

RESEARCH GRANTS COUNCIL**Application for Allocation from
the General Research Fund for 2010 / 2011****Application Form (GRF1)**

[Explanatory Notes (GRF2) give detailed instructions on how to complete the application form, including responsibilities of PIs/Co-Is/institutions and implications of the information required in the form. Please read the GRF2 when filling in the application]

PART I SUMMARY OF THE RESEARCH PROPOSAL

[To be completed by the applicant(s)]

1. Particulars of the Project:**Name and Academic Affiliation of Principal Investigator:**

<u>Name</u>	<u>Post</u>	<u>Unit/ Department/ Institution</u>
Dr Liu, James Nga-kwok	Associate Professor	Department of Computing/The Hong Kong Polytechnic University

(a) ii) **Title of Project:** Automatic extraction of fuzzy domain ontology for content management and information retrieval in e-publication systems

(b) i) **Primary Field:** Others - Computing Science & Information Technology (e-Commenrce, e-Logistics and Web Technology) & Code 2299

Secondary Field: Artificial Intelligence & Code 2201

(b) ii) A maximum of five keywords to characterise the work of your proposal
(maximum 30 characters for each keyword)

1) Web ontology 2) Fuzzy concept extraction 3) Content management 4) e-publication

Allocation Requested from the General Research Grant:

[The amount shown here should be the same as shown in Section 3(a)(xii) of Part II below]

Project Duration:	36 Months
Total cost of the project:	750,600
Less :	
(a) Institutional funding for provision of research experience for undergraduate student	0
(b) Other research funds secured from other sources	0
Amount requested *:	HK\$ 750,600
Plus :	
On-costs (15% of the amount requested) :	HK\$ 112,590
Grand total amount requested :	HK\$ 863,190

(c). i. Nature of Application *

- New [i.e. PI and/or Co-I(s) applying for RGC funds on this research topic for the first time]. Please give further details in Part II item 2.
- Re-submission [i.e. PI and/or Co-I(s) have previously applied for RGC funds on this research topic but application not supported]. Please give further details in Part II Section 6.
- On-going [i.e. PI and/or Co-I(s) extending work previously funded by the RGC]. Please give further details in Part II Sections 9-10.

ii. Special funding template (Applicants can select more than one box)

- Application for Clinical Research Fellowship Scheme (Please also complete an additional form (Enclosure I) and see (Enclosure II)) (only available for applications under B & M Panel)
- Application for individual research (see Enclosure III) (only available for applications under H Panel)
- Application for longer-term research grant (see Enclosure IV)
- Application for employment of relief teacher under Humanities Sub-Panel (see Enclosure V)(only available for applications under H Panel with the primary field in the four specified disciplines)
- Application for provision of research experience for undergraduate student (see Enclosure VI)

(d). Abstract of Research comprehensible to a non-specialist (a maximum of one A-4 page in standard RGC Format for attaching PDF documents or a maximum 400 words for direct input in the text box):

The popularity of e-publication is growing rapidly for the past few years, which potentially contains a huge amount of knowledge. It is to be anticipated that the e-publication systems can provide useful knowledge and more appropriate and correct results for information search to their clients. However, in contrast to the exponential speed of the growth of data and information, the data and information processing capabilities do not keep up the pace. One major concern is the missing of an effective automated solution to transform data into meaningful information or knowledge. Although ontology, which enables advanced functionality in knowledge systems and forms the knowledge base for future innovations, can be used for developing content management system and intelligent search engine for e-publication, the engineering of the knowledge construct using relation tagging of concepts is very time consuming and expensive involving a large amount of human resources. It appears that human tagging is not a feasible way to build up the solution space, and it becomes the bottleneck of knowledge acquisition. We are basically dealing with this bottleneck, which is also a classical knowledge engineering problem.

Overall, the automatic extraction of ontology is quite challenging due to the unstructured nature of information contents and inherent semantic ambiguities in natural language. To serve our specific domain of interest, the ontology should capture domain dependent information. A more scalable, systematic and automatic approach to ontology construction for fuzzy domain is to be developed. For this purpose, several essential problems are to be addressed, including concept learning, concept extraction, semantic formation, relation extraction, fuzzy representation, knowledge discovery, etc. In this project, we will mainly focus on: 1) exploiting ontology paradigm in support of information extraction and retrieval, and 2) investigating the formulation of concept learning from a selected domain. The application domain of this project is targeted on e-publication, due to its popularity and significance to sharing knowledge in our real life.

Furthermore, we will also pay efforts to the issue of correctness and verification of the ontologies extracted automatically, by exploiting the integration of the ontology extractor with the Fuzzy Ontology Coloured Map (FOCM) to be extended from our previous approved CERG project on the problems of content management and recommendation. This will help the construction of correct ontology for content management and information retrieval in the application systems.

2. Particulars of PI and Co-Is

(a). Investigator(s) information:

Name(s) and Academic Affiliation(s) of Applicant(s):

	Name	Post	Unit/ Department/ Institution
Principal Investigator (PI): (with title)	Dr Liu, James Nga-kwok	Associate Professor	Department of Computing/The Hong Kong Polytechnic University
Co-Investigator(s) [Co-I(s)]:(with title)	Dr Lee, Raymond	President	IATOPIA Technology Ltd
	Dr You, Jane J.	Associate Professor	Dept of Computing/The Hong Kong Polytechnic University

2(b). Is the PI a new appointee within two years of full-time paid appointment to his/her first substantive academic/research position in a university at the time of submission of the proposal?

Yes No

2(c). Curriculum vitae (CV) of applicant(s).

[For the PI and each Co-I, please attach one **A-4 page** CV (with a maximum of 400 words) per person in the following format.]

i) Name:ii) Academic qualifications:iii) Previous academic positions held(with dates):iv) Present academic position:v) Previous relevant research work:vi) Publication records :Section A - Five most representative publications in recent five yearsSection B - Five representative publications beyond the recent five-year period with the latest publication entered first.vii) Others (please specify):

i) **Name:** Dr James N. K. Liu

ii) **Academic qualifications:**

- 1991 PhD, La Trobe University, Australia
- 1988 MPhil, Murdoch University, Australia
- 1982 BSc (Hons), Murdoch University, Australia

iii) **Previous academic positions held:**

- 2000-Present Associate Professor, Department of Computing, Hong Kong Polytechnic University
- 1994-2000 Assistant Professor, Department of Computing, Hong Kong Polytechnic University
- 1988-1989 Lecturer, Department of Computer Science, La Trobe University, Australia

iv) **Present academic position:** Associate Professor

v) **Previous relevant research work:**

James Liu's PhD research was in expert system verification and validation. He has created a formal description technique based on Petri Nets theory to model expert systems. The methodology is used for verifying the correctness, consistency, and completeness of production based systems. It can also address knowledge with uncertainty. The model has subsequently been extended to cater for frame based systems. Over the years, James has been doing applied research in relevant areas including intelligent systems, news extraction, machine learning techniques, data mining, agent technology and Web ontology.

vi) **Publication records:**

Section A

1. Toby H.W. Lam, James N.K. Liu and Raymond S.T. Lee (2009) MASTER: An Intelligent Ontology-based Multi-Agent System For Sightseer, *International Journal of Software Engineering and Knowledge Engineering*, Vol. 19, No. 2, pp. 137-157.
2. Edward Lim, Raymond S.T. Lee and James N.K. Liu (2008) Knowledgeseeker – an Ontological Agent-based System for Retrieving and Analyzing Chinese Web Articles, in *IEEE Proceedings of the International Conference on Fuzzy Systems (FUZZ2008) as part of 2008 IEEE World Congress on Computational Intelligence (WCCI2008)*, Hong Kong, June 1-6, 2008
3. James N.K. Liu and Raymond W.M. Kwong (2007), Automatic Extraction and Identification of Chart Patterns towards Financial Forecast, *Special Issue Soft Computing for Time Series Prediction*, 7(4), pp. 1197-1208 Elsevier,
4. James N.K. Liu, Meng Wang and Bo Feng (2005) iBotGuard: An Internet-based Intelligent Robot Security System Using Invariant Face Recognition Against Intruder, *IEEE Trans. On SMC - Part C*, 35(1), pp. 97-105.
5. Liu, James N.K. and You, Jane (2003), "Smart Shopper: An Agent Based Web-mining Approach to Internet Shopping ", *Special Issue on Knowledge Discovery and Data Mining by IEEE Transactions on Fuzzy Systems*, 11(2), pp. 226-237.

Section B

1. Raymond S.T. Lee and James N.K. Liu (2001), "NORN Predictor – Stock prediction using a neural oscillatory-based recurrent network", *International Journal on Computational Intelligence and Applications (IJCIA)*, 1(4), pp. 439-452.
2. Liu, J.N.K., Li, B.N.L. and Dillon, T.S. (2001), "An improved Naïve Bayesian Classifier technique coupled with a novel input solution method", *IEEE Trans. on SMC, Part C*, **31**(2), pp. 249-254.
3. Lee, R.S.T. and Liu, J.N.K. (2000), "Tropical Cyclone Identification and Tracking System using integrated Neural Oscillatory Elastic Graph Matching and hybrid RBF Network Track mining techniques", *IEEE Transaction on Neural Network*, **11**(3), pp. 680-689.
4. Liu, J.N.K. and Sin, K.Y. (1997), "Fuzzy neural networks for machine maintenance in Mass Transit Railway System", *IEEE Transactions on Neural Networks*, **8**(4), pp. 932-941.
5. Shiu, S.C.K., Liu, J.N.K. and Yeung, D.S. (1997), "Formal description and verification of hybrid rule/frame based expert systems", *Expert Systems with Applications*, **13**(3), pp. 215-230.

