capital of China was NanJing. Those different values (i.e. True or False) along different time points are extensions of this expression. In other words, the evaluation of the above expression is time-dependent. A natural extension is to consider expressions that depend on more than one dimension, such as time, space, audience, and so on.

Example 2 The meaning of the expression:

E: the overseas indexes during this period close 10% below their highs can be interpreted when the possible worlds spanned by the dimensions overseas and period are defined. The Table below gives a possible extension of the expression E when the periods are months in a given year, and overseas stock markets are Amsterdam, Brussels, Frankfurt, and London. By varying the year we get 3-dimensional extension.

Table 1: Example 1

	Ja	*Fe*	*Mr*	*Ap*	*Ma*	*Jn*	*Jl*	*Au*	*Se*	*Oc*	*No*	*De*
Amsterda	*F*	*F*	*F*	*F*	*T*	*T*	*T*	*T*	*T*	*F*	*F*	*F*
Brussels	*F*	*F*	*F*	*T*	*T*	*T*	*T*	*T*	*T*	*F*	*F*	*F*
Frankfurt	*F*	*F*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*F*	*F*
London	*F*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*	*T*

2.2 Formalizing Context in AI

Contexts in AI were introduced by Weyhrauch (1980) [51] and subsequently developed by McCarthy and Buva c (1998) [28] and Giunchiglia (1993) [17]. Surveys of the formalizations and the usage of contexts can be found in Sharma (1995) [38], Akman and Surav (1996) [2], Bouquet et al. (1999) [5], Bonzon et al. (2000) [6] and Akman et al. (2001) [3]. Context serves an important purpose in AI and Intelligent Information Processing (IIP). The classic example of the earliest IIP that failed to meet safety criteria is MYCIN [39]. It was observed by McCarthy [2]. MYCIN system advises physicians on treating bacterial infections of the blood and meningitis. When MYCIN was first introduced context was not part of system's query processing phase. When it was given the query "what is the treatment for *Chlorae Vibrio*" it recommended "two weeks of *tetracycline*" treatment. What it failed to inform the physician was that a massive *dehydration* during the course of the treatment would occur. While the administration of *tetracycline* would cure the bacteria, the patient would die long before that due to diarrhea. Here is an instance where the context of correct usage was not given to the physician, which ultimately made the system *unsafe* and hence not trustworthy. A contextual MYCIN will explicitly state the context for correct administration of medications. In AI context is formalized using propositional or predicate logic.

Contexts are abstract objects (representation free) and are first-class citizens. Consequently, contexts are freely used in logical formulas, without explicitly defining contexts. In some sense, in the logical approach a context itself is defined by a set of formulas that are true by the truth assignments in that context. In [28] McCarthy, who introduced a logical framework in AI for studying context, gave three reasons justifying his approach.

- *Axiomatization:* The use of contexts simplifies axiomatizations. Axioms from one context can be *lifted* to more general contexts.
- *Vocabulary and Interpretation:* Contexts allow the use of specific vocabulary and information. Terms that are used in one context have particular meaning, which they will not have in general.
- *Building AI Systems:* A hierarchy of AI systems can be built be transcending from one context to another.

According to Giunchiglia (1993) [17], the notion of context formalizes the idea of localization of knowledge and reasoning. Intuitively speaking, a context is a set of facts (expressed in a suitable language, usually different for each different set of facts) used locally to prove a given goal, plus the inference routines used to reason about them (which can be different for different sets of facts). A context encodes a perspective about the world. It is a partial perspective as the complete description of the world is given by the set of all the contexts. It is an approximate perspective, in the sense described in McCarthy (1979) [27], as we never describe the world in full detail. Finally, different contexts, in general, are not independent of one another as the different perspectives are about the same world, and, as a consequence, the facts in a context are related to the facts in other contexts.

The work in Giunchiglia and Serafini (1994) provides a logic, called Multi Language Systems (ML Systems), formalizing the principles of reasoning with contexts informally described in Giunchiglia (1993). In ML systems, contexts are formalized using multiple distinct languages, each language being associated with its own theory (a set of formulas closed under a set of inference rules). Relations among different contexts are formalized using bridge rules, namely inference rules with premises and consequences in distinct languages. Recently, Ghidini and Giunchiglia (2001) proposed Local Models Semantics (LMS) as a model-theoretic framework for contextual reasoning, and use ML systems to axiomatize many important classes of LMS. From a conceptual point of view, Ghidini and Giunchiglia argued that contextual reasoning can be analyzed as the result of the interaction of two very general principles: the principle of locality (reasoning always happens in a context); and the principle of compatibility (there can be relationships between reasoning processes in different contexts). In other words, contextual reasoning is the result of the (constrained) interaction between distinct local structures. A good survey of context formalization in AI and a comparison between different formalizations can be found in [2]. According to this exposition, context is either treated within some logical framework or within situation theory. Both approaches deal with abstract contexts and focus only on contextual reasoning.

3. Context in Languages

We review the role of context in intensional programming languages (IPL) and in ¸ λ calculus.

3.1 Formalizing Context in AI

The intensional programming paradigm has its foundations on intensional logic. It retains two aspects from intensional logic: first, at the syntactic level, are contextswitching operators, called *intensional operators*; second, at the semantic level, is the use of *possible world semantics*. By making difference between intension and extension, IPL provides two different levels for programming. On the higher level, it allows to represent/express problems in a declarative manner; on the lower level, it solves problems without loss of accuracy.

IPL deals with *streams* of entities which could be numbers, or strings of characters, or any computable structure. These streams are first class objects in intensional languages and functions can be applied to these streams. Because of the infinite nature of IPL, it is especially appropriate for describing the behavior of systems that change with time or physical phenomena that depend on more than one parameter (such as time, space, temperature, etc). It is also an appropriate language for use in business applications that generate data streams, or textual streams, or media streams. Examples include stock market transactions and credit card transactions which are mostly data streams of records where each record contains information on a transaction, call center transactions that generate textual streams of conversations, and multi-media streams that are generated by cable companies to distribute movies on demand. The streams are processed by accessing certain semantic units and interpreting it in different contexts.

There is no notion of type in an IPL. The operators on the stream contents are assumed to be given when one writes the stream functions. The natural logical view of a stream is an infinite sequence, and in writing programs one does not worry about the physical representations of stream contents. This abstraction enables one to understand an IPL program from the statements in it, without any reference to its implementation. The computational model for IPLs is known as *eduction*. That is, an implementation starts computing the first element that satisfies a given context, then the second, and so on. A context for expression evaluation, as informally understood in Example 2, is described by a set of dimensions (attributes) and a *finality* (goal). The finality is domain-dependent and is chosen so that a finite set of dimensions would suffice to realize that goal. For example in processing call center streams *understanding a conversation* may be the finality, and the attributes may be a set of *key words* chosen in advance to meet the goal. As another example, in processing streams of user interactions with web, the finality may be *understanding user patterns* and the attributes may be *ActivityLocation*, *ActivityDuration*, and *VolumeofDataTransfer*. Both finality and the attributes defining a context are implicitly used in evaluating the extensions from a stream.

3.2 Lambda Calculus

In programming languages, context is a meta concept: *static* context introduces constants, definitions, and constraints, and *dynamic* context processes the executable information for evaluating expressions. In [35] context is introduced in the lambda calculus and an argument is made for introducing context as first class objects in programming languages. Their motivation for introducing context in the theory of lambda calculus is to develop a programming language with first-class contexts that has advanced programming features for manipulating open terms. We are motivated along similar lines for introducing context in Lucid. However there are significant differences in the semantics of context between the two approaches.

A context in the lambda calculus is defined as a term with a"hole" in it. The hole in a context can be filled with a term which may involve free variables. To avoid inconsistent hole filling within the scope of lambda binding the holes are labeled, hole abstraction, and context application are separated. Informally, a context C with a hole in it, written C[•], will become the term C[M] when the hole is filled with the term M. The formal way of writing this in λ calculus is M' $\odot$M , where the term M' abstracts the hole in the context. The term M that abstracts the hole labeled X itself is written as δX.M'. For example, the context C[•] = (λ x.[•] +y)3 is represented by the term M'= (δX(λ x.X + y)3). The term obtained by filling the hole in M' with x+ z is written (δX.(λ x.X+y)3) $\odot$ (x+z).

In our theory context plays two roles: one role is as a reference to an item in a multi-dimensional stream, and the other role is as a descriptor of situations at which expressions are evaluated. A stream of contexts may be constructed and a context expression may be evaluated at a context. In Lucx expressions and contexts exist independently. A context may be defined without any regard to any specific expression and hence it may be used to evaluate different Lucx expressions. Similarly, an expression can be evaluated at different contexts. It is possible to define context dependent expressions in Lucx. Such expressions may be evaluated at a context distinct from any other context used in its definition. We can define nested contexts, and dependent contexts. These features offer a variety of flexible ways to programming different applications.

3.3 Lucid and Lucx

Lucid was originally invented as a *Program Verification Language* by Ashcroft and Wadge [1]. And later it evolved into a dataflow language [52]. The basic intensional operators are *first*, *next*, and *fby*. The four operators derived from the basic ones are *wvr*, *asa*, *upon*, and *prev*, where *wvr* stands for *whenever*, *asa* stands for *as soon as*, *upon* stands for *advances upon*, and *prev* stands for *previous*. Lucid is a *stream* (i.e. infinite entity) manipulation language. All the above operators are applied to streams to produce new streams. The definitions of these operators [30] are shown as follows

Definition 1 If X = (x0,x1,…, xi,…) and Y = (y0,y1,…,yi,…), then

(1) <u>first</u> X = (x0,x0,…,x0,…)

(2) <u>next</u> X = (x1,x2,…,xi+1,…)

(3) X <u>fby</u> Y = (x0,y0,y1,…,yi-1,…)

(4) X <u>wvr</u> Y = if <u>first</u> Y then X
<u>Fby</u> (nextX <u>wvr</u> <u>next</u> Y)
Else (next X <u>wvr</u> <u>next</u> Y)

(5) X <u>asa</u> Y = <u>first</u> (X <u>wvr</u> Y)

(6) X <u>upon</u> Y = X <u>fby</u> (if <u>first</u> Y
then (next X upon next Y)
else (X upon next Y))

(7) <u>prev</u> X = X@(#1) 2

Example 3 illustrates the definitions on a stream *A* whose elements are integers, and a stream *B* whose elements are boolean. In a boolean stream the symbols 1 and 0 indicate true and false respectively. The symbol *nil* indicates an undefined value.