(vii) **Others**

James was a research scientist at the Aeronautical Research Laboratory, Defence Science & Technology Organizations (DSTO), Australia during 1990-94. His work involved several defence projects investigating the techniques and application of artificial intelligence for aircraft operations. Prior to that, James was a computer scientist at the Defence Signal Directorate, Australia during 1988-90. He was involved in defence projects on the techniques and application of decryption algorithms for signal processing.

i) **Name:** Dr Raymond Lee

ii) **Academic qualifications:**

- PhD from Hong Kong Polytechnic University in 2000
- MSc from Hong Kong Polytechnic University in 1997
- BSc from Hong Kong University in 1989

iii) **Previous academic positions held:**

2005-2006 Oct Associate Professor, Department of Computing, Hong Kong Polytechnic University
2000-2005 Assistant Professor, Department of Computing, Hong Kong Polytechnic University
1998-2000 Lecturer, Department of Computing, Hong Kong Polytechnic University

iv) **Present academic position:** President of IATOPIA Technology Ltd

v) **Previous relevant research work:**

Raymond Lee has published over 85 publications and the author of FOUR books/research monographs in the areas of E-commerce, Pattern Recognition, Intelligent Agent Technology, Chaotic Neural Networks and Fuzzy Ontology systems.

vi) **Publication records:**

Section A

1. Lam, Toby H.W., Lee, **Raymond S.T.** and Liu, James N.K. (2008), An Ontology-based Intelligent Mobile System for Tourist Guidance., Computational Intelligence: A Compendium, Springer-Verlag, pp. 381-406.
2. **Lee, Raymond S. T.** and Loia, Vincenzo (Eds), (2007), Computation Intelligence for Agent-based Systems, Springer-Verlag, Heidelberg,
3. **Lee, Raymond S. T.** (2006), Fuzzy-Neuro Approach to Agent Applications (From the AI Perspective to Modern Ontology), Springer-Verlag, Heidelberg, ISBN: 3-540-21203-5.
4. Janice W.Y. Hui, Toby H.W. Lam and **Raymond S.T. Lee** (2005), "The Design and Implementation of an Intelligent Negotiation Shopping System," An International Journal of Multiagent and Grid Systems 1 131-146.
5. **Lee, Raymond S. T.** and Liu, James N. K. (2003), Invariant Object Recognition Based on Elastic Graph Matching: Theory and Applications, IOS Press, ISBN: 1-58603-296-8.

Section B

1. Lee, R.S.T. and Liu, J.N.K. (2002), "SCENOGRAM - SCENE analysis using composite Neural Oscillatory-based elastic GRAPH Model", International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 16(2), pp. 215-237.
2. Raymond S.T. Lee and James N.K. Liu (2001), "NORN Predictor – Stock prediction using a neural oscillatory-based recurrent network", *International Journal on Computational Intelligence and Applications (IJCIA)*, 1(4), pp. 439-452.
3. Lee, R. S. T. and Liu, James N. K. (2001), "An Elastic Contour Matching Model for Tropical Cyclone Pattern Recognition", *IEEE Trans. Systems, Man, and Cybernetics*, Volume 31: Part B, No. 3, pp. 413-417.
4. Lee, R.S.T. and Liu, J.N.K. (2000), "Tropical Cyclone Identification and Tracking System using integrated Neural Oscillatory Elastic Graph Matching and hybrid RBF Network Track mining techniques", *IEEE Transaction on Neural Network*, 11(3), pp. 680-689.
5. Lee, R.S.T. and Liu, J.N.K. (1999), "An Automatic Satellite Interpretation of Tropical Cyclone Patterns Using Elastic Graph Dynamic Link Model", *International Journal of Pattern Recognition and Artificial Intelligence*, 13(8), pp. 1251-1270.

(vii) **Others**

Although Raymond left PolyU in late 2006, he has been maintaining his research interest in a number of areas and developing applications in electronic business and mobile commerce.

BRIEF CURRICULUM VITAE

Name: Jia YOU

Academic qualifications: **Ph.D.**, La Trobe University, Australia, 1992
BEng., Xi'an Jiaotong University, P.R. China, 1986

Previous academic positions: Senior Lecturer, Griffith University, Australia (01/1996 – 02/2002)
Research Fellow, Universite Paris XI (12/1993 – 11/1994)
Lecturer, The University of South Australia, Australia (04/1993 – 12/1995)

Present Academic Position: Associate Professor, The Hong Kong Polytechnic University

Previous relevant research work:

Image processing, medical imaging, pattern recognition, retinal image analysis, content-based image retrieval, biometrics computing, multimedia systems, data mining and data warehousing.

Publication records:

A. Five most representative publications in recent five years

1. M. Niemeijer, B. Zhang, X.Q. Wu, J. You, Q. Li, F. Karray and M.D. Abramoff, *et. al.*, "Retinopathy On-line Challenge: Automatic detection of microaneurysms in digital color fundus photographs", accepted for publication by *IEEE Trans. on Medical Imaging*, 2009.
2. B. Zhang, X.Q. Wu, J. You, Q. Li and F. Karray, "Detection of microaneurysms using multi-scale correlation coefficients", accepted for publication by *Pattern Recognition*, 2009.
3. L. Zhang, Q. Li, J. You and D. Zhang, "Modified matched filter with double-side thresholding and its application to proliferative diabetic retinopathy screening", *IEEE Trans. on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 528-534, 2009.
4. L. Liu, D. Zhang and J. You, "Detecting wide lines using isotropic nonlinear filtering", *IEEE Trans. Image Processing*, vol. 16, no. 6, pp. 1584-1595, 2007.
5. J. You, W.K. Kong, D. Zhang and K.H. Cheung, "On hierarchical palmprint coding with multi-features for personal identification in large databases", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 234-243, 2004.

B. Five most representative publications beyond the recent five-year period

6. W.K. Kong, K.H. Cheung, D. Zhang, M. Kamel and J. You, "An analysis of biohashing and its variants", *Pattern Recognition*, vol. 39, pp. 1359-1368, 2006.
7. W.X. Li, J. You and D. Zhang, "Texture-based palmprint retrieval using a layered search scheme for personal identification", *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 891-898, 2005.
8. D. Zhang, W.K. Kong, J. You and M. Wong, "On-line palmprint identification", *IEEE Trans. PAMI*, vol. 25, no. 9, pp. 1041-1050, 2003.
9. J. You, W.X. Li and D. Zhang, "Hierarchical palmprint identification via multiple feature extraction", *Pattern Recognition*, vol. 35, no. 4, pp. 847-859, 2002.
10. J. You and P. Bhattacharya, "A wavelet-based coarse-to-fine image matching scheme in a parallel virtual machine environment", *IEEE Trans. Image Processing*, vol. 9, pp. 1547-1559, 2000.

Awards:

- **The Second Place** of SPIE Medical Imaging Int. Competition Retinopathy Online Challenge (2009)
- The Hong Kong Polytechnic University Outstanding Professional Services and Innovation Award (2004)
- **Hong Kong Industry Award**, Palmprint Identification System, the Certificate of Merit under the Consumer Design Category, Hong Kong Government (2003)

Patent:

- D. Zhang, J. You et al., "Method of Palmprint Identification", US Patent No. US7,496,214 B2, Date of patent: Feb. 24, 2009.

Book:

- J. You and D. Zhang, "Advanced Multimedia: Algorithms, Systems and Applications", to be published by John Wiley & Sons, Inc. in 2009 (the manuscript has been delivered to the publisher)

Professional service:

- member of the editorial board for *International Journal of Image Graphics* since 2001, program chair of 2008 Int. Conf. Medical Biometrics, and PC members of many international conferences.

PART II DETAILS OF THE RESEARCH PROPOSAL

[To be completed by the applicant(s)]

RESEARCH DETAILS

1. Project Objectives

(Maximum 800 words in total for the project objectives and long-term impact.)

(a) Long-term impact:

[Identify the key issues and problems being addressed, and state the possible outcome of the research project in terms of its relevance, significance and value]

Nowadays, with the popularity of e-publication, such as Google books, it becomes an increasingly important approach and channel to obtain information and knowledge in our real life. In the current World Wide Web, people usually find information from e-publications by using search engines, which mainly focus on text-based retrieval and usually based on keyword matching, such as Google, Yahoo or Baidu. However, sometimes it is not easy for us to obtain information what we really want. For example, if a fan wishes to know what movies associated with some famous actress at present, who were the directors of these movies? Any interesting news or comments reported in the media about this actress? Nowadays, queries such as these are quite difficult to address and is still not possible to expect a computer to do the whole process for us automatically. One reason for this is that web pages are designed to display information to users, instead of computers. There is a missing of an effective automated solution to transform data into meaningful information or knowledge. This limits the computer to analyzing the content of the web and associated concepts.

Although the semantic web becomes popular, such as Google YouTube, the engineering of the knowledge construct using relation tagging of concepts is very time consuming and expensive involving a large amount of human resources. For example, the IMDB movie management site, and Wikipedia website, they both want to discover the knowledge and allow their clients to search for the information across the knowledge base. However, to achieve this goal, they must face a big challenge which is how to tag and link up those concepts. It appears that human tagging is not a feasible way to build up the solution space, and it becomes the bottleneck of knowledge acquisition. We are dealing with this bottleneck, which is also a classical knowledge engineering problem.

Domain ontology enables advanced functionality in knowledge management systems and forms the specific knowledge base for future innovations. For examples, it can be used for developing the intelligent search engine, content management system like Wikipedia etc. Those reasons let us have the utmost incentive to develop the autonomic solution for extracting the knowledge. To achieve such goal, we anticipate develop an automatic ontology extractor for fuzzy knowledge domain, which can be further applied to the content management and information retrieval. The integration of fuzzy theory makes it more powerful to deal with imprecise or vague information, whereas much of our real-life

knowledge is involved with imprecision and uncertainty.

Furthermore, the correctness of the extracted fuzzy domain ontology is critical for the effective application. Thus, we also pay efforts to the issue of correctness and verification of ontology to ensure an effective and reliable application system based on the extracted ontology. The Fuzzy Ontology Coloured Map (FOCM), to be extended from our previous approved CERG project on the problems of content management and recommendation, will be integrated into the automatic fuzzy ontology extractor. It enhances the capability of constructing correct ontology for content management and information retrieval applications.

The application background of this project is focused on e-publication systems. The research results to be obtained concerning the problems discussed above, on the aspect of theory, will be applied to and evaluated in the e-publication systems. I.e., a concrete e-publication system with the capability of content management and information retrieval, based on the automatic extracted fuzzy domain ontology using the proposed approach, will be developed. It is to be anticipated with improved performance on content management and information retrieval. Thus, besides the contribution on theoretic problems in ontology engineering, this project is also of important significance in practice.

[Please list the objectives in point form below]

(b) Objectives

1. Investigation of automatic or semi-automatic ontology engineering techniques. It includes data collection and preprocessing, concept learning and extraction, relation extraction and evaluation, structure formation and representation, and knowledge discovery and inference.
2. Collection and reuse of selected domains of Web contents. The extracted fuzzy ontological knowledge for selected domains of Web contents by our proposed automatic extractor is anticipated to be shared and reused in application.
3. Design of concept learning mechanism. The mechanism of concept learning is to be designed for the automatic fuzzy domain ontology extraction.
4. Formulation of algorithms. Algorithms for automatic fuzzy domain ontology extraction are to be formulated in detail, including concept extraction, semantic formation, fuzzy representation, and knowledge discovery.
5. Exploiting the potential integration with FOCM. With the integration of FOCM, the extracted ontology is anticipated to be verified using a high level net approach.
6. System development for content management and information retrieval. Application systems on e-publication for content management and information retrieval are to be developed based on the extracted fuzzy domain ontologies.

(c) A maximum of two non-text pages of attached diagrams, charts, photos and others etc, if any.

Attached 2 pages(s) as follows

2. Background of Research

(A) Work done by others:

In the current World Wide Web, web pages are designed to display information to users, instead of computers. This limits the computer to analyzing the content of the web and associated concepts. Thus, Berners-Lee and other researchers proposed a new form of web, Semantic Web (SW), which would make the web content accessible and understandable to both machines and people. We note that the core element of SW is ontology which can facilitate capture and construction of domain knowledge and enable representation of skeletal knowledge to facilitate integration of knowledge bases irrespective of the heterogeneity of knowledge source (e.g. [2], [76]).

In general, ontology can be a formal representation of concepts and their interrelationships [18]. It can take the simple form of taxonomy (i.e. knowledge encoded in some hierarchical structure) or a vocabulary with standardized machine interpretable terminology supplemented with natural language definitions. It can be used to describe a logical domain theory with very expressive, complex, and meaningful information. As regarding domain ontology, it specifies the knowledge for a particular type of domain [12]. This kind of ontologies generalize over application tasks in domains such as medical, tourism, banking, finance, entertainment, etc. During the development, the ontology is often specified in a declarative form by using semantic markup languages such as Resource Description Framework (RDF) [70] and Ontology Web Language (OWL) [65]. These markup languages are able to model the web content in a machine-readable way which assists the processes for information gathering and searching in an automatic way.

Lately, [57] have extended the idea of ontology-based knowledge representation to include fuzzy measure for a set of inter-concept relations defined in ontology. The measure of these relations are used to judge the context of a set of entities, the context of a user and the context of the query for the purpose of intelligent information retrieval. A fixed set of commonly encountered semantic relations have been identified and their combinations are used to generate fuzzy relations. [68] proposed the FOGA framework for fuzzy ontology generation. The framework consists of fuzzy formal concept analysis, fuzzy conceptual clustering, fuzzy ontology generation, and semantic representation conversion. Nevertheless, the quality of clustering is dependent on assignment of meaningful labels to initial class names, attributes and relations. This is done manually and requires domain expertise. The system is also not designed to extract fuzzy relational concepts from unstructured or semi-structured text documents. Similarly, the fuzzy domain ontologies about various events covered by some Internet news, which were utilized for news summarization purpose in [33] and [34], were manually developed by human domain experts. Recently, a particular interest in non-taxonomic fuzzy relationship called correlation has been shown ([4], [5]), which is a binary and symmetric fuzzy relationship that allows one to specify the semantic link among concepts of the fuzzy ontology. In their developed systems, the correlations, first assigned by experts, are updated after querying or when a document has been inserted into a database.

As such, ontologies are useful in many areas but then the engineering of ontologies is very

expensive and time consuming. This is despite the fact that several tools including Ontolingua [13], OilEd [3], [64], Protégé [67], and OntoEdit [78] are developed for the construction and management of ontologies. These tools are being applied by domain expert to the construction of ontologies. We do need to develop better automatic or semi-automatic ontology engineering techniques for extracting fuzzy domain ontology. The development is vital as it deals with the knowledge acquisition bottleneck which is a classical knowledge engineering problem. Overall, the automatic extraction of domain specific ontologies is quite challenging due to the unstructured nature of information contents and inherent semantic ambiguities in natural language.

In literature, different automatic extraction techniques have been proposed and implemented in several contexts. Some are based on the statistical analysis of terms in documents ([8], [9], [14]) while others are based on syntactical features. There are also studies exploiting both linguistic and statistical features to extract phrases (e.g. [16], [31], [32]). The former features are used to filter word sequences (noun and adjective phrases). The latter exploits the occurrences and emphasis terms that may appear as nested within longer terms. They also extract ontology from various data types such as textual data ([59], [63]), dictionary ([8], [62]) and relational schemata ([72], [77]). Furthermore, [82] implemented a genetic algorithm to find the optimal set of key-phrases that display the lowest dispersions level. [10] presented an automatic taxonomy learning algorithm to extract concept hierarchies from a text corpus. It is based on formal concept analysis, a kind of conceptual clustering technique to generate concept lattice. [14] used a Markov model approach to extract noun phrases. Candidates could be matched to a dictionary to train the Markov Model. These candidate terms are words or list of n words that can be automatically extracted from a corpus of resources (documents, service descriptions, etc) as part of the concept hierarchy or ontology depending on their calculated importance.

Moreover, the study on automatic ontology matching, which is closely relevant to the topic of ontology extraction, will also give some inspiration to the design of mechanism and algorithms for automatic ontology extraction in this project. In recent years, many automatic ontology matching tools have been developed to automate the matching process, such as OntoDNA[25], ASMOV [24], and Falcon-AO [23]. These tools share many common features, whereas the major difference lies in the algorithms used to perform all these matching phases [27]. We will make reference to these different mechanisms and algorithms, and some of them have the potential to be modified and employed for ontology extraction in this project.

(B) Work done by PI and Co-Investigators:

The PI has been working in the field of artificial intelligence since 90's. He builds intelligent systems and extends AI technologies to solve time-series forecasting problems for many applications ([15], [35], [38], [44], [45], [51], [52], [55], [56], [57]). Recently, the PI and Co-Is have been involved in the development of an intelligent agent-based platform, iJADE¹ from which a number of systems have already been developed including an authentication system, a surveillance system, a web mining

¹ <http://www.ijadk.org>

system, a stock advisor system, a weather forecasting system, and an ontology-based free walker which is a location-aware application utilizing the Global Positioning System to gather geographic information ([28], [29], [36], [37], [38], [40]). In addition, the research team has proposed a fuzzy extension for the current hard-constraint markup language – Fuzzy Ontology Map (FOM). Subsequently, several applications have been developed including the construction of FOM profile for personalization, information filtering, product management and recommendation, tourist guidance (e.g. [30], [46], [87]). These systems are often integrated with different intelligent agents to enhance the functionality of the application development.

Besides that, the PI's previous work to support multi-perspective visualizations and morphing capability according to user interactions, feature extraction and selection of a target object, object segmentation from background, object encoding, searching and identification solution for electronic commerce, expert system modeling with uncertainty and related verification and reliability problems ([41], [42], [43], [47], [48], [49], [50], [75]), coupled with the research on text analysis, multi-lingual news digest and retrieval ([6], [19], [39], [53], [54], [58], [83], [85], [86]) provide a solid foundation to the proposed system development for automatic extraction of domain ontology and are specifically in the best interest of electronic commerce and the Web community. Subsequent research work has been carried out in the area of ontology engineering since last year as presented in the additional references [87-94]. The previously approved CERG grant, which started in January 2009 can on one hand, allow the research team to exploit the use of FOM to model fuzzy concepts, and on the other hand, the use of high level Petri Nets which is a kind of formal description technique, to check for the correctness and consistency of the ontology system. One of the Co-Is is actively involved in the commercialization of content management. The present proposal for the formulation of a domain ontology extractor will facilitate the reuse of technology in ontology construction of specific domain of interest and supplementing to the steps towards the successful development of the Web system.

Together, these individual investigators form a solid team qualified to take on different aspects of the research required by the proposed project.

3. Research Plan and Methodology

We need to look into the means of preserving the basic structural knowledge format for storing domain knowledge, but allowing for update of information at the same time. We borrow the idea of information extraction (IE) based summarizations which is more capable of extracting and merging information from various resources, but they are often domain dependent due to the use of linguistic rules designed to extract information of specific content. Adaptive IE system (e.g. [11]) can be employed to ease the problem by identifying new extraction rules induced from example annotations supplied by users. We'll examine various means to make the training process more effective.

IE that incorporates with ontologies is expected to be able to support information integration [71] and increase domain portability [60]. The domain ontology should capture domain dependent information including those important attributes and actions, relationship of a concept to its associated sub-concepts, etc. Clustering methods can be used on test corpuses to aid users construct primitive

ontologies to represent the main topics in the corpus. Templates could then be generated from the ontology and guide the IE process. Ontologies produced by this approach though limited to the content of the corpus, are feasible when we confine our focus of interest per specific application (e.g. movie domain). We'll explore tools that can help develop ontologies semi automatically with user driven annotations (e.g. [80]). We consider generate hierarchically arranged terms from the chosen domain and when presented in an ontology interchange language they can be inspected and refined in an ontological engineering tool like Protégé, using a semantic schema such as OWL . Note that Protégé is a free, open source ontology editor and knowledge-base framework. The advantages for linking with such tool are to increase the visibility and editing ability for the ontology.

The process of ontology extraction and the representation of ontology will be integrated with fuzzy logic to enhance its capability of knowledge discovery and description from imprecise or vague information. For example, fuzzy concept clustering techniques are to be employed for concept extraction, instead of preassigning of meaningful labels. The membership values of the relations between concepts can be a function of frequency of co-occurrence of concepts. We believe that the frequency of use of words correlates with acceptability of those words as part of the vocabulary. The mathematical baseline for these fuzzy relations is detailed in [84]. A fuzzy ontology could then be created by using the RDF/OWL document with the fuzzy concept clustering. It supports fuzzy searching and leads to the development of several applications as a result (e.g. [28], [79], [81]).