Example 3 *:*

A	=	1	2	3	4	5	
B	=	0	0	1	0	1	
first A	=	1	1	1	1	1	
next A	=	2	3	4	5		
prev A	=	*nil*	1	2	3	4	5
AfbyB	=	1	0	0	1	0	1
A wvr B	=	3	5				
A asa B	=	3	3	3			
A upon B		1	1	1	3	3	5

With the operators defined above, Lucid only allows sequential access into streams. That is, the $(i + 1)th$ element in a stream is only computed once the *ith* element has been computed. To enable subcomputations to take place in arbitrary dimensions and all indexical operators to be parameterized by one or several dimensions, two basic intensional operators are added. One is *intensional navigation* (@.*d*), which allows the values of a stream to vary along the dimension *d*. Another is *intensional query* (#.*d*), which refers to the current position (i.e. tag value) along the dimension *d*. This way, it is possible to access streams randomly.

Example 4 illustrates the definitions of these two operators on two streams *A* and *B* along the *time* dimension.

Example 4

A = 1 2 4 8 16 32 64 128 …

B = 1 2 3 0 6 7 4 5 …

A @.time B = 2 4 8 1 64 128 16 32 …

#.time = 0 1 2 3 4 5 6 7 …

The major distinction between contexts in AI and in IPL is that in the former case they are *rich objects* that are not *completely expressible* and in the later case they are *implicitly* expressible. Hence it is possible to write an expression in Lucid whose evaluation is context-dependent. However, a context in the current version of Lucid can not be explicitly manipulated. This restricts the ability of Lucid to be an effective programming language for programming diverse applications. So we have extended Lucid by adding the capability to explicitly manipulate contexts. This is achieved by introducing *context* as a first class object in the language. That is, contexts can be declared, assigned values, used in expressions, and passed as function parameters. The language thus extended, is called as *Lucx* [45] (Lucid extended with *contexts*)(the x is used as the x in TeX). Thus, the rationale for introducing context in Lucid is quite analogous to the introduction of context to enrich knowledge base in AI. However, our notion of context differs significantly from McCarthy's. In our study context is both *finite* and *concrete*. It is finite in the sense that only a finite number of dimensions are allowed in defining a context. However it does not impose any limitation on handling infinite streams, because with every dimension an infinite *tag set* is introduced in the language. A full account of context-based evaluation of expressions in Lucx is given in [45].

4. Context in Systems

Context-aware adaptation is regarded as the most important feature for pervasive and ubiquitous services [50, 20, 25]. Web services [26] and mobile computing applications [53, 24] immensely benefit with a formal context model. It is in this context that we review the role of context. Context modeling and context-dependent interpretive actions are important in HCI [13, 15, 49]. However in all these works context is not formalized. In this section, after we review context formalism, we explain how our formal definition provides a rigorous platform for developing context-aware systems.

4.1 Formalizing Context

We formalize context as *a typed relation*, a set of ordered pairs of (*d*, *x*) where *d* is a dimension, T_d is the type of *d* and $x : T_d$.

Definition 2 Let DIM denote the set of all possible dimensions, and $T = \{T_d \mid d \in DIM\}$ be the set of types associated with the dimensions. A context c is a finite relation $\{f(d, x) \mid d \in DIM \wedge x : T_d\}$. The degree of the context c is |dom c|. The empty relation corresponds to Null context. The degree of Null context is 0.

A context having only one (dimension, tag) pair is called a micro context. Let G denote the set of contexts over {DIM, T}. The set of micro contexts is $M = \{c \mid c \in G; \mid c \mid = 1\}$. The set of simple contexts is $S = \{c \mid c \in G$, c is a partial function$\}$. Clearly, a simple context c of degree 1 is a micro context. A context which is not simple is referred to a non-simple context. The basic functions dim and tag are to extract the set of dimensions and the values associated with the dimensions in a context. That is, if $c = f<d_1,x_1>,\ldots,<d_k, x_k>\}$, then we may write $c = \{m_i \mid m_i = <d_i,x_i>\}$, $dim(c) = \{d_1, d_2,\ldots d_k,\}$, and $tag(c) = \{x_1, x_2,\ldots x_k\}$. For the tuple (d, x) in a micro context c we use the functions dim_m and tag_m to extract the tuple components: $dim_m(c) = d$ and $tag_m(c) = x$.

4.2 Context Operators

In this section, context operators are discussed. A context being a relation we borrow the notation and meaning of those relational operators that are available in mathematics. Rest of them we define, using set theory notation. Using these context operators contexts can be managed dynamically and flexibly. The syntax of context expressions are also formally defined. In order to evaluate context expression correctly, precedence rules for context operators are provided as well.

Context operators are : *override* $\oplus$, *difference* $\ominus$, *choice* | , *conjunction* $\bigcap$, *disjunction* $\bigcup$, *undirected range* $\rightleftharpoons$, *directed range* $\rightharpoonup$, *projection* $\downarrow$, *hiding* $\uparrow$, *substitution* / , *comparison* =, $\subseteq$, $\supseteq$. The *difference* $\ominus$, *conjunction* $\bigcap$, *disjunction* $\bigcup$, and *comparison* =, $\subseteq$, $\supseteq$ operators are set operators. The rest of the operators are explained and formally defined below.

Definition 3 Override $\oplus$ This operator takes two contexts $c1 \in G$, and $c2 \in S$ and returns a context $c \in G$, which is the result of the conflict -free union of c1 and c2, as defined below:

$$\_\oplus\_ : G \times S \rightarrow G,$$

$c= c1 \oplus c2 = \{m \mid (m \in c1 \wedge dim_m(m) \notin dim(c2)) \vee m \in c2\}$

Definition 4 Choice | This operator accepts a finite number of c1ck of contexts and non-deterministically return one of the cis. The definition c= c1|c2|…,|ck implies that c is one of the ci, where $1 \leq i \leq k$:

$$\_|\_ : G \times G \times \ldots \times G \rightarrow G,$$

Definition 5 Projection. This operator takes a context $c \in G$ and a set of dimensions $D \subset DIM$ as arguments and filters only those micro contexts in c that have their dimensions in set D.

$$\_\downarrow\_ : G \times D \rightarrow G$$

$$c \downarrow D = \{m \mid m \in c \wedge dim_m(m) \in D\}.$$

Example 5 :

Let c1 = { (d,1) , (e,4) , (f,3)}, D={d,e}

then c1 $\downarrow$ D= { (d,1) , (e,4)}

Definition 6 Hiding. This operator enables a set of dimensions D to be applied on a context $c \in G$ to remove all the micro context s in c whose dimensions are in D:

$$\_\uparrow\_ : G \times D \rightarrow G ,$$

$$c \uparrow D=\{m \mid m \in c \wedge dim_m(m) \notin D\}$$

Example 6 :

Let c1 = { (d,1) , (e,4) , (f,3)}, D={d,e}

then c1 $\uparrow$ D = {(f,3)}

Definition 7 Substitution. This operator takes a general context and a simple context as arguments and produces a context which is the result of replacing a sub-context of the general context with a sub-context of the simple context if their domains are equal.

$$\_/\_ : G \times S \rightarrow G ,$$

$$c/s = (c \uparrow dim\ s) \bigcup (s \downarrow dim\ c)$$

Example 7 :

Let c1 ={(d, 1), (e, 4),(d, 3)},

c2 ={(d, 4),(f, 3)}, then c1/c2 = {(e,4),(d,4)}

Definition 8 Undirected range. This operator takes two contexts c1, $c2 \in G$ as arguments and returns a set of simple contexts. The tag set U is assumed to be totally ordered. We give a constructive definition here.

$$\_\rightleftharpoons\_ : G \times G \rightarrow \mathcal{P} S$$

Steps for constructing the final result are shown as follows:

1. Let S' be the set of simple contexts, which is the result of ($c1 \rightleftharpoons c2$).
2. For each pair of $m1 \in c1$, $m2 \in c2$, and $dim_m(m1) = dim_m(m2)$, do the following:
   (a) Define $a = \min\{tag_m (m1), tag_m (m2)\}$ and $b = \max\{tag_m(m1), tag_m(m2)\}$
   (b) Define the subrange t_{ba} = a..b.
   (c) Construct the set Y1:
   $Y1 =\{ (d_{ba} ,x) \mid d_{ba} = dim_m(m1) = dim_m(m2), x \in t_{ba}\}$
3. Y={Y1,Y2,…Yp}, Where Yi(i = 1,…,p), are the sets of micro context s constructed in Step 2. Define for $Yi \in Y$, $first(Yi) =\{dim_m(m) \mid m \in Yi\}$, and, $second(Yi) =\{tag_m(m) \mid m \in Yi\}$. If there exists Yi, $Yj \in Y$ such that first(Yi) = first(Yj), for $i \neq j$, we replace the sets Yi and Yj by their union $Yi \bigcup Yj$, and repeat this process until the first(Yi)s for Yi Y are distinct
4. For $Yi \in Y$, construct the set Z of contexts: $Z=\{\{(first(Y1),x1), (first(Y2), x2),\ldots, (first(Yp), xp)\} \mid (x1, x2, \ldots, xp) \in \prod^{p}_{i=1} second(Yi))\}$.
5. Define: $Xc1 = c1 \uparrow \bigcup_{Yi \in Y} first(Yi)$.
6. Define: $Xc2 = c2 \uparrow \bigcup_{Yi \in Y} first(Yi)$..
7. Construct S': $S'= \{\{z\} \bigcup X_{c1} \bigcup X_{c2} \mid z \in Z\}$.

Basically, the result consists of three parts:

1. For each pair $m1 \in c1$, $m2 \in c2$ which shares the same dimension, constructs a set Yi, this is done in step 2 and step 3. The result of union the set Yi, done in step 4, consists of the first part : Z.
2. All the other micro context s of c1 which have different dimensions consists of the second part : Xc1.
3. Similarly, all the other micro context s of c2 which have different dimensions consists of the third part : Xc2.