Our approach identifies and extracts terms and ontologies from a set of domain documents. Moreover, a more scalable, systematic and automatic approach to ontology construction is to be developed. Figure 1 gives a high level view of building domain ontology with a traditional system verse an alternative system involving the development of an automatic ontology concepts extractor. We will use knowledge domain supplied by the City Entertainment's Film Bi-Weekly Magazine via our project co-investigator of IATOPIA Research Lab as test bed for our project development. With those challenges in mind, we set our main focus of the project on the following:

- 1) Exploiting ontology paradigm in support of information extraction and retrieval. This is to facilitate knowledge access, sharing, and reuse information.
- 2) Investigating the formulation of concept learning from a selected domain. This is to facilitate concept extraction, semantic formation, fuzzy representation, knowledge discovery.
- 3) Exploiting the potential integration with FOCM. This is to allow for knowledge inference and system checking.

Figure 1. High Level View of Domain Ontology Construction

Methodology & Project Framework:

The project has the following stages:

- Preparation Stage (6 months): System planning and domain data preparation
- Analysis Stage (15 months): Concept learning, extraction, modelling and analysis

- Coding Stage (10 months): System integration and program coding
- Evaluation Stage (10 months): Experimental testing and comparison

Preparation Stage

Figure 2 illustrates the framework of the proposed e-publication system for content management and information retrieval, which consists of four main components: 1) provide source of knowledge, 2) concept learning, 3) create ontology model, and 4) information retrieval.

Figure 2. System Framework Overview

The system first requires user to provide the source of knowledge, this knowledge is not well structured, i.e. in a raw format that is not machine readable and understandable, such as text document written in natural language or web logs from user representing the users' preference. This source of knowledge in raw format is treated as the most prior input to the domain ontology extraction process. The second component of the framework is concept learning. This is the main and core part of the framework. This project will mainly focus on this component, and it will be discussed in detail at the analysis stage. The third component is the representation of ontology model, it defines how the domain knowledge is represented in machines, and also is easily understood by human. Finally, the information retrieval component defines the information retrieval methods, which involve two main processes respectively the concept mapping and the knowledge querying. The concept mapping defines how different concept matches and calculated, and the knowledge querying defines the method of retrieving knowledge. All of which are defined for building ontology-based application such as concept-based search engines and concept-based document classification and categorization.

In this project, we will use the City Entertainment's Film Bi-Weekly Magazine as content data to develop, test and evaluate the proposed approaches and system. It will provide some structural information of the movie domain content for start. However, we do need to pre-process these data for our project, including coding, markup, tagging, etc.

Analysis Stage

Figure 3 shows how knowledge units cluster and relate within a concept graph. Concepts that are mutually dependent or common to a particular domain may be hierarchically grouped into a concept cluster. Links between knowledge units (term, concept, concept cluster, concept graph) are concept relations. The largest knowledge unit is a concept graph, and this represents the sum-total of the units and relations extracted from a document set for a particular domain or topic. We will focus on the upper-level ontology (e.g. [73]) which is an ontology that can provide a set of generic concepts which can be shared and reused by different users. It describes the general concepts across different domains. While learning and extracting these knowledge units and their relations, considerations shall include:

- Concept interdependence when calculating different levels of importance

- Content grouping by interest with a certain level of similarity
- Grouping similar contents into fuzzy clusters (e.g. [21], [74])

Figure 3. Term, Concept, Concept Cluster, and Concept Graph

With these considerations, we will make references including the semantic indexing approach by [20] to associate concepts to texts and weigh these associations according to the occurrences of concept labels in texts; the Growing Self-Organization Map (GSOM) by [7] to develop a hybrid model for mining the hidden semantics from text at different levels of abstract to alleviate the shortcomings of the self-organizing map SOM approach by [12]. We consider supplement background linguistic knowledge derived from corpuses (e.g. WordNet [61], EuroNet, HowNet [17], [22]) that are associated with the segmentation process to improve the clustering performance. In addition, we'll explore domain dependent heuristics (e.g. [1], [69], [71]) for comparing and merging extracted information from multiple documents.

As shown in Figure 3, there are different levels of knowledge unit that are to be extracted. For this purpose, a bottom-up learning approach is proposed for concept learning, as shown in Figure 4. All of these sub-processes correspond to identifying different levels of knowledge unit, thus the knowledge is learnt from the smallest unit towards the largest unit, and finally the domain ontology is generated.

Figure 4. Concept Learning Process

The first step extracts a list of candidate terms, which are used for concept learning process. However, this candidate term list has no meaning and relationship to any conceptual knowledge unit in the model. Thus the next process applied to the candidate term list is the term-to-concept relationship mapping. It selects and separates every term in the candidate term list to its related domain concepts. Taxonomical relationship is also classified at the step of term-to-concept relation mapping. An illustrative example of term-to-concept mapping with fuzzy relationship is shown in Figure 5.

Figure 5. Term-to-Concept Mapping with Fuzzy Relationship

After the term-to-concept mapping, a list of candidate concepts have been created. Then the concept selection process is applied to the candidate list, and further, the relation between the selected concepts is to be learned. In the concept relation learning, considerations may include: frequency of co-occurrence of the concepts, similarity measure based on the terms contained in the concepts (e.g. mutual information [66] and expected cross entropy [26]), χ^2 statistical independence test, etc. Then, semantically similar concepts are to be grouped into a tight semantic group, based on the concept

relation learning. The idea of concept clustering is to group concepts with high weighted relations into a sub graph while separating out other concepts to create a new sub graph of low weighted relations. Clusters are automatically created without explicitly requiring the number of clusters to be defined in the result. The taxonomical relationships between concepts are also created in the process of concept clustering. An illustrative example is shown in Figure 6. With the results of concept learning, containing all the processes described above, a fuzzy domain ontology is finally built via transformation from the clustered concepts and their relations.

Figure 6. Concept Clustering and Taxonomical Relationship

In addition, after the ontology has been extracted, its correctness is also to be assessed by integrating with the FOCM, which is based on the high level Petri nets approaches. We'll make use of our previous research experience of using State Controlled Petri Nets formalism for verifying hybrid rule / frame-based expert systems (e.g. [42], [43], [50], [75]) to examine the ontology construct that can be integrated with concept hierarchy, property inheritance and possible embedded heuristics in the association.

Coding Stage

This stage is implementation of the model, it includes:

- ✧ Program coding and the integration of different modules in the system, including the data input module, the concept learning module, and the fuzzy ontology conversion module.
- ✧ System simulation and testing based on domain data prepared in previous stage.

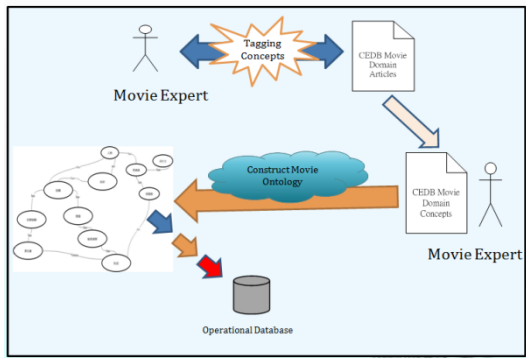
Evaluation Stage

This stage will focus on the system evaluation of the model.

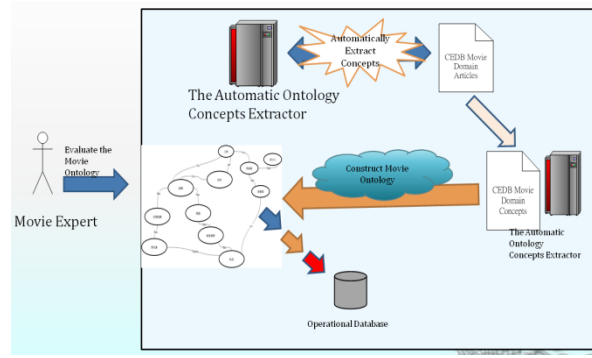
- ✧ Experimental testing of the model in terms of its ability in:
 - ◆ Providing concepts information
 - ◆ Extracting ontology constructs
 - ◆ Feedback from domain experts.
- ✧ Potential integration of the FOCM for the verification of ontology structure.

Note that some of the tasks involved in the above stages can be conducted in parallel depending on the availability of resources at the time.

(d) Reference (All references should be provided in full and include all authors.)