Example 8 :

Let $c1 =\{(e, 3), (d, 1)\}$, $c2 =\{(e, 1),(d, 3)\}$,
$c3 =\{(e, 3)\}$, $c4 =\{(f, 4)\}$,
$c5 =\{(e, 1), (f, 4)\}$
then $c1 \rightleftharpoons c2=\{\{(e,1),(d,1)\}, \{(e,1),(d,2)\}, \{ (e,1),(d,3)\}, \{(e,2),(d,1)\},\{(e,2),(d,2)\}, \{ (e,2),(d,3)\}, \{(e,3),(d,1)\}, \{(e,3),(d,2)\}, \{ (e,3),(d,3)\}$
$c3 \rightleftharpoons c4 =\{(e,3),(f,4)\}$
$c3 \rightleftharpoons c5 =\{\{(e,1),(f,4)\},\{(e,2),(f,4)\}, \{(e,3),(f,4)\}\}$

Definition 9 *Directed Range. This operator takes two contexts* $c1 \in G$ *and* $c2 \in S$ *and returns a set of contexts:*

$$\_\rightharpoonup\_ : G \times S \rightarrow \mathcal{P} G$$

We change only Step 2 of the method described for the undirected range(Page 7) to obtain the result:

(a) Define $a = tag_m (m1)$, $b = tag_m (m2)$, *if* $tag_m (m1) < tag_m(m2)$, *else ignore the pair m1, m2.*

(b) Define the subrange t_{ba} *= a..b.*

Example 9 :

Let $c1 = \{(d,1)\}$, $c2 = \{(d,3), (f,4)\}$,

then $c1 \rightharpoonup c2 = c1 \rightleftharpoons c2 = \{(d,1),(f,4)\}, \{(d,2),(f,4)\}, \{(d,3),(f,4)\}\}$,
$c2 \rightharpoonup c1 = \{(f,4)\}$

4.2 Context Expression

Informally, a context expression is an expression involving context variables and context operators. Let c be a context variable, and D be a set of dimensions. A formal syntax for context expression C is shown in Figure 1(left column). A context expression that satisfies those syntactic rules is a well-formed context expression.

In order to provide a precise meaning for a context expression, we define the precedence rules for all the operators. Figure 1(right column) shows the operator precedence from the highest (top row) to the lowest (bottom row). Parentheses will be used to override this precedence when needed. Operators having the same precedence will be applied from left to right.

syntax			precedence
$C ::=$	c	$C = C$	$\downarrow, \uparrow, /$
|	$C \supseteq C$	$C \subseteq C$	|
|	$C \mid C$	C / C	$\cap, \cup$
|	$C \oplus C$	$C \ominus C$	$\oplus, \ominus$
|	$C \cap C$	$C \cup C$	$\rightleftharpoons, \rightharpoonup$
|	$C \rightleftharpoons C$	$C \rightharpoonup C$	$=, \subseteq, \supseteq$
|	$C \downarrow D$	$C \uparrow D$	

Fig. 1 Formal Syntax of Context Expressions and Precedence Rules for Context Operators.

Example 11 *Given a well -formed context expression* $c3 \uparrow D \oplus c1 \mid c2$, *where* $c1 = \{(x,3),(y,4),(z,5)\}$, $c2 = \{(y,5)\}$, *and* $c3 = \{(x,5),(y,6),(w,5)$, $D = \{w\}$, *the evaluation steps are shown as follows:*

[Step1]. $c3 \uparrow D = \{(x,5),(y,6)\}$[Definition 6, Page 6]

[Step2]. $c1|c2 = c1$ *or* $c2$ [Page 6] [Step3]. Suppose in Step2, $c1$ is chosen,

$c3 \uparrow D \oplus c1 = \{(x,3),(y,4),(z,5)\}$ [Definition 3, Page 6]

else if $c2$ is chosen,

$c3 \uparrow D \oplus c2 = \{(x,5),(y,5)\}$[Definition 3, Page 6]

4.3 Context Set Operators

In Lucx we avoid higher-order sets of contexts, and allow only sets of simple contexts. Hereafter, by "set of contexts" we refer only to "set of simple contexts". There are two kinds of such operators: *lifting* operators, and *relational* operators.

4.3.1 Lifting Operators

Definition 11 Projection. *For* $s \in \mathcal{P}\, S$, *and* $D \subseteq DIM$. *The projection operator constructs a set of contexts* $s' \in \mathcal{P}\, S$, *where s' is obtained by projecting each context from s on to the dimension set D.*

$$\_\downarrow\_ : \mathcal{P}\, S \times \mathcal{P}\, DIM \rightarrow \mathcal{P}\, S$$

$$s' = s \downarrow D = \{c \downarrow D \mid c \in s\}$$

Definition 12 Hiding. *For* $s \in \mathcal{P}\, S$, *and* $D \subseteq DIM$. *The hiding operator constructs* $s' \in S2$, *where s' is obtained by hiding each context in s on the dimension set D.*

$$\_\uparrow\_ : \mathcal{P}\, S \times \mathcal{P}\, DIM \rightarrow \mathcal{P}\, S$$

$$s' = s \uparrow D = \{c \uparrow D \mid c \in s\}$$

Definition 13 Substitution. *This operator produces a set of contexts s, s*P *S, for a given set of contexts s, s* P *S, a dimension and a tag value belonging to that dimension:*

$$\_/\_ : \mathcal{P}\, S \times (DIM \times U) \rightarrow \mathcal{P}\, S$$

$$s' = s / \langle d', t' \rangle = \{c / \langle d', t' \rangle \mid c \in s\}$$

Definition 14 Choice. *This operator accepts two sets of contexts* $s1$, $s2$ *and non-deterministically returns one of them. The definition* $s = s1|s2$ *implies that s is either* $s1$ *or* $s2$.

$$\_|\_ : \mathcal{P}\, S \times \mathcal{P}\, S \rightarrow \mathcal{P}\, S$$

Definition 15 Override. *For every pair of context sets* $s1$, $s2$, $s1, s2 \in \mathcal{P}\, S$ *this operator returns a set of contexts s,* $s \in \mathcal{P}\, S$, *where every context* $c \in s$ *is computed as* $c1 \oplus c2$, $c1 \in s1$, $c2 \in s2$.

$$\_\oplus\_ : \mathcal{P}\, S \times \mathcal{P}\, S \rightarrow \mathcal{P}\, S$$

$$s = s1 \oplus s2 = \{c1 \oplus c2 \mid c1 \in s1 \wedge c2 \in s2\}$$

Definition 16 Difference. *For every pair of context sets* $s1$, $s2$, $s1, s2 \in \mathcal{P}\, S$ *this operator returns a set of contexts s,* $s \in \mathcal{P}\, S$, *where every context* $c \in s$ *is computed as* $c1 \ominus c2$, $c1 \in s1$, $c2 \in s2$.

$$\_\ominus\_ : \mathcal{P}\, S \times \mathcal{P}\, S \rightarrow \mathcal{P}\, S$$

$$s = s1 \ominus s2 = \{c1 \ominus c2 \mid c1 \in s1, c2 \in s2\}$$

Lifting the *undirected range* $\rightleftharpoons$ and *directed range* $\rightharpoonup$ to sets of contexts will produce higher-order sets. So, we do not define lifting for these two operators. However, since the results of applying these two operators are sets of contexts, the lifting operators can be applied to the results.

4.3.2 Relational Operators

We define the three relational operations $\boxtimes$ (join), $\boxdot$ (intersection), and $\boxplus$ (union) for sets of contexts. In the following definitions, c denotes a context, $si \in \mathcal{P}\, S$ and $\Delta i = \bigcup_{c' \in si} dim(c')$.

Definition 17 Join.

$$\_\boxtimes\_ : \mathcal{P}\, S \times \mathcal{P}\, S \rightarrow \mathcal{P}\, S$$

$s = s1 \boxtimes s2 = \{c1 \cup c2 | c1 \in s1 \wedge c2 \in s2 \wedge c1 \downarrow \Delta 3 = c2 \downarrow \Delta 3\}$, *where* $\Delta 3 = \Delta 1 \cap \Delta 2$.

Definition 18 Intersection.

$$\_\boxdot\_ : \mathbb{P}\, S \times \mathbb{P}\, S \rightarrow \mathbb{P}\, S$$

$s = s1 \boxdot s2 = \{c1 \cap c2 | c1 \in c2 \wedge c2 \in s2\}$.

We can prove that $s = s1 \boxdot s2 = (s1 \boxtimes s2) \downarrow \Delta 3$, *where* $\Delta 3 = \Delta 1 \cap \Delta 2$.

Definition 19 Union.

$$\_\boxplus\_ : \mathbb{P}\, S \times \mathbb{P}\, S \rightarrow \mathbb{P}\, S$$

$s = s1 \boxplus s2$ is computed as follows:

$\Delta_1 = \bigcup_{c \in s_1} dim(c)$, $\Delta_2 = \bigcup_{c \in s_2} dim(c)$, *and* $\Delta_3 = \Delta_1 \cap \Delta_2$

1. Compute $X1: X1 = \{c_i \cup cj \uparrow \Delta_3 | c_i \in s_1 \wedge c_j \in s_2\}$

2. Compute $X2: X2 = \{c_j \cup ci \uparrow \Delta_3 | ci \in s_1 \wedge c_j \in s_2\}$

3. The result is : $s = X1 \bigcup X2$.

Earlier we have shown that the results of $ci \rightleftharpoons cj$ and $ci \rightharpoonup cj$ are sets of contexts. So the relational operators $\boxtimes$ (join), $\boxdot$ (intersection), and $\boxplus$ (union) can also be applied to the expressions $ci \rightleftharpoons cj$ and $ci \rightharpoonup cj$, where *ci* and *cj* are contexts.

4.4 Context Set Expressions

Informally, a context set expression is an expression involving sets of contexts and context set operators. Let *s* ranges over a set of contexts, *S* over a context set expression and D over a dimension set. A formal syntax for context set expression *S* is shown in Figure 2(left column).

syntax				precedence
S ::=	s	|	$S \mid S$	$\downarrow, \uparrow, /$
|	$S \oplus S$	|	$S \ominus S$	|
|	$S \downarrow D$	|	$S \uparrow D$	$\oplus, \ominus$
|	$S \boxtimes S$	|	$S \boxdot S$	$\boxtimes, \boxdot, \boxplus$
|	$S \boxplus S$	|	$S/\langle d, t\rangle$	

■ Fig. 2 Formal Syntax of Context Set Expressions and Precedence Rules for Context Set Operators.

In order to precisely calculate a context set expression, we define the precedence rules for the context set operators. These are shown in Figure 2(right column) (from the highest precedence at the row to the lowest precedence in the bottom row). Parentheses will be used to override this precedence when needed. Operators having the same precedence will be applied from left to right.

4.5 Box Notation

In many applications it is of special interest to consider a set of contexts, all of which have the same dimension set and the tags corresponding to the dimensions in each context satisfy a given constraint. We use the notation *Box* to denote such a set when the dimension set is $\Delta = \{d_1, \ldots, d_k\} \subset DIM$ and *p* is a logical expression.