(a) Traditional system



(b) Automatic ontology concept extraction system

Figure 1. High Level View of Domain Ontology Construction

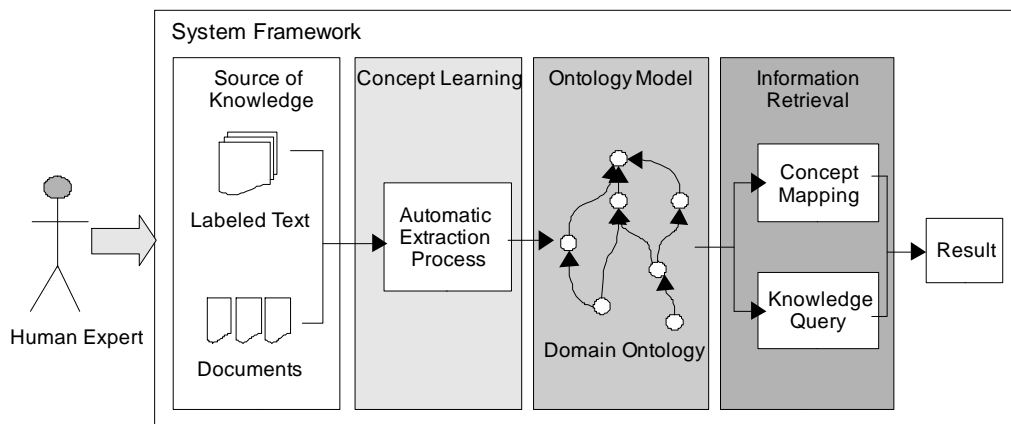


Figure 2. System Framework Overview

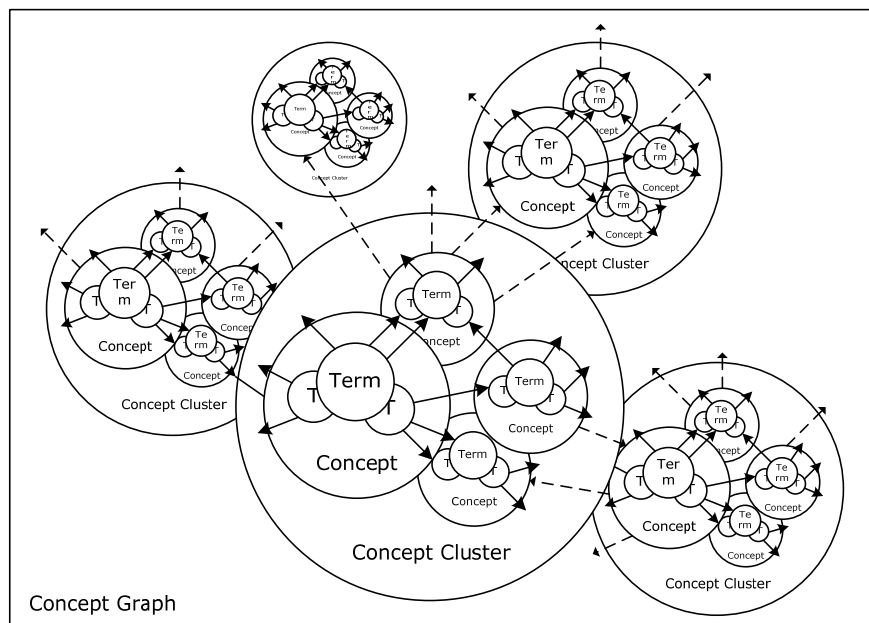


Figure 3. Term, Concept, Concept Cluster, and Concept Graph

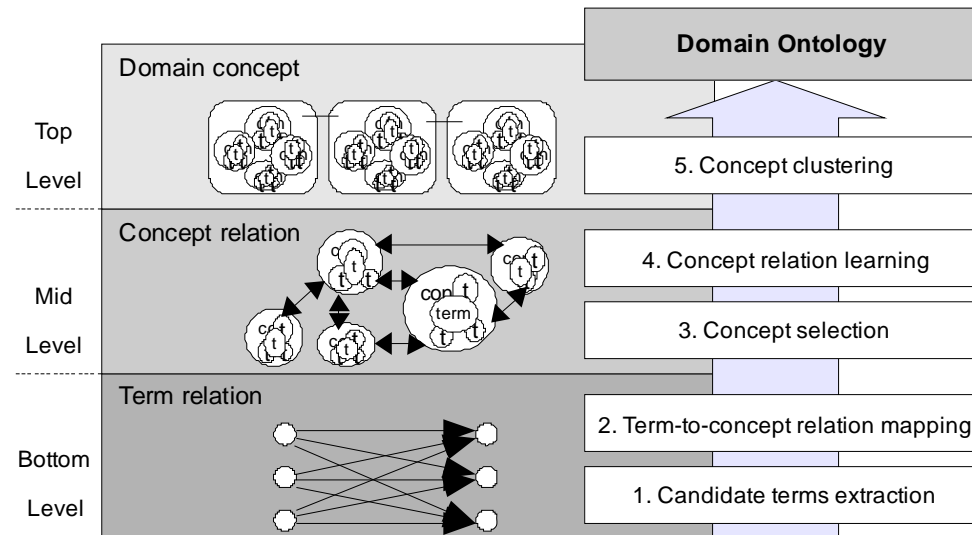


Figure 4. Concept Learning Process

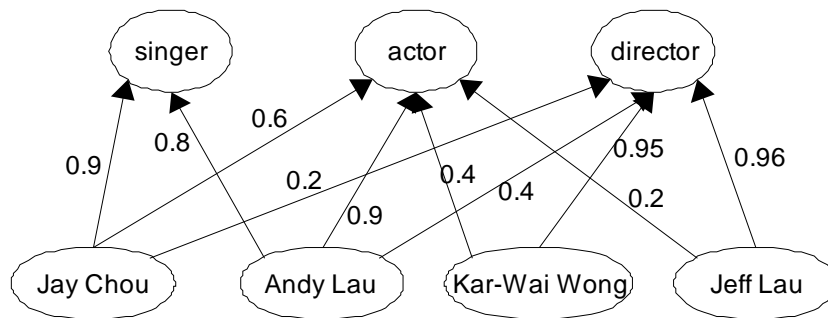


Figure 5. Term-to-Concept Mapping with Fuzzy Relationship

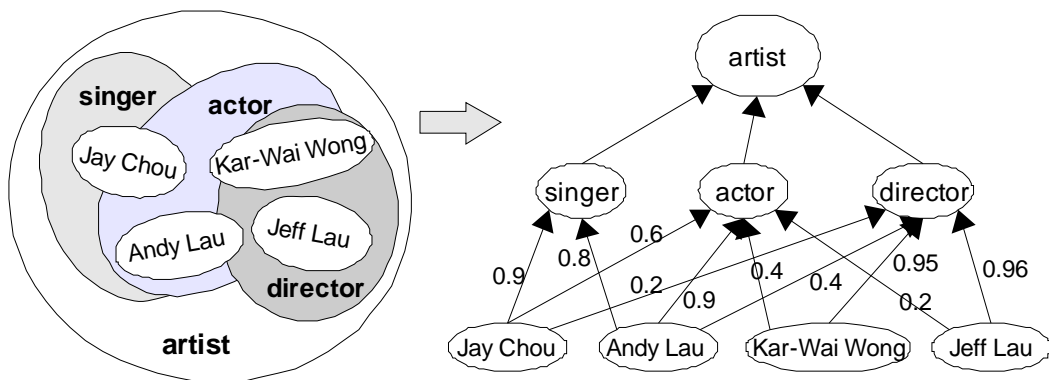


Figure 6. Concept Clustering and Taxonomical Relationship

- [1]. Alani, H., Kim, S., Millard, D.E., Weal, M.J., Hall, W., Lewis, P.H. and Shadbolt, N.R. (2003), Automatic Ontology-Based Knowledge Extraction from Web Documents, *IEEE Intelligent Systems*, Vol. 18, No. 1., pp. 14-21.
- [2]. Baziz, M., Boughanem, M., Aussenac-Gilles, N., Chrisment, C. (2005), Semantic Cores for Representing Documents in IR. *Proceedings of the 2005 ACM Symposium on Applied Computing SAC'05*, pp. 1011-1017.
- [3]. Bechhofer, S., Horrocks, I., Goble, C. and R. S. (2001), OilEd: A Reasonable ontology editor for the semantic web. *Proceedings of Joint German/Australian Conf. on Artificial Intelligence*.
- [4]. Calegari, S. and Farina, F. (2007). Fuzzy ontologies and scale-free networks analysis. *International Journal of Computer Science and Applications*, IV(II), pp. 125-144.
- [5]. Calegari, S. and Sanchez, E. (2008), Object-fuzzy concept network: an enrichment of ontologies in semantic information retrieval , *Journal of the American Society for Information Science and Technology*, v 59, n 13, pp. 2171-85.
- [6]. Chan, Manchung, Luo, Weidong and Liu, James N.K. (2007), Design and Implementation of a High Performance Distributed News Retrieval System, in *Proceedings of ICITM2007*, January 3-5, 2007, Hong Kong, pp. 26-33.
- [7]. Chen, S., Alahakoon, D. and Indrawan, M. (2005), Background knowledge driven ontology discovery. *Proceedings of the 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service*, pp. 202-207.
- [8]. Chen, L., Fankhauser, P., Thiel, U. and Kamps, T. (2005), Statistical relationship determination in automatic thesaurus construction, *Conference on Information and Knowledge Management*, pp. 267-268.
- [9]. Church K. and Hanks P. (1990), Word association norms, mutual information and lexicography, *Computational Linguistics*, Vol. 16, No. 1, pp. 22-29.
- [10]. Cimiano, P., Hotho, A. and Staab, S. (2005), Learning concept hierarchies from text corpora using formal concept analysis, *Journal of Artificial Intelligence Research*, 24:305-339.
- [11]. Ciravegna, F. (2001), Adaptive Information Extraction from Text by Rule Induction and Generalization. *Proceedings of 17th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pp. 1251-1256, Seattle, USA.
- [12]. Dittenbach, M., Berger, H., and Merkl, D. (2004), Improving domain ontologies by mining semantics from text. *Proceedings of the 1st Asia-Pacific Conference on Conceptual Modelling (APCCM2004)*, pp. 91-100.
- [13]. Farquhar A., Fikes, R. and Rice J. (1997), The ontolingua server: A tool for collaborative ontology construction. *International Journal of Human-Computer Studies*, Vol. 46, pp. 707-727.
- [14]. Feng F. and Croft W.B. (2001), Probabilistic techniques in phrase extraction, *Information Processing Management*, Vol. 37, pp. 199-220.
- [15]. Feng, L., Dillon, T. and Liu, J. (2001), "Inter-transactional association rules for prediction and their application to studying meteorological data", *Data and Knowledge Engineering*, Elsevier, 37, pp. 85-115.
- [16]. Frantzi K., Ananiadou S. and Mima H. (2000), Automatic recognition of multi-word terms: the C-value / NC-value method, *International Journal on Digital Libraries*, Vol. 3, pp. 115-130.
- [17]. Gan, Kok-Wee, Wang, Chi-Yung and Mak, Brian. (2002), Knowledge-Based Sense Pruning Using the HowNet: An Alternative to Word Sense Disambiguation. *International Symposium on Chinese Spoken Language Processing, ISCSLP 2002*, Taipei, Taiwan.
- [18]. Gruber, T. (1993). *What is an ontology?* <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
- [19]. Hang, X., Liu, J., Ren, Y. and Dai, H. (2005), *An Incremental FP-Growth Web Content Mining and its Application in Preference Identification*, in *Proceedings of Ninth International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES2005)*, September 14-16, 2005, Melbourne, Australia.
- [20]. Hernandez, N. and Mothe, J. (2004), An approach to evaluate existing ontologies for indexing a document corpus, *International Conference on Artificial Intelligence: Methodology, Systems, Applications, Semantic Web Challenges*, pp. 11-12.
- [21]. Hotho, A., Staab, S. Maedche A. (2001), Ontology-based Text Clustering, Workshop "Text Learning: Beyond Supervision".
- [22]. HowNet – Computation of Meaning, [Online]. Available : <http://www.keenage.com/>
- [23]. Hu, W. and Qu, Y. (2008), Falcon-AO: A practical ontology matching system, *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 6, no. 3, pp. 237-239.
- [24]. Jean-Mary, Y. R. and Kabuka, M.R. (2007), Asmov results for OAEI 2007, In: *Proceedings of the 2nd Ontology Matching Workshop*, Busan, South Korea.
- [25]. Kiu, C.C. and Lee, C.S. (2007), OntoDNA: Ontology Alignment Results for OAEI 2007, In: *Proceedings of the 2nd Ontology Matching Workshop*, Busan, South Korea.
- [26]. Koller, D. and Sahami, M. (1997), Hierarchically Classifying Documents Using Very Few Words, *Proc. 14th Int'l Conf. Machine Learning (ICML '97)*, pp. 170-178.
- [27]. LACHTIM, F.A., MOURA, A.M.C. and CAVALCANTI, M.C. (2009), Ontology matching for dynamic publication in semantic portals. *Journal of Brazilian Computer Society*, v. 15, n. 1. pp. 27-43.
- [28]. Lam, Toby H.W. and Lee, Raymond S.T. (2006), "An Intelligent Ontology-based Agent Framework for Semantic Web Service: Ontia iJADE", *Knowledge-Based Intelligent Information and Engineering Systems (LNCS 4252)*, pp. 637 – 643, Springer-Verlag, Berlin Heidelberg.
- [29]. Lam, Toby H.W. Lee, Raymond S.T. and Liu, James N.K. (2008), An Ontology-based Intelligent Mobile System for Tourist Guidance., *Computational Intelligence: A Compendium*, Springer-Verlag, pp. 381-406.
- [30]. Lam, H.W. and Liu, James N.K. (2008), Modeling a Personal Profile by Fuzzy Ontology Map and Its Applications, in *IEEE Proceedings of the International Conference on Fuzzy Systems (FUZZ2008) as part of 2008 IEEE World Congress on Computational Intelligence (WCCI2008)*, Hong Kong, June 1-6, 2008.
- [31]. Lau, R.Y.K., Hao, J.X., Tang, M. and Zhou, X. (2007), Towards context-sensitive domain ontology extraction, *Proceedings of the 40th Hawaii International Conference on System Sciences*, pp. 1-10.
- [32]. Lau, R.Y.K., Song, D., Li, Y., Cheung, T.C.H. and Hao, J.X. (2009), Toward a Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning, *IEEE Transactions on Knowledge and Data Engineering*, v 21, n 6, pp. 800-813.

- [33]. Lee, C.S., Chen, Y.J. and Jian, Z.W. (2003), Ontology-based fuzzy event extraction agent for Chinese e-news summarization, *Expert Systems with Applications*, 25, pp. 431–447.
- [34]. Lee, C.S., Jian, Z.W. and Huang, L.K. (2005), A Fuzzy Ontology and Its Application to News Summarization, *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 35, no. 5, pp. 859-880.
- [35]. Lee, R.S.T. and Liu, J.N.K. (2000), "Tropical Cyclone Identification and Tracking System using integrated Neural Oscillatory Elastic Graph Matching and hybrid RBF Network Track mining techniques", *IEEE Trans. on Neural Network*, **11**(3), pp. 680-689.
- [36]. Lee, Raymond S.T. and Liu, James N.K. (2001), "NORN Predictor – Stock prediction using a neural oscillatory-based recurrent network", *International Journal on Computational Intelligence and Applications (IJCIA)*, 1(4), pp. 439-452.
- [37]. Lee, R.S.T. and Liu, J.N.K. (2004), "iJADE Web-miner: An Intelligent Agent Framework for Internet Shopping", *IEEE Trans. on Knowledge and Data Engineering*, 16(4), pp. 461-473.
- [38]. Lee, Raymond S.T. and Liu, James N.K. and You, J. (2004), "iJADE WeatherMAN – A weather forecasting system using intelligent multi-agent based fuzzy-neuro network", *IEEE Trans. On SMC – Part C*, 34(3), pp. 369-377.
- [39]. Li, Yan, Shiu, Simon C. K., Pal, Sankar K. and Liu, James N. K. (2006), "A Rough Set-based Case-Based Reasoner for Text Categorization," *International Journal of Approximate Reasoning*, 41, pp. 229-255.
- [40]. Lim, Edward H.Y. and Lee, Raymond S.T. (2007), iJADE InfoSeeker: On using intelligent context-Aware agents for retrieving and analyzing Chinese Web articles. *Studies in Computational Intelligence*, 72, pp. 127-153.
- [41]. Liu, N.K. (1993), "Formal description technique for the verification of fuzzy knowledge base redundancy and subsumption", in *IEEE Proceedings of the 1st New Zealand International Conference on Artificial Neural Networks and Expert Systems*, Dunedin, New Zealand, November 1993, pp. 142-145.
- [42]. Liu, N.K. (1995), "A formal description techniques for verifying redundancy and subsumption in expert systems", *Australian Journal of Intelligent Information Processing Systems*, 2(1), pp. 24-35.
- [43]. Liu, N.K. (1996), "Formal verification of some potential contradictions in knowledge base using a High Level Net approach", *Applied Intelligence*, 6(4), pp. 325-344.
- [44]. Liu, J.N.K. (1997), "Quality prediction for concrete manufacturing", *Automation in Construction*, 5, pp. 491-499, Elsevier Science.
- [45]. Liu, N.K. (2001), "An intelligent business advisory system for stock investment", *Encyclopedia of Microcomputers*, 27, pp. 161-184.
- [46]. Liu, James N.K. (2007), Fuzzy Ontology Based System for Product Management and Recommendation, *International Journal of Computers*, Vol. 1, No. 3, pp. 117-124.
- [47]. Liu, N.K. and Dillon, T.S. (1992), "Modelling uncertainty in expert systems", in *Proceedings of the 2nd Pacific Rim International Conference on Artificial Intelligence*, Seoul, Korea, September 1992, pp. 610-616.
- [48]. Liu, N.K. and Dillon, T.S. (1991), "An approach towards the verification of expert systems using Numerical Petri Nets", *International Journal of Intelligent Systems*, 6, pp. 255-276.
- [49]. Liu, N.K. and Dillon, T.S. (1988), "Detection of consistency and completeness in expert systems", in J.S. Gero and R. Stanton, Eds., *Artificial Intelligence Developments and Applications*, North-Holland, pp. 119-134.
- [50]. Liu, N.K. and Dillon, T.S. (1995), "Formal description and verification of production systems", *International Journal of Intelligent Systems*, 10(4), pp. 399-442.
- [51]. Liu, James N.K. and Kwong, Raymond W.M. (2007), Automatic Extraction and Identification of Chart Patterns towards Financial Forecast, Special Issue Soft Computing for Time Series Prediction, 7(4), pp. 1197-1208, Elsevier.
- [52]. Liu, N.K. and Lee, K.K. (1997), "An intelligent business advisor for stock investment", *Expert Systems*, 14(3), Blackwell Publishers, UK, pp. 129-139.
- [53]. Liu, James N.K., Luo, Weidong and Chan, Edmond M.C. (2005), Design and Implement a Web News Retrieval System, Lecture Notes in Computer Science 3683, Springer pp. 149-156 (in *Proceedings of Ninth International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES2005)*, September 14-16, 2005, Melbourne, Australia).
- [54]. Liu, James N.K., Liu, Luo, W. D. and Chan, Edmond M.C. (2007), Design and Implementation of a Parallel News Retrieval System, accepted to appear in *International Journal of Knowledge-based and Intelligent Engineering Systems*, IOS Press.
- [55]. Liu, James N.K. and Leung, Tommy T.S. (2007), A Web-based CBR Agent for Financial Forecasting, ICFAI's (Institute of Chartered Financial Analysts of India) Professional Reference Book: Artificial Intelligence In Real World, ICFAI University Press.
- [56]. Liu, J.N.K. and Sin, D.K.Y. (2000), "A data mining approach for maintenance scheduling", *International Journal of Engineering Intelligent Systems*, 8(2), pp. 119-126.
- [57]. Liu, J. and Yung, K. W. (1996), "An intelligent system for respiratory protection", *Expert Systems with Applications*, 11(3), pp. 309-322.
- [58]. Liu, J., Zhou, L. and Wu, Y. (1999), "A hybrid method for abstracting newspaper articles", *Journal of the American Society for Information Science*, 50(13), pp. 1234-1245.
- [59]. Maedche A. and Staab S. (2001), Ontology learning for the Semantic Web, *IEEE Intelligent Systems*, Vol. 16, No. 2, pp. 72-79.
- [60]. Maedche, A., Neumann, T.G. and Staab, S. (2002). Bootstrapping on ontology-based information extraction system. *Intelligent Exploration of the Web*. P. Szczepaniak, et al., Heidelberg, Springer.
- [61]. Miller, G.A., Beckwith R., Fellbaum, C., Gross, D. and Miller, K.J. (1990), Introduction to WordNet: An on-line lexical database, *Journal of Lexicography*, Vol. 3, No. 4, pp. 234-244.
- [62]. Morin, E. (1999), Automatic acquisition of semantic relations between terms from technical corpora, *Proceedings of the 5th International Congress on Terminology and Knowledge Extraction (TKE 1999)*, Innsbruck, Austria, Aug. 1999, pp. 268-278.
- [63]. Navigli, R., Velardi, P. and Gangemi, A. (2003). Ontology learning and its application to automated terminology translation, *IEEE Intelligent Systems*, Vol. 18(1), pp.22-31.
- [64]. OilEd, <http://oiled.man.ac.uk>
- [65]. OWL (2004). *OWL Web Ontology Language Overview*, <http://www.w3.org/TR/owl-features>
- [66]. Perrin, P. and Petry, F. (2003), Extraction and Representation of Contextual Information for Knowledge Discovery in Texts, *Information Sciences*, vol.151, pp. 125-152.