Note that in *p*, we are allowing the dimensions as variables, denoting the current tags. That is, if $p(d_1, d_2) = d_1 < d_2$, it means the current tag of d_1 is less than the current tag of d_2 in the context that has dimensions d_1 and d_2. A formal definition follows:

Definition 20 *A Box set (or a Box for short) is a set of simple contexts with the same domain. Let* $\phi \neq \{d_1, \ldots, d_k\} \subseteq DIM$ *be a set of dimensions and p be an expression in which the* d_i $(1 \leq i \leq k)$ *may occur as variables. Then* $Box[d_1, \ldots, d_k \mid p] = \{c \in S \mid dim(c) = \{d_1, \ldots, d_k\}$ *and p is true when, for each I,* d_i *is assigned the value* $c(d_i)\}$.

The dimension $\Delta(b)$ *of an nonempty box b is the dimension of any (all) its elements.*

The set of Box sets (or Boxes for short) are all sets of simple contexts all of which have the same domain. It is easy to show that anything defined by the Box notation is a Box.

4.6 Using Context Formalism in System Development

The two key terms in the study of context-aware systems are *context* and *awareness*. Awareness is of two kinds. One kind is the internal monitoring of the system, called *self-awareness* or *internal awareness*. System contexts are dynamic and consequently self-awareness varies from context to context. The other kind is the external monitoring of the system, called *external-awareness*. External awareness, also known as perception, is normally achieved through sensors and other stimuli, say from users or other system elements. External contexts change as and when the system environment changes, and such changes cause changes to external awareness. The system must use the knowledge it gained from its perception, apply it to the changing internals, and react by either triggering an internal state change or providing an external service. Hence, we must use a context formalism in which both self-awareness and externalawareness can be represented and reasoned about. Using context calculus we can compute dynamically different contexts, combine external and internal context, and calculate an internal context corresponding to an observed external context. Without the formalism such calculations can only be done in an ad hoc manner.

Context calculus has been implemented in C#. This context toolkit is portable and can be used as a plug-in for any context-aware application development. The component-based architecture given in [43] illustrates our approach to using context formalism for developing context-aware systems. Such an approach can be adapted to any context-aware application, including service-oriented systems [47], web services [48], and trustworthy systems [46].

References

[1] Edward A. Ashcroft, William W. Wadge. Lucid - A Formal System for Writing and Proving Programs. SIAM J. Comput. 5(3): 336-354, 1976.

[2] V. Akman, and M. Surav. Steps toward formalizing context AI Magazine 17, 55C72,1996.

[3] V. Akman, P. Bouquet, R. Thomason, and R. Young, eds. Modeling and Using Context Proceedings of CONTEXT2001: Third International and Interdisciplinary Conference on Modeling and Using Context (Dundee, Scotland, 27C30 July 2001), Lecture Notes in Artificial Intelligence, Vol. 2116, Berlin: Springer Verlag.

[4] V. Alagar, and K. Wan. Context Based Enforcement of Authorization for Privacy and Security in Identity Management. IFIP International Federation for Information Processing. Springer Series Vol. 261, Editors: de Leeuw, E., Fischer Hubner, S., Tseng J. C. and Borking, J. p. 25-38, 2008.

[5] P. Bouquet, L. Serafini, P. Brezillon, M. Benerecetti, and F. Castellani, eds. Modelling and Using Context. Proceedings of CONTEXT99: Second International and Interdisciplinary Conference on Modelling and Using Context (Trento, Italy, 9C11 September 1999), Lecture Notes in Artificial Intelligence, Vol. 1688, Berlin: Springer-Verlag.

[6] P. Bonzon, M. Cavalcanti, and R. Nossum, eds. Formal Aspects of Context. Applied Logic Series, Dordrecht: Kluwer Academic Publishers,2000.

[7] B. Brumitt, B. Meyers, J. Krumm, A. Kern, S. Shafer. Easy-Living: Technologies for Intelligent Environments. Proceeding of the 2nd International Symposium on Handheld and Ubiquitous Computing (HUC 2000), Bristol, England, 2000.

[8] O. Bucur, P. Beaune, and O. Boissier. Representing Context in an Agent Architecture for Context-Based Decision Making. In Proceedings of CRR'05 Workshop on Context Representation and Reasoning. 2005.

[9] R. Carnap. Meaning and Necessity. Chicago University Press, 1947. Enlarged Edition 1956.

[10] K. Cheverst, K. Mitchell, and N. Davies. Design of an object model for a context sensitive tourist GUIDE. Computers & Graphics, 6(23):883-891, 1999.

[11] H.H. Clark and T.B. Carlson. Context for Comprehension. In J. Long and A. Baddeley, editors, Attention and Performance IX, pp.313-330, Lawrence Erlbaum Associates, Hillside, NJ, 1981.

[12] E. Cobden. Speeches46, 1849.

[13] A. K. Dey, G. D. Abowd, D. Salber. A Conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Human-Computer Interaction, Vol. 16, pp. 97-166, 2001.

[14] D.Dowty, R.Wall, and S.Peters. Introduction to Montague Semantics. Reidel Publishing Company, 1981.

[15] W. K. Edwards. Putting Computing in Context: A Infrastructure to Support Extensible Context-Enhanced Collaborative Applications. ACM Transactions on Computer-Human Interactions, Vol. 12, No.4, pp. 446-474, 2005.

[16] F. Giunchiglia, L. Serafini, E. Giunchiglia, and M. Frixione. Non-omniscient belief as context-based reasoning, pp. 548C554 in Proceedings of the 13th International Joint Conference on Artificial Intelligence, Chambery, France, 1993.

[17] F. Giunchiglia. Contextual reasoning. Epistemologia, Special Issue on I Linguaggi e le Macchine XVI, 345:C364, 1993.

[18] P. D. Gray, and D. Salber. Modeling and Using Sensed Context Information in the Design of Interactive Applications. Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction (EHCI 2001), Toronto, Canada, May 2001.

[19] W. G. Griswold, R. Boyer, S. W. Brown, and T. M. Truong. A Component Architecture for an Extensible, Highly Integrated Context-Aware Computing Infrastructure. Proceedings of the 25th International Conference on Software Engineering (ICSE 2003), Portland, USA, May 2003.

[20] V. Grassi, and A. Sindico. Towards Model Driven Design of Service-Based Context-Aware Applications. ESSPE'07, Dubrovnik, Croatia, pp. 69-74, September 2007.

[21] A. Held, E. Schill, and S. Buchholz. Modeling of context information for pervasive computing applications. In International Conference on Information Systems, Analysis and Synthesis, 2002.

[21] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster. The Anatomy of a Context-Aware Application. Mobile Computing and Networking, pp. 59-68, 1999.

[22] K. Henricksen, J. Indulska. A Software Engineering Framework for Context-Aware Pervasive Computing. Proceedings of the 2nd IEEE Conference on Pervasive Computing and Communications (PerCom 2004), Orlando, USA, Mar 2004.

[23] K. Henricksen, J. Indulska, and A. Rakotonirainy. Modeling Context Information in Pervasive Computing Systems. Proceedings of the 1 st International Conference on Pervasive Computing, Zurich, Switzerland, Aug 2002.

[24] J. Kjeldskov, and M. B. Skov. Exploring contextawareness for ubiquitous computing in the health care domain. Pers Ubiquit Comput, Vol. 11, 2007, pp.549-562.

[25] Z. Mammar, D. Benslimane, and N.C. Narendra. What can CONTEXT do for Web Services? CACM, Vo. 49, No. 12, pp. 98-103, 2006.

[26] J. McCarthy Formalizing Common Sense. Papers by John

McCarthy, V. Lifschitz, ed., Norwood, NJ: Ablex Publishers, 1990.

[27] J. McCarthy, and S. Buva~c. Formalizing context (Expanded notes) Computing Natural Language, pp. 13-50

[28] C. Mundie, P. de Vries, P. Haynes, and M. Corwine.

Trustworthy Computing - Microsoft White Paper. Microsoft Corporation, October 2002.

[29] Joey Paquet. Intensional Scientific Programming Ph.D. Thesis, D´epartement d'Informatique, Universite Laval, Quebec, Canada, 1999

[30] A. Ranganathan, R. H. Campbell, A. Ravi, and A. Mahajan. ConChat: A Context-Aware Chat Program. IEEE Pervasive Computing (vol. 1, no. 3), pp. 51-57, 2002.

[31] A. Ranganathan, J. Al-Muhtadi, RH. Campbell. Reasoning about Uncertain Contexts in Pervasive Computing Environments. IEEE Pervasive Computing, Vol. 3, No. 2, pp. 62-70, 2004.

[32] G. Rey and J. Coutaz. Contextor: capture and dynamic distribution of contextual information. In Proceedings of the 1 st French-speaking conference on Mobility and ubiquity computing, pages 131-138, New York, NY, USA, 2004. ACM Press.

[33] G.-C. Roman, C. Julien, and Q. Huang. Network abstractions for context-aware mobile computing. In ICSE02: Proceedings of the 24th International Conference on Software Engineering, pages 363-373, New York, NY, USA, 2002. ACM Press.

[34] M. Sato, T. Sakurai, and Y. Kameyama. A simply typed context calculus with first class environments. Journal of Functional and Logic Programming, No. 4, pp. 359-374, 2002.

[35] A. Schmidt, K. A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven, and W. Van de Velde. Advanced Interaction in Context. Proceedings of the 1 st International Symposium on Handheld and Ubiquitous Computing (HUC), Karlsruhe, Germany, Sep 1999.

[36] N. A. B. N. Schilit and R. Want. Context-aware computing applications. In 5th IEEE Workshop on Mobile Computing Systems and Applications, 2003.

[37] Sharma, N., 1995, On formalising and reasoning with contexts, Technical Report Technical Report 352, Department of Computer Science, The University of Queensland, Australia.

[38] E. Shortcliffe. MYCIN: Computer-based Medical Consultations. Newyork: Elsivier

[39] T. Strang, C. Linnhoff-Popien, and K. Frank. CoOL: A Context Ontology Language to enable Contextual Inter - operability. in Proceedings of 4th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems (DAIS2003), volume 2893 of Lecture Notes in Computer Science (LNCS), pages 236-247, Paris/France, November 2003. Springer Verlag.

[40] T. Strang and C. Linnhoff-Popien. A context modeling survey. In First International Workshop on Advanced Context Modelling, Reasoning And Management, Nottingham, England, Sept. 2004.

[41] R.Thomason, editor. Formal Philosophy, Selected Papers of R.Montague. Yale University Press, 1974.

[42] K. Wan, V.S. Alagar, J. Paquet. An Architecture for Developing Context-aware Systems. In the Proceedings of 2nd International Workshop on Modeling and Retrieval of Context (MRC2005), LNCS Springer-Verlag, Vol. 3946, pp. 48-62, 2005.

[43] K. Wan, and V. Alagar. An Intensional Programming Approach to Multi-agent Coordination in a Distributed Network of Agents. In Proceedings of the 3rd International Workshop on Declarative Agent Languages and Technologies (DALT). Utrecht, The Netherlands, LNCS Vol. 3904, p. 205-222, 2005.

[44] K. Wan. Lucx: Lucid Enriched with Context. Ph.d Thesis, Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada, January 2006.

[45] K. Wan, K. and V. Alagar. An Intensional Functional Model of Trust. Trust Management II. Editors: Y. Karabulut, J. Mitchel, P. Hermann, C.D. Jensen, IFIP Springer Series, p. 69-85, 2008.

[46] K. Wan, V. Alagar. A Context-aware Trust Model for Service-oriented Multi-agent Systems. Proceedings of 1 st International Workshop on Quality-of-Service Concerns in Service Oriented Architectures (QoSCSOA 2008), LNCS Vol.5472: Eds: G. Feuerlicht and W. Lamersdorf, pp. 221-236, 2009.

[47] K. Wan, M. Mohammad, and V. Alagar. A Formal Model of Business Application Integration from Web Services (Position Paper). Current Trends in Theory and Practice of Computer Science (SOFSEM09),LNCS Vol.5404, pp. 656-667, 2009.

[48] T. Winograd. Architecture for Context. Human-Computer Interaction, Vol. 16, pp. 401-419, 2001.

First A. Author KaiYu Wan received her ph. degree of Computer Science from Concordia University in 2006, master degree in 1999 and bachelor degree in 1996 from NanJing University of Science and Technology. Currently she is a Lecturer at Department of Computer Science, East China Normal University. She has published 20 referred papers. Her research interests are in development of systems such as Trustworthy Systems, Component-Based Software Systems, Context-Aware System and Multi-Agent Systems, and in Programming Languages and Compilers. She enjoys reading and traveling.

ICT in Universities of the Western Himalayan Region in India: Status, Performance- An Assessment

Dhirendra Sharma[1] and Vikram Singh[2]

[1]University Institute of Information Technology, Himachal Pradesh University, Shimla, Himachal Pradesh 171 005, India.

[2]Department of Computer Science and Engg, Ch. Devi Lal University, Sirsa, Haryna 125 055, India.

Abstract

The present paper describes a live project study carried out for the universities located in the western Himalayan region of India in the year 2009. The objective of this study is to undertake the task of assessment regarding initiative, utilization of ICT resources, its performance and impact in these higher educational institutions/universities. In order to answer these, initially basic four- tier framework was prepared. Followed by a questionnaire containing different ICT components 18 different groups like vision, planning, implementation, ICT infrastructure and related activities exhibiting performance. Primary data in the form of feedback on the five point scale, of the questionnaire, was gathered from six universities of the region. A simple statistical analysis was undertaken using weighted mean, to assess the ICT initiative, status and performance of various universities. In the process, a question related to Performance Indicator was identified from each group, whose Coefficient of Correlation was calculated. This study suggests that a progressive vision, planning and initiative regarding academic syllabi, ICT infrastructure, used in training the skilled human resource, is going to have a favourable impact through actual placement, research and play a dominant role at the National and International level.

Keywords: *Information and Communication Technology (ICT), Initiative, Status, Performance, Assessment, Impact, Performance Indicator.*

1. Introduction

During the past decade there has been a very rapid development in activities related to Information and Communication Technology (ICT), (Sadagopan 1998), in various Universities and Institutions of higher learning. Such a transformation was triggered by more and more awareness of microwaves and tele-communication. Its pervasiveness has affected almost all fields from micro- to astronomical level via class rooms. ICT in respective fields makes education extremely effective, efficient and engaging due to its power of problem solving and application.

The impetus at the national level was set in motion by the Task Force 'Technology Information Forecasting and Assessment Council' (TIFAC), for Technology Vision 2020. One of the important pilot documents "India 2020, A Vision: for the New Millennium" (1998) prepared by Technology Information Forecasting and Assessment Council (TIFAC), under the leadership of Dr. A.P.J.Kalam, provided a blue print of the `Technology Vision for India'. According to this document ICT can enhance critical thinking, information handling skills, level of perception, problem solving capability and adding value to research in educational institutions. It not only highlighted the importance of higher educational institutions/ universities to take prompt and appropriate initiatives in this direction, but has given further direction for planning ICT strategies in which the role of higher educational institutions in India is going to be very crucial. It emphasized that the Indian human resource has great learning capabilities with a core competence in technology along with the spirit of entrepreneurship and competitiveness. It strongly advocated the development of human resource cadre and the role of higher educational Institutions that will be the foundation of the technological advancement in the country.

This prompted Government of India to take a major ICT initiative to lay down the ICT policy for whole of the country, which is reflected in the planning and implementation by the Ministry of Human Resource Development (MHRD) at various levels particularly through higher educational institutions/ universities of the country. At the national level, it is being implemented vigorously in the form of e-governance and is being coordinated by National Informatics Centers (NIC) throughout the country. University Grants Commission (UGC), All India Council of Technical Education, Department of Science & Technology (DST), and the Indian Council of Agricultural Research (ICAR) are also encouraging higher education Institutions and Universities in this direction. Latest telecommunication policy (Tandon 2002) is quite consistent with the initiatives mentioned above.

In Jan. 2009, the Cabinet Committee on Economic Affairs, Government of India, gave its approval for a new scheme by the name of National Mission on Education through Information and Communication Technology (NMEICT) submitted by the MHRD, Govt. of India, for the 11th Five Year Plan. This mission has been envisaged to leverage the potential of ICT, in providing high quality personalized and interactive knowledge modules for all the learners in higher education institutions in ‘any time any where mode’. Further, it has to ensure on-line access to all the ICT resources viz. e- contents, the connectivity and virtual laboratories, to 18,000 colleges and all the universities of the country and thus to bridge the digital divide.

There has been another body, National Association of Software and Services Companies (NASSCOM), in the field of ICT, dealing with the ICT infrastructure, research & development and trade since 1988. According to recent report of NASSCOM (Strategic Review-2009), it has emphasized the availability of skilled and employable talent. In order to achieve this concentrated effort to enhance talent availability and quality are needed from all the concerned sections namely the government academia and the industry. The role of academia is supposed to be very critical in setting up research laboratories. Specific initiatives like faculty development program, upgrading the curriculum, launching internship program and academia- industry collaboration can help to bridge the gap towards the development of talent. NASSCOM is particularly interested in the ICT- business process outsourcing (ICT-BPO) industry in India, which has become a grown economy of the country. Further, this body is a partner with the government of India and various state governments of the country. It has also played a crucial role in the formulation of ICT policies which endeavors to narrow down the digital divide in the country and enables all citizens to enjoy the benefits of ICT.

At the international level, United Nations Educational, Scientific and Cultural Organization UNESCO (2007- 08) has also prepared a document for Asia– Pacific countries, for the implementation of ICT programs through the higher educational institutions. Various ICT strategies of driving the higher education towards excellence in ICT is described in the above document, in support of the core areas of higher education. i.e., learning, teaching, research and training programs.

All these innovative ideas could propagate through the concept of diffusion and adoption. It seems that adoption of the latest information and communication technology is making all the difference particularly in the universities. During the last two decades there has been several attempts to understand information technology diffusion. At the empirical level, the concept of diffusion was reviewed by Fichman in 1992. He has dwelled upon innovation diffusion theory, in particular, how to improve technology assessment, adoption and implementation. A framework was discussed in terms of classical diffusion adopted mainly by individuals and organizations alongwith achieving critical mass beyond which the innovation is universally adopted. A lot of scientific work exists on the diffusion of innovation and its adoption given by Attewell (1992), Rogers (1995), Farquhar & Surry (1994), Anderson et al (1998) and Wilson et al (2000). Attewell (1992) put the work on diffusion in two categories a) Adoption studies and b) Macro diffusion studies. The former is primarily concerned with understanding of innovations and its assimilation during a time of adoption. The latter is concerned with characterizing the rate a pattern of adoption of technology. This type of work was understood in terms of mathematical models of the diffusion process by Mahajan & Peterson (1985) and Mahajan et al. (1990). Innovation diffusion has also been understood in terms of mathematical approaches given by Karmeshu and Pathria (1980), Lal et al (1988), Karmeshu et al (1992) and by Karmeshu and Jain (1995).

In recent years, a very significant work on ‘Global Diffusion of the Internet’ by Wolcott and Goodman (2003) appeared in Indian context. Walcott and Goodman (2003), presented a vision of new India as a measure of IT power, fully integrated with a global economy. The key to this vision was obviously the internet for enabling the technology based changes. They provided an analytical framework which broadly consisted of dimensions and determinants. Dimensions contained six variables namely organizational infrastructure, geographic dispersion, connectivity infrastructure, pervasiveness, sectoral absorption and sophistication of use. These variables are supposed to capture the state of internet within a country at a given point of time. Each of these variables was judged at five different levels. Determinants represent various factors, to understand the observation of the “state”. While discussing the IT action plan, they further elaborated on distinctive features of the internet in India, for a continued dynamic growth, in three steps (a) The government policy (b) The nature of relationship between the government, the state owned telecommunications service providers and private sectors as a critical variable. (c) The policy makers try to strike a balance between the interest of the society and those of individuals. Later, this framework has been applied to understand the initiative, status and performance in Ghana by Foster, et al (2004) in Togo by Bernstien and Goodman (2005) and in Kenya by Ochara et al (2008). Such a comprehensive analysis was carried out in a broader context at the national level for different countries.

In view of the above initiatives from different quarters, most of the universities and higher educational institutions have started adopting the latest technology, focusing on the development of the skilled human resource, as part of its responsibility, in the field of ICT, encompassing all the disciplines, particularly in technical fields, productively and constructively. Most of the universities are undertaking the job of training the technical personnel's in the respective specializations or in the area of its excellence. These programs have been going on for the last almost one decade.

Almost after a decade, one may wonder how efficient the system, particularly the university/higher educational institutions has become. Initiatives to have an assessment of ICT infrastructure, its utilization and overall performance and efficiency, of any higher educational institution, are quite essential, in identifying, planning and achieving the ICT strategies at the respective levels. Essentially, this has formed the basis of motivation to undertake the study of the ICT initiatives, status and its academic performance in various universities in the western Himalayan region of the country and then to analyze, using simple statistical methods, the ICT based academic/ technological status and performance of different universities, and assess in a comparative manner, their vision, planning, initiative, status activities and impact.