- [67]. Protégé (2001). <http://protege.stanford.edu>. Technical report.
- [68]. Quan, T.T., Hui, S.C. and Cao, T.H. (2004), FOGA: A fuzzy ontology generation framework for scholarly semantic Web, Proceedings of the 2004 Knowledge Discovery and Ontologies Workshop (KDO'04), Pisa, Italy.
- [69]. Radev, D.R. and McKeown, K.R. (1998). Generating natural language summaries from multiple on-line sources. *Computational Linguistics*, 24(3), pp. 469-500
- [70]. RDF (2004). Resource Description Framework (RDF): Concepts and Abstract Syntax, www.w3.org/TR/rdf-concepts/
- [71]. Reidsma, D., Kuper, J., Declerck, T., Saggion, H. and Cunningham, H. (2003), Cross document annotation for multimedia retrieval. EACL Workshop on Language Technology and the Semantic Web, Budapest.
- [72]. Rubin, D.L., Hewett, M., Oliver, D.E., Klein, T.E. and Altman, R.B. (2002), Automating data acquisition into ontologies from pharmacogenetics relational data sources using declarative object definitions and XML, *Proceedings of the 7th Pacific Symposium on Biocomputing*, Hawaii, USA, Jan. 2002, pp. 88-99.
- [73]. Salton, G. and McGill, M. J. (1983), *Introduction to modern information retrieval*. s.l. : McGraw-Hill.
- [74]. Sebastiani, Fabrizio (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1-47.
- [75]. Shiu, S.C.K., Liu, J.N.K. and Yeung, D.S. (1997), "Formal description and verification of hybrid rule/frame based expert systems", *Expert Systems with Applications*, 13(3), pp. 215-230.
- [76]. Sridharan, B., Tretiakov, A., Kinshuk (2004), Application of Ontology to Knowledge Management in Web based Learning. IEEE International Conference, pp. 663-665.
- [77]. Stojanovic, L., Stojanovic, N. and Volz, R. (2002), Migrating Data-Intensive Web Sites into the Semantic Web, *Proceedings of the ACM Symposium on Applied Computing (SAC)*, Madrid, Spain, March 2002, pp. 1100-1107.
- [78]. Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R. and Wenke, D. (2002), Ontoedit: Collaborative ontology development for the semantic web. Proceedings of International Semantic Web Conference.
- [79]. Tho, Q.T., Hui, S.C., Fong, A.C.M., Cao, T.H. (2006), "Automatic Fuzzy Ontology Generation for Semantic Web", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18(6), pp. 842- 856.
- [80]. Vargas-Vera, M., Motta, E., Domingue, J., Buckingham Shum, S. and Lanzoni, M. (2001), Knowledge extraction by using an ontology-based markup & semantic annotation, 1st Int. Conf. on Knowledge Capture, pp. 5-12, Victoria, B.C. Canada.
- [81]. Widuanto, D.H., and Yen, J. (2001), A Fuzzy Ontology-Based Abstract Search Engine and Its User Studies, *Proceedings of the 10th IEEE International Conference on Fuzzy Systems*, Melbourne, Australia, Dec 2001, pp. 1291-1294.
- [82]. Wu J.L. and Agogino A.M. (2004), Automatic keyphrase extraction with multi-objective genetic algorithms, Proceedings of the 37th Annual Hawaii International Conference on System Sciences, pp. 401-403.
- [83]. Wu, Y., Liu, J.N.K. and Wang, K.Z. (1998), "A method of automatic analysis of text structure", in *Proceedings of International Conference on Chinese Information Processing (ICCIIP'98)*, November 18-20, 98, Beijing, PRC, pp. 304-309.
- [84]. Zadeh, L.A. (2003), Web Intelligence and Fuzzy Logic – The concept of Web IQ (WIQ), *Invited talk at the 2003 IEEE/WIC Int. Conference on Web Intelligence (WI 2003)*,
- [85]. Zhang, X., Liu, J. and Atwell, E. (1997), "A multilingual information retrieval tool hierarchy for a WWW virtual corpus", in *Proceedings of the 2nd Workshop on Multilinguality in Software Industry: The AI Contribution (MULSAIC'97) of the International Joint Conference on Artificial Intelligence (IJCAI-97)*, August 23-29, 1997, Nagoya, Japan, pp. 17-23.
- [86]. Zhou, L. and Liu, J. and Yu, S.W. (1998), "Automatic extraction of word pairs based on heuristics and statistical method", *International Journal on Computer Processing of Oriental Languages*, 11(4), pp. 339-351.

Additional references: work done by PI in the last year

- [87]. Toby H.W. Lam, James N.K. Liu and Raymond S.T. Lee (2009) MASTER: An Intelligent Ontology-based Multi-Agent System For Sightseer, *International Journal of Software Engineering and Knowledge Engineering World Scientific*, vol. 19, No. 2, pp. 137-157.
- [88]. Raymond S.T. Lee, James N.K. Liu, Karo S.Y. Yeung, Alan H.L. Sin and Dennis T.F. Shum (2009), Agent-based Web Content Engagement Time (WCET) Analyzer on e-Publication System, accepted to appear in Proceedings of the 9th International Conference on Intelligent Systems Design and Applications, November 30 – December 2, Pisa, Italy.
- [89]. Hongwei Wang and James N.K. Liu (2009), Analysis of Semantic Heterogeneity Using the Ontological Structure Based on Description Logics, in Proceedings of ICNC'09 - FSKD'09, Special Session on Ontology Methodologies and Applications (OMA), Tianjin, China, August 14-16, 2009, pp. 216-221.
- [90]. Hongwei Wang, James N.K. Liu and Honghua Dai (2009), Acquisition of Domain Ontology from Relational Schemes Based on Reverse Engineering, in Proceedings of ICNC'09 - FSKD'09, Special Session on Ontology Methodologies and Applications (OMA), Tianjin, China, August 14-16, 2009, pp. 211-215.
- [91]. Edward H.Y. Lim, James N.K. Liu and Raymond S.T. Lee (2009), Knowledge discovery from text learning for ontology modeling, in Proceedings of ICNC'09 - FSKD'09, Special Session on Ontology Methodologies and Applications (OMA), Tianjin, China, August 14-16, 2009, pp. 227-231.
- [92]. Edward H.Y. Lim, Hillman W.K. Tam, Sandy W.K. Wong, James N.K. Liu and Raymond S.T. Lee (2009), Collaborative Content and User-based Web Ontology Learning System, in Proceedings of FUZZ-IEEE 2009, ICC Jeju, Jeju Island, Korea, August 20-24, 2009.
- [93]. Edward Lim, Raymond S.T. Lee and James N.K. Liu (2008) Knowledge seeker – an Ontological Agent-based System for Retrieving and Analyzing Chinese Web Articles, in IEEE Proceedings of the International Conference on Fuzzy Systems (FUZZ2008) as part of 2008 IEEE World Congress on Computational Intelligence (WCCI2008), Hong Kong, June 1-6, 2008
- [94]. James N.K. Liu and M.K. Ho (2008) An Intelligent Agent-based Systems for Multilingual Financial News Digest, Lecture Notes in Artificial Intelligence 4953, Springer, pp. 822-831. Proceedings of the 2nd KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications (AMSTA 2008), Korea, March 27-28, 2008, Springer Verlag, pp. 822-831.

3(a). Estimated cost and resource implications:

[Detailed justifications should be given in order to support the request for each item below]

(a maximum of 200 words for each box)

	Year 1 (HK\$)	Year 2 (HK\$)	Year 3 (HK\$)	Year 4 (HK\$)	Year 5 (HK\$)	Total (HK\$)
(i) Staff [see enclosure VI] #						
<u>Rank</u>						
Monthly salary x No. x Months						
<u>Research Assistant</u> 216,000						
12,000 * 1 * 6	72,000					
12,000 * 1 * 12		144,000				
(ii) Other Personnel (Payment to Co-I is prohibited)						
Monthly salary x No. x Months 0						
<u>Other Personnel</u>						
(iii) Relief Teacher						
[see enclosure III for individual research and Enclosure V for relief support under Humanities Sub-Panel]						
<u>Rank</u>						
Monthly salary x Months						
(iv) Equipment ^						
[please itemize and provide quotations for each item costing over HK\$200,000]						
N/A	0	0	0	0	0	0
(v) General Expenses [please itemize]						
- reference books and conference proceedings	6,000	6,000	6,000	0	0	18,000
- data collection and preprocessing						

(vi) Research Expenditure for Undergraduate Student(2,500 / month)

Monthly salary x No. x Months

0

(vii) stipends to Research Postgraduate Student @

Monthly salary x No. x Months

480,600

13,350 * 1 * 12

160,200

13,350 * 1 * 12

160,200

13,350 * 1 * 12

160,200

(viii) Conference Expenses
[see Explanatory Notes]

Budget provision for attending international conferences or workshops of related research interest.

12,000

12,000

12,000

0

0

36,000

Total cost of the project

250,200

322,200

178,200

0

0

750,600

On costs (An additional 15% of the amount requested from RGC)

112,590

Grand Total amount requested

863,190

[Please refer to Part II Section 3 of GRF2 for items permissible in the budget]

PI/Co-Is are not regarded as staff and must not be remunerated with salaries.

^ RGC fund must not be used to purchase personal electronic devices such as cell phones,iPod, MP3 Players, digital cameras and PDAs except with sound justifications, in addition, the RGC will not provide funding for the purchase or use of standard equipment (e.g. desktop PCs, servers, laptop computers, printers, scanners) and standard software licence/dataset (e.g. Windows, MS-Office) known to be available or reasonably expected to be provided in the institutions concerned.

@ Standard rate of stipends of the institution concerned should be used.