The University Grants Commission (UGC), New Delhi, took steps to establish an autonomous institution, National Assessment and Accreditation Council (NAAC), for comprehensive assessment of various universities and to place them according to certain ranking. NAAC (2007) has developed a framework for higher education based on the promotion and sustenance of quality of teaching-learning, research and training programs. Their most significant core value is quest for excellence/ innovations using the latest technological trends and fostering global competence among students. They have devised seven assessment criteria namely, curricular, teaching- learning, research & application, innovative, infrastructure, student support and leadership/governance, aspects, to capture the micro- level quality indicators by using differential weightages. Some other national agencies TOI group 2007, also tried to assess, independently, various agriculture, horticulture and technical institutions. These recommendations and respective gradations are available at the national level. This information may be utilized for comparison with the findings of this paper in the context of ICT.

Recently Mokhtar et al (2007), presented the state of academic computing in Malaysian colleges in which they tired to answer the central questions regarding indicators of assessments and performance of academic computing. They adopted the value chain concept originally proposed by Porter (1985), in connection with some business idea, to describe the relationship between academic ICT activities. This framework consisted of two groups namely primary activities and support activities. Primary activities included use of ICT in learning, teaching, research and training as the core service area and then for enhancing the value in each. It further included remote access to data, faster and more precise data processing, simulation of complex systems, collaboration between research groups. The support activities were linked with primary activities to improve the effectiveness/ efficiency and contain ICT vision, policies and standards, ICT infrastructure, ICT services. Backer and Mohamed (2008) elaborated on the benefits of ICT in teaching. These studies have certain limitation keeping same immediate objectives in mind, in that country. However, the analysis presented by Mokhtar et al and Bakar & Mohamed was academic in nature, carried out for colleges only and not for universities and higher educational institutions of training and research.

We plan to undertake this work for universities of western Himalayan region in India. This region is typically a difficult terrain in which people have to really struggle for livelihood, health and education. It is in this context the government of India is paying special attention to alleviate the level of education using the full potential of ICT, in this region. In view of this, a little different framework for the universities of this region is required. A framework will be developed for assessing the initiative, status and performance of different universities. Initiative will imply vision and planning of the ICT programs in the universities. ICT status would mean complete ICT infrastructure including local area network facility, internet network security, mobile computing access, system application software, website and information system, teaching display technologies, ICT technical staff, ICT budget allocation, e- library, e- placement/ alumni portal. ICT performance will be related with various activities based on ICT infrastructure. It would include teaching; learning and ICT based research at the conventional level and at the advanced level in the form of ICT training programs. This performance is going to have an impact of the ICT programs, functional in the universities, in the form of actual placements of the outgoing students, research publications and actual problem solving. This outline of the framework will be discussed explicitly in the next section.

The objective of this study is to undertake the task of assessment regarding initiative, utilization of ICT resources, its performance and impact in these higher educational institutions/ universities. In order to achieve this, an overall status report is prepared using the framework of this paper. A questionnaire was prepared accordingly, related to various ingredients of ICT, initiative, status and performance. Then a survey was conducted in different universities located in the western Himalayan region, as per the questionnaire. The analysis and comparison of the initiative, status and performance through performance indicators in respective universities has been carried out. The corresponding benchmarking is expected to identify the future ICT strategies in order to improve their status as well as the product in the form of professionally skilled human resource in the field of ICT. The reference period of the study has been the year 2009.

The paper is organized in six sections. After the introduction in section 1, the basic framework will be presented in section 2. The methodology, preparation of questionnaire, a survey and analytic procedure will be given in section 3. Section 4 will deal with the results obtained from the analysis of the primary data. Summary and conclusions are given in Section 5. Future strategies are discussed in the final Section.

2. Basic Framework

Before discussing the methodology it seems proper to have a framework on the basis of which the study of ICT initiative, status, performance and impact in the field of ICT in different universities and higher technical institution, is carried out.
The important building blocks are supposed to be Performance Indicators (PI). One group of authors (Riley and Nuttall 1994) suggests that Performance Indicators must be something quantitative, whereas the other group takes a wider view in favour of qualitative and descriptive statement for PI. The latter view is adopted by International Standards Organisation (1998). Both these measures, quantitative as well as qualitative, will be used here that will allow us to have a complete view of richness and diversity of the ICT based academic/ technological performance and the related activities. With these it should be possible to assess and judge the effectiveness, efficiency and impact of an institution in terms of ICT performance to have the possible projection to finally decide about the future strategies.

In deciding the basic framework, the question is what are the ICT ingredients/ components and Performance Indicators for assessing the ICT initiative, status and their performance at different levels. The basic ingredients of ICT are clearly classified among four tiers, containing 117 questions divided in 18 groups, A-R. These groups, in their respective tier, depict different academic components of ICT and the related activities. Schematic representation of four tier framework is given in Fig. 1.

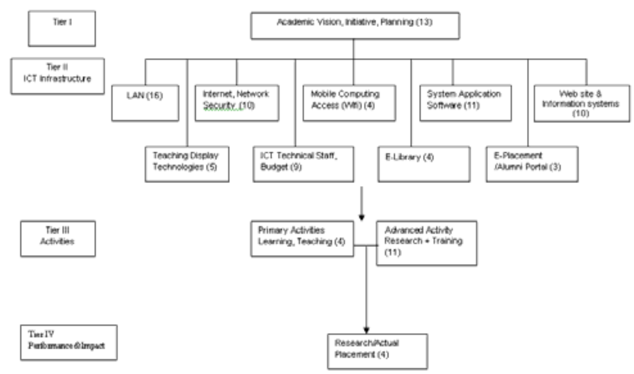

Fig. 1: Four Tier framework. Each box contains the main ICT component along with the number of questions in it

The basis of formulating this framework has been the recommendations contained in the documents available at the national and international level, as mentioned in the introduction. This framework is supposed to be most suited to the university and higher educational institutions which are mainly the centre of research where knowledge is created. In each group one of the questions, the last one, is related to the performance. All these performance related questions, are placed in another group, S. These groups, alongwith the Performance Indicator in each, are presented in Table I:

Table I: Various Groups of ICT Components in an Institution.

S.No	Group	ICT details at University level	Performance Indicators	No. of Questions	Tier in Framework
1.	A	ICT Vision	Implementation	03	TIER-I
2	B	ICT Initiatives	Execution	04	
3	C	ICT Planning	Achievement	06	
4.	D	ICT Infrastructure	Utilization	13	TIER-II
5.	E	Local Area Networks Facility	Users	16	
6.	F	Internet, Network Security	Speed	10	
7	G	Mobile Computing access	Access 24x7 basis	04	
8.	H	System Applications S/W	Overall user's Satisfaction	11	
9.	I	Web site and information systems	Online communication	09	
10.	J	Teaching Display Technologies	Availability	05	
11.	K	ICT Technical Staff	Trouble Shooting	05	
12.	L	ICT Budget allocation	Utilization	04	
13.	M	E- Library	Availability	04	
14.	N	E-Placement/Alumni Portal	Alumni Portal	03	
15.	O	ICT based Teaching and learning	Effective use of ICT	04	Tier III
16.	P	ICT Based research	Satisfaction level	07	
17	Q	ICT Training programmes	No. of programmes organized	04	
18.	R	Impact of ICT: Research & Actual Placement	Placement /Research	04	TIER IV
19.	S	Performance	Success rate	18	

3. Methodology

The universities located in the western Himalayan region imply those spreading over the state of Himachal Pradesh and Jammu & Kashmir. There are mainly 7 universities located in the western Himalayan region of India, out of which 6 universities have been selected purposely. Five universities belong to the state or the central government, and one is a private university but deemed university approved by the University Grants Commission under the aegis of Ministry of Human Resource Development, Government of India. These are Himachal Pradesh University (Shimla), Y.S. Parmar Horticulture University (Solan), C.S.K. Agriculture University (Palampur), National Institute of Technology (Hamirpur), Jammu University (Jammu) belonging to the government and J.P. University (Solan) as deemed university.

In order to have the information regarding ICT initiative, status and performance in the universities of this region, the questionnaire was given to the concerned head of the department of ICT/ IT/ Computer Science in the institution/ universities personally. The concerned head was requested to provide information/ feed back as per the questionnaire. Thus the information was gathered personally by the authors.

In the structured questionnaire, the feedback to various queries were on five point scale arranged in a particular order that revealed the natural flow from the lowest level to the highest level, in increasing order of sophistication. That is why it was thought reasonable to give a weightage of 1 to 5 respectively to each level in increasing order.

Simple standard statistical tools were used to analyse the data groupwise to find the Weighted Mean, Standard Deviation and Coefficient of Variation (CV) keeping in mind the weight of respective levels. The CV indicates the variation around its weighted mean in the series, the lesser is the CV, more consistent and stable is the series. However, the weighted mean was found to be a better measure as compared to CV, to understand the trend in a particular group. The use of median was another alternative but in view of the fact that the data was not having extreme variation and due to the limitation of the median method, weighted mean was logically preferred. Another statistical quantity 'Coefficient of Correlation' among various groups was also calculated which has also been used to ascertain the proper groupings in the questionnaire.

A simple tabular analysis was carried out to find out the results. In view of the fact that queries were replied on five point scale with the respective weight from one to five, firstly weighted mean of each group was calculated followed by overall mean in the respective tier. The weighted mean was found by picking up the proper value of the response in a particular level on the five point scale (say L1 to L5). Then dividing it by the total value of the levels, i.e. sum of 1 to 5, multiplied by the number of questions in the group. This is repeated for all the groups for six universities of the region. These are presented in Table II. The Pearson's Coefficient of Correlation of each group with the Performance Indicator group is given in the last column of the Table II.

Table II: Weighted Mean and Pearson's Correlation Coefficients of various groups with Group R

Tier	Groups	J.P.U Solan	NIT Hamirpur	CSK.U Palampur	J.U. Jammu	Y.S.P.U. Nauni	H.P.U, Shimla	Correlation
Tier I	Group A	0.3	0.33	0.17	0.3	0.27	0.33	0.60
	Group B	0.27	0.27	0.2	0.22	0.18	0.28	0.70
	Group C	0.29	0.28	0.21	0.21	0.24	0.32	0.48
Weighted. Mean		0.29	0.30	0.19	0.24	0.23	0.31	
Tier II	Group D	0.33	0.33	0.28	0.22	0.22	0.3	0.70
	Group E	0.32	0.33	0.21	0.23	0.24	0.2	0.93
	Group F	0.27	0.28	0.23	0.26	0.19	0.28	0.64
	Group G	0.2	0.29	0.23	0.24	0.17	0.17	0.60
	Group H	0.07	0.33	0.07	0.27	0.16	0.07	0.49
	Group I	0.24	0.24	0.16	0.18	0.19	0.19	0.93
	Group J	0.15	0.27	0.15	0.14	0.07	0.12	0.80
	Group K	0.24	0.33	0.18	0.16	0.13	0.18	0.92
	Group L	0.33	0.27	0.2	0.2	0.09	0.11	0.77
	Group M	0.23	0.33	0.23	0.2	0.23	0.17	0.71
	Group N	0.31	0.27	0.13	0.17	0.23	0.17	0.79
Weighted Mean		0.24	0.30	0.10	0.20	0.17	0.18	
Tier III	Group O	0.27	0.27	0.25	0.2	0.26	0.15	0.38
	Group P	0.2	0.27	0.07	0.18	0.07	0.23	0.95
	Group Q	0.2	0.2	0.09	0.12	0.07	0.08	0.91
Weighted Mean		0.22	0.25	0.14	0.17	0.13	0.15	
Tier IV	Group R	0.3	0.32	0.17	0.23	0.15	0.22	0.63
	Group S	0.28	0.32	0.17	0.21	0.18	0.2	

4. Results and Discussions:

The weighted mean for Tier I, vision, initiative and planning, for different universities has been presented in Fig.2, which depicts a comparative view among the universities.

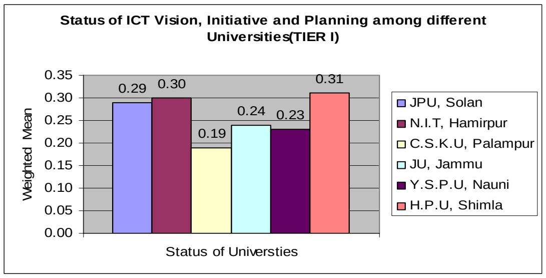

Fig.2 Tier I.

This diagram shows that there is not much difference corresponding to the vision and planning, particularly in the first three universities i.e. JP University, NIT and H.P.University. At the level of vision, initiative and planning all the universities thought big with a minor difference. Standard deviation and the coefficient of variation were also calculated. A range of coefficient of variation (%) was found to be 14.54%, 15.13%, 18.04%, 24.05, 32.66% and 39.44% respectively for H.P.University, J.P.University, NIT, Hamirpur, Nauni University, Jammu University and Palampur University. It reveals that first three universities are having more consistency in ICT vision, initiative and planning than the later three universities. Because for the same vision, the ICT initiatives and planning were at different levels. Further, the level of vision also differed from one university to another.

The overall weighted mean of all the different groups related to Tier II which mainly deals with overall ICT infrastructure among different universities in the region, has been presented in Fig. 3.

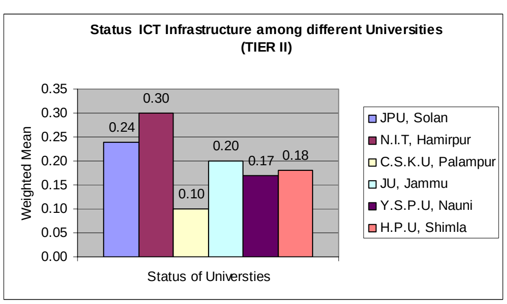

Fig. 3. Tier II

From this diagram, it may be seen that NIT Hamirpur is having the largest ICT infrastructure followed by J.P. University, Jammu University, H.P.University, Nauni University and Palampur University. This difference may be mainly attributed to advanced ICT facilities available in NIT, Hamirpur like better Local Area Network (LAN) with internet facilities, video-conferencing facility available with each faculty, Mobile (Wi-Fi) computing facility, IP telephony and assured access to all the digital tools/ resources on the campus. These facilities are also available in the hostels and faculty accommodation as well, along with an effective web site and information system. In this respect NIT Hamirpur has an edge over other universities. Other contributing factors are the effectiveness of ICT training progammes, organized for the faculty, students and professionals as the drive for human resource development programs. Finally, the placement of professionals in reputed establishments adds much more weightage in this direction.

This figure further shows that the difference between Jammu University and Himachal Pradesh University is very little. It is particularly because of the fact that Jammu University is having higher internet bandwidth, better mobile computing facilities alongwith e-library as compared to those available in Himachal Pradesh University.

Tier III, is concerned with the main ICT activities which can be further divided into primary activities of teaching and learning and advanced activities like sophisticated training & research using ICT for developing professional skills. These are displayed in fig 4.

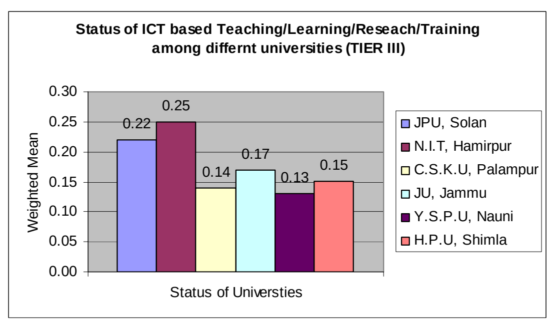

Fig.4 Tier III.

This category deals with various academic activities. Teaching and learning using ICT based tools is one of the effective approaches of procuring knowledge. Researches using e-journals and the material available on the internet have become the lifeline of any good research activity. These ICT tools have provided the academic community a more versatile powered instrument. Further, awareness and training programs at the advanced level to produce professionally skilled technological human resource have become more and more effective using these ICT tools. In this respect, NIT Hamirpur has established superiority over others in this category also, followed by J.P.University, Jammu University and H.P.University etc.

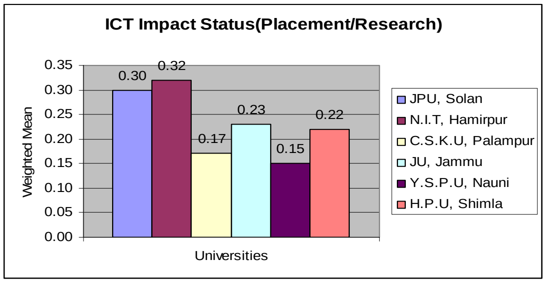

Fig. 5. Tier IV

Tier IV of the framework exhibits the impact of universities at the local (societal) level, at the national and international level as well. The impact is reflected mainly through two components. One is outstanding research supported with patents for the fruitful activities and for the benefit of mankind. Another is the placement of outgoing students trained in professional courses, to reputed universities, institutes, industries and other organizations. The impact factor has been given in Fig.5, which reveals that NIT Hamirpur is again having an edge over other universities in the western Himalayan region of India, followed by J.P.University, Jammu University and Himachal Pradesh University. The main reason for this variation is the value addition through research, particularly supported by the advanced ICT infrastructure. Further, large number of good research publications per year, along with successful collaborative research work also becomes an important factor for the superiority of one university over the other. In addition, actual placement contributes the most to Tier IV in the form of linkages of the professionals with industries.

Towards the end, it was thought reasonable to have Pearson's Coefficient of Correlation (CC) of different groups with the group S depicting the Performance Indicators. These Correlation Coefficients have been given in Table II and displayed in Fig. 6.

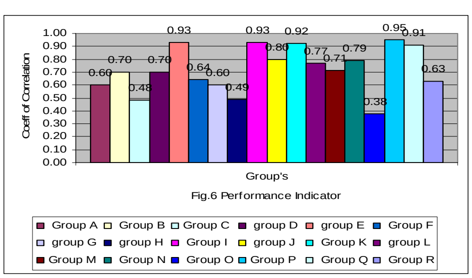

Fig. 6. Correlation with performance Indicators.

On its perusal, the CCs are found to vary between 0.38 and 0.95. Its value, >90, was found for five groups. These groups are E (LAN), I (Website and Information System), K (ICT Staff), P (ICT based Research) and Q (ICT training programs). It speaks for the obvious that the advanced internet facility used in research and training programs has a better correlation with the performance of an institution. This is found quite consistent with the analysis presented above. The next five groups having CCs between 0.70- 0.89, are found to be due to D (ICT Infrastructure), J (Teaching Display Technologies), L (ICT Budget Allocation), M (e-Library) and N (Alumni portal and placement). There is another set of five groups having CCs between 0.6-0.69, which is found just acceptable. These are related with ICT vision, initiative, internet & network security, mobile computing facility and impact of ICT. The remaining three groups having correlations below 0.60 are not acceptable due to inadequate planning in ICT based teaching programs.

5. Summary and Conclusion

In this study we have presented an assessment from the primary data gathered as feedback to a questionnaire divided into 18 different relevant groups in respect of ICT activities starting from vision, initiative, planning, implementation, ICT infrastructure, to the performance/ success of various ICT initiatives in six universities in the western Himalayan region. Performance Indicator of each of the groups was identified that was utilized to gauge the success of ICT initiatives at different stages of the framework.

The vision and planning covered in Tier I are not enough if these are not supported with the proper initiative by a visionary at the highest level in the university system with full support from the government along with a small but dynamic team involved with it. It may be emphasized that a good academic curriculum at the teaching level and better faculty are the key factors which are most essential to impart the latest in the field of ICT.

It further requires equally good infrastructure (Tier II) hardware, software and fast access to the internet in the ICT laboratory.

After all the universities are meant (Tier III) to create knowledge and the professionally skilled dedicated manpower trained not at the local level but at the international level. Professional spirit with dedication of a person is further gauged with a scale on how much busy his academic schedule remains for understanding and for problem solving.

The better performance always results in sound impact (Tier IV) at the national and international level in terms of outstanding research publications and actual placements.

It may be concluded from the analysis of all the components of the four- tier framework that NIT Hamirpur has a clear- cut superiority over the other universities in this region, due to the sound vision, excellent ICT infrastructure and efficient ICT based activities and fast adoption of innovation in the field of ICT, whose impact became known at the national and international level. The other Universities should follow the live example of NIT Hamirpur.

Finally, it is the performance in each of the groups that really matters. The correlation of Performance Indicators (group S) with all the other eighteen groups echoes the same. These correlation coefficients further provided the strength to the conceptualization of groups presented in the questionnaire used in this work.

Our findings about the overall ICT status is on the basis of performance and impact do match with the reported ranking given to universities by NAAC and other agencies. This finding is quite interesting which proclaims that overall progress shown by any of the universities is mainly due to the respective progress through initiative, planning, infrastructure and performance in the field of ICT. One could not think of scoring better ranking at the national or international level without the proper development in ICT which has become synonym with the progress of the university.

6. Future Strategies

The pattern revealed in this paper in respect of initiative, status and performance of different universities in the targeted area of the country points towards future strategies as follows:

Although ICT initiatives in different universities are quite encouraging, its extension must be envisaged in all higher educational institutions of the country. It must include the advanced training programs to prepare professionally skilled manpower using ICT for enhancing the core competence of scientists in the field of ICT and making these programs more meaningful.

These skilled professionals should function as nucleus to further extend such programs to create more skilled human resource.

They must plan with determination, of removing the sufferings of humanity using the earned knowledge to fulfill the objectives of `Transforming the Society' as envisaged in the Kothari Commission report of 1966 and the Technology Vision 2020 document prepared by TIFAC for India

Different courses in the form of e-content inclusive of video, in Indian universities are made available as freeware similar to those of MIT open courseware, as envisioned by NMEICT. It will have an effective long term sustainable impact on society by way of advance ICT enabled education and empowerment of people in the western Himalayan region of the country.

Acknowledgments

Authors express their sincere thanks to Dr. S.P. Saraswat, Agro-Economic centre, H.P.University, Shimla, for very fruitful discussion during data analysis.

References

[1] A. R. Bakar & Shamsiah Mohamed. "Teaching using Information and Communication Technology: Do Trainee Teachers have the Confidence?" International Journal of Education and Development using ICT, Vol 4, No.1, (2008), 374

[2] T. Anderson, S. Vernhagen. et al., "Faculty Adoption of Teaching and Learning Technologies : Contrasting Earlier Adopters and Mainstream Faculty." The Canadian Journal of Higher Education, 28 (1998), pp. 71-98.

[3] P, Attewell, "Technology Diffusion and Organisation Learning: The Case of Business Computing." Organisation Science, 3, (1992), pp 1-19.

[4] A. Berstein, and A.E. Goodman, " Global Diffusion of the Internet VI, The Internet in Togo" Communication of the Association for Information Systems, 15, 2005,pp 371-392).

[5] J.D. Farquhar, & D.W Surry., Adoption Analysis: An Additional Tool for Instruction Developers. Educational and Training Technology 31, pp. 19-25(1994).

[6] G.R. Fichman, Information Technology Diffusion: A Review of Empirical Research, Proceedings of the Thirteenth International Conference on Information Systems (ICIS), December, Dallas, 195-206 (1992).

[7] R.G. Fichman, "The Diffusion and Assimilation of Information Technology Innovations in Framing the Domains of IT Management: Projecting the Future Through the Past", Robert Z., editor, Pinnaflex Publishing (2000)

[8] W. Foster, S.E Goodman, E Osiakwan, and A.Bernstein, "Global Diffusion of the Internet IV: The Internet in Ghana", Communication of the association for Information Systems, 2004, Vol 13, 654-681.

[9] International Standards Organisation. ISO 11620: Information and Documentation- Library Performance indicator. (International Standards Organisation, Geneva), 1998

[10] A.P.J Kalam,.& V .Rajan, India 2020– A Vision for the New Millennium (Penguin Books India), 1998.

[11]Karmeshu & R.K. Pathria, , Stochastic Evolution of a Non-linear Model of Diffusion of Information, Journal of Mathematical Sociology, 7, pp-59-71. 1980

[12]Karmeshu, C.L. Sharma, and V.P. Jain, Nonlinear Stochastic Models of Innovation Diffusion with Multiple Adoption Levels, 51, 1992, pp 229-241

[13] L.S Kothari, Kothari Commission Report (Government of India), 1966

[14] Kulik & Kulik Effectiveness of Computer-based Instruction: An Updated Analysis, Computer in Human Behavior, 7, 75-94. 1991

[15] V.B. Lal, Karmeshu & S. Kaicker, Modeling Innovations, Diffusions with Distributed Time-lag, Technological Forecasting and Social Change, Elsevier Science, 34, 1988, pp.34-113.

[16] V. Mahajan, & R. Peterson, Models for Innovation Diffusion, Braverly Hills, C.A Sage Publications. 1985

[17] V. Mahajan, E. Muller, and F.M. Bass, New Product Diffusion Models in Marketing: A Review and Directions for Research, Journals of Marketing, 54, 1990, pp.1-26

[18] S.A. Mokhtar, R.A Alias, and A.A Rahman,., "Academic Computing at Malaysian Colleges", J. Education and Development Using ICT 3, 2007, 312

[19]MIT open courseware http://ocw.mit.edu/ OcwWeb/web/home/home/index.htm (Accessed on 12/08/2009)

[20]National Assessment & Accreditation Council, New Methodology of Assessment & Accreditation. 2007

[21] NASSCOM Strategic Review, 2009.

[22] N.M. Ochara, J.P.V. Belle, and I. Brown, Global Diffusion of the Internet XIII: Internet Diffusion in Kenya and its Determinants- A Longitudinal Analysis. 23, 2008,pp.123-150.

[23] M.E Porter, Competitive Advantage (Free Press, New York), 1985

[24] K.A Riley, and D.L. Nutall., Edited. Measuring Quality: Education Indicators- United Kingdom and International Perspectives. (The Falmers Press, London) pp.17-40. 1994

[25]P. Resta,. (Ed.), Information and Communication Technologies in Teacher Education, A Planning Guide.(UNESCO, Paris), 2002

[26] E. Roger, Diffusion of Innovations, New York, The Free Press, 1995

[27] L.,Roy, & D. Raitt,., The Impact of IT on Indigenous Peoples, The Electronic Library, 21, 411-13, 2003

[28] S. Sadagopan, "Historical Perspective in IT In India- A status report, April 24,1998. http://www.allindia.com/infotechtechdesk/art1a.htm, 1998

[29]SEUSISS PROJECT., "Survey of European Universities Skills in ICT of Students and staff - Final Report".(Accessed on 08/05/09, from http:// www.intermedia.uib.no/seusiss/) 2003

[30]Tadon.S, ICT Country Status Report: India, 2002. ww.tc.apii.net/~apt2003/India/ Mystatus.htm (Accessed on 02nd November 2009)

[31]UNESCO, ICT Policies of Asia Pacific, 2007-08.

[32] W. Foster, S.E. Goodman, , E. Osiakwan, and A. Bernstein, "Global Diffusion of the Internet VI, The Internet in Togo", Communication of the Association for Information Systems, 15, 2005, pp 371-392.

[33] B.L. Wilson, L.Sherry et al, "Adoption of Learning Technologies in Schools and Universities" Handbook on information technologies for education and training. H.H.Adelsberger, B.Collis and J.M.Pawlowski. New York, Springer- verlag. 2000

[34] P.Wolcott, and S,E. Goodman, , "Global Diffusion of the Internet-I: India: Is the Elephant Learning to Dance?", Communications of Association for the Information Systems 113, 2003, pp 560-646

Dhirendra Sharma is pursuing his Ph.D from the Department of Computer Science and Engineering, Ch. Devi Lal University, Sirsa, Haryana, India. He has obtained his MBA from Maastricht School of Management, Maasticht, Netherlands, M.S from BITS, Palani, India and Masters in Physics from Himachal Pradesh University, Shimla, India. His areas of interest are ICT in Educational Institutes, computer networking (wired and wireless), Sensor Networks, and Open source web content management. He played a very important role in the Design and Implementation of Campus Wide Optical Fibre Network backbone of Himachal Pradesh University, Shimla. He is having more than 10 years of teaching experience in addition to his 5 years in IT Industry.

Dr. Vikram Singh is Ph.D in Computer Science from Kurukshetra University, Kurukshetra, India. Presently he is working as Professor and Head in the Department of Computer Science & Engg and Dean Faulty of Engineering, Ch. Devi Lal University, Sirsa – 125055.Haryana, (India), since 2004 onwards. Before this he was working with Kurukshetra University, Kurukshetra. His areas of research are Computer Networks, E-Governance, Simulation tools, etc. He is having more than 17 years of teaching/research experience and has written two books and having more than 30 publications in international and national journals/conference proceedings.

IJCSI CALL FOR PAPERS MAY 2010 ISSUE

Volume 7, Issue 2

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.
See authors guide for manuscript preparation and submission guidelines.

Accepted papers will be published online and authors will be provided with printed copies and indexed by Google Scholar, Cornell's University Library, ScientificCommons, CiteSeerX, Bielefeld Academic Search Engine (BASE), SCIRUS and more.

Deadline: 31st March 2010
Notification: 30th April 2010
Revision: 15th May 2010
Online Publication: 31st May 2010

- Evolutionary computation
- Industrial systems
- Evolutionary computation
- Autonomic and autonomous systems
- Bio-technologies
- Knowledge data systems
- Mobile and distance education
- Intelligent techniques, logics, and systems
- Knowledge processing
- Information technologies
- Internet and web technologies
- Digital information processing
- Cognitive science and knowledge agent-based systems
- Mobility and multimedia systems
- Systems performance
- Networking and telecommunications
- Software development and deployment
- Knowledge virtualization
- Systems and networks on the chip
- Context-aware systems
- Networking technologies
- Security in network, systems, and applications
- Knowledge for global defense
- Information Systems [IS]
- IPv6 Today - Technology and deployment
- Modeling
- Optimization
- Complexity
- Natural Language Processing
- Speech Synthesis
- Data Mining

All submitted papers will be judged based on their quality by the technical committee and reviewers. Papers that describe research and experimentation are encouraged.
All paper submissions will be handled electronically and detailed instructions on submission procedure are available on IJCSI website **(http://www.IJCSI.org)**.
For other information, please contact IJCSI Managing Editor at **editor@ijcsi.org**
Website: **http://www.IJCSI.org**

IJCSI

The International Journal of Computer Science Issues (IJCSI) is a refereed journal for scientific papers dealing with any area of computer science research. The purpose of establishing the scientific journal is the assistance in development of science, fast operative publication and storage of materials and results of scientific researches and representation of the scientific conception of the society.

It also provides a venue for researchers, students and professionals to submit on-going research and developments in these areas. Authors are encouraged to contribute to the journal by submitting articles that illustrate new research results, projects, surveying works and industrial experiences that describe significant advances in field of computer science.

Indexing of IJCSI:

1. Google Scholar
2. Bielefeld Academic Search Engine (BASE)
3. CiteSeerX
4. SCIRUS
5. Docstoc
6. Scribd
7. Cornell's University Library
8. SciRate
9. ScientificCommons

www.IJCSI.org