- (b). (i) I declare that no equipment is required
- (ii) I declare that the equipment indicated at 3(a)(iv) above is not available in the institution
- (iii) I declare that all or some of the equipment(please provide details

in the following text box) indicated at 3(a)(iv) above is available in the institution but cannot be used by me in view of the following reasons(a maximum of 200 words)

Reasons : (a maximum of 200 words)

(c). Declaration on the research-related software licence / dataset

- (i) I declare that no research-related software licence/dataset is required
- (ii) I declare that the research-related software licence/dataset indicated at 3(a) above is not available in the institution
- (iii) I declare that all or some of the research-related software licence/dataset (please specify and provide details in the following text box) indicated at 3(a) above is available in the institution but cannot be used by me in view of the following reasons(maximum 200 words)

Reasons : (Maximum 200 words)

Justifications for each category/item of the budget in the section of "Estimated Cost" in

Part II:

[Detailed justifications should be given in order to support the request]

(maximum 200 words for each box)

(a) Staff

Due to the specific need of preprocessing a large quantity of domain information content, a RA is required to support the data collection and

investigation work. The major roles of this person include:

- assist in preparing the domain contents in digital format
- assist in data preprocessing including content markup and program concept clustering
- assist in the simulation process and analysis
- assist in system testing

(b) Other Personnel

N/A

(c) Relief Teacher

N/A

(d) Equipment

N/A

(e) General Expenses

General expenses are budgeted to support data collection and preprocessing, help purchase related books and conference proceedings as to keep the related research knowledge up-to-date.

(f) Research Expenditure for Undergraduate Student

N/A

(g) Stipends to Research Postgraduate Student

The project work provides indepth research opportunities for training PhD students. The provision of stipends is necessary to attract good quality research staff and provide incentive for a sustainable research development

The major roles of this RS include:

- assist in the theoretical work for the R&D of ontology engineering
- assist in the design concepts association framework
- assist in the simulation process for knowledge extraction
- assist in model testing and system evaluation

(h) Conference Expenses *

The provision of conference attendance is to facilitate the publishing of research results and professional exchange of views in some international forums.

* A standard provision of \$12,000 per year for each funded project may be provided for travelling in connection with attending conferences/meetings outside Hong Kong upon request.

**4. Existing facilities and major equipment available for this research project:
(a maximum of 400 words)**

The PI and the associated Co-Is have basic computing facilities in their offices / research Lab. In addition, the external collaborator will assist in the preparation of specific domain information for our model testing and analysis.

5. Funds secured or to be secured

(a). Other research funds already secured for this research proposal:

[This amount will be deducted from the total cost of the project in Section 3 of Part 1 above.]

SourceAmount
(HK\$)*

(b). Other research funds to be or are being sought for this research proposal.

[If funds under this item are secured, the amount of the GRF to be awarded may be reduced]:

SourceAmount
(HK\$)

DECLARATION OF SIMILAR OF RELATED PROPOSALS

[Please refer to GRF2 for implications for 7(b) - 7(d)]

6 Re-submission of a proposal not supported previously

(a). Is this proposal a re-submission or largely similar to a proposal that has been submitted to but not supported by the RGC or other funding bodies ?

Yes No

If yes, please state the funding agency(ies) and the funding programme(s):

GRF

Reference No(s).: [for RGC projects only]: 524609

Project title(s) : [if different from Item 1(a) of part I above.]

Automatic extraction of fuzzy domain ontology

Date(month/year) of application: 10/2008

Outcome: not supported

(b). If this application is the same as or similar to one(s) submitted but not supported previously, what were the main concerns / suggestions of the reviewers then ?

- (1) Some declared objectives are quite vague and should have been described in more detail.
- (2) The bibliographical first "preparation" stage should have been finished before applying to this grant.
- (3) Due to the amount of potential implications, a three-year project is suggested, if possible.
- (4) Regarding the description of the work done by other researchers, the reviewer does not agree with the description: "The management of the ontology is fulfilled by the software agent which takes over the services required to define and update the ontology items."
- (5) The reviewer is not clear about the relationship between the previously given CERG grant and the present application. How do they complement each other?
- (6) Figures 1 and 2 are not precise, which makes them hard to interpret. The figure should explicitly show the data sources and depict both data and control flow.
- (7) The reviewer expected that the section on Research Plan and Methodology should provide more information about the plan to extract ontological knowledge in a fuzzy way.
- (8) The reviewer is not clear how to deal with the overlapping of concepts within a

concept graph (e.g. Figure 3) for the research.

(c). Please give a brief response to the points mentioned at 6(b) above, highlighting the major changes that have been incorporated in this application.

Based on the valuable comments of previous reviewers, we have been able to revise our project proposal with specific attention to corresponding comments as follows:

(1) Those objectives are now described in detail.

(2) With the clarification of those objectives in detail, we realize that the preparation stage denotes system planning and domain data preparation. A large effort will be anticipated to deal with data collection and pre-processing of domain information contents in digital form. Most of these domain contents are not in readily format for the proposed system development.

(3) As advised by the reviewer and the fact that the project can provide in-depth research opportunities for training PhD students, we are now making provision for a three-year project instead.

(4) The said description is now deleted from the revised proposal.

(5) Additional description is now provided in the revised proposal. As stated, the previous given CERG grant can, on one hand allow us to exploit the use of Fuzzy Ontology Map to model fuzzy concepts, and on the other hand, the use of high level Petri Nets which is a kind of formal description technique to check for the correctness and consistency of the ontology systems. This has the potential to help make a better Ontology Extractor.

(6) Explicit details and descriptions are added to show the framework of the proposed e-publication system (Figure 2), the concept learning process (Figure 4), etc.

(7) Additional information and illustrative examples (Figures 5, 6) are now added to show some typical relationships such as the term-to-concept mapping with fuzzy relationship, the concept clustering, the taxonomical relationship that domain ontology could be involved for project consideration.

(8) Illustrative examples are now added (e.g. Figures 3, 6) to show the possible means of modelling the overlapping of concepts and their relationships.

7. Previous similar or related research proposals

(a). Is there similar or related research being carried out at your institution(s)?

Yes

No

If yes, please give a brief account including names of investigators, departmental and institutional affiliations, project title(s) and nature of the project(s): (a maximum of 400 words)

(b). Is/are there **any other** proposal(s) being submitted by **PI or Co-Is** to the RGC (including those submitted by **PI or Co-I** through other institutions) in this funding exercise?

Yes

No

If yes, please give the following details -

Proposal(s) submitted by the PI in the capacity of Co-I

No.	Project Title	Name of PI	Explanation
1	The fusion of multi-channel medical features for non-invasive diabetic care	Dr Jane You	With the advances in sensor and medical imaging technologies, there have been various devices developed for diabetic diagnosis. However, most of the existing non-invasive approaches are based on the assessment of the specific clinical signs which are shown in the particular part of the body, for example, special sensors for blood glucose level reading, ophthalmological methods for diabetic retinopathy analysis, ultrasonic scanners for diabetes-related foot ulcer and neuropathy detection. In addition, some of these systems are subject to other limitations such as lack of the quantitative description of the disease progression/regression, high cost of some equipment for data collection, time consuming and special professional skills required for data analysis, inconsistent assessment result among different observers, health side effect of radiation and high cost. Although the main symptoms of diabetes occur in many of the body systems including the circulatory, nerve,

digestive, endocrine and urinary systems, an integrated system which makes full use of multiple features from different parts of the body for the comprehensive examination of diabetes has not yet been available. It remains a challenging task to develop a high performance non-invasive system to overcome the current limitations for effective diabetic identification and analysis.

This project aims to tackle the key issues of non-invasive diabetic care by fusion of multi-channel medical features from eye, foot and blood vessels for effective diabetic screening, diagnosis and monitoring. In contrast to the existing approaches, we propose to investigate, design, analyze, implement, evaluate and prototype a new computer-aided system, which includes a set of unique sensors to capture multi-spectrum medical data from eye, foot and body tissue -- near infrared, visible and ultrasonic; a general approach to multiple feature extraction and representation, a diabetic indexing scheme for multi-channel feature fusion and assessment; a novel discrimination function for classification. Our previous work on diabetic retinopathy detection, biometrics computing and ultrasonic diabetic foot analysis have laid the basis for further development of the relevant research issues of this grant application. The findings throughout this proposed project will be not only beneficial to medical applications to improve the quality of life and life expectancy, but also important to multidisciplinary research in a diversity of areas including imaging, multimedia information systems and pattern recognition with excellent potentials for economic productivity and growth. Feasibility study, system prototyping and a series of benchmarking testing for diabetic screening, diagnosis and monitoring will be carried out in the project.

(c). Is there similar or related proposal being submitted by PI or Co-I(s) to other competitive funding schemes of the UGC / other public or government funding agencies?

Yes No

If yes, please give the following details -

(i) The funding agency(ies) and the funding programme(s):

(ii) Reference No(s): (for RGC funding schemes only)

(iii) Proposal Title(s):

(iv) A brief account of the proposal(s) including month & year of application and an explanation on the differences between the proposal(s) concerned and this application: (a maximum of 400 words)

(d). Is there similar or related project by PI or Co-I(s) already funded by other competitive funding schemes of the RGC / other public or government funding agencies?

Yes No

If yes, please give the following details -

(i) The funding agency(ies) and the funding programme(s):

(ii) Reference No(s): (for RGC funding schemes only)

(iii) Proposal Title(s):

(iv) A brief account of the proposal(s) including month & year of application and an explanation on the differences between the proposal(s) concerned and this application: (a maximum of 400 words)

(iv) A brief account of the proposal(s) including month & year of application and an explanation on the differences between the proposal(s) concerned and this application: (a maximum of 400 words)

COLLABORATION PLAN

8. Plan(s) for collaboration in this application:

[Indicate the role and the specific task(s) the PI and each Co-I, if any, is responsible for.]

[Letter(s) of collaboration can be attached to section 14]

The hours to be spent by the PI on the project each week are (4,3,2) in each of the three years. Those spent by the co-investigators (Dr Raymond Lee and Dr Jane You) are

(1,1,1) and (1,2,1) respectively. There will be regular discussion with the technical team of our external collaborator (i.e. Dr Raymond Lee) being the facilitator at IATOPIA through our placement programme and joint research projects.

The PI, Dr James Liu will act as the project manager overseeing the progress of the while project as well supervision of project staff including research student, if applicable. He is an expert in knowledge base construction, intelligent techniques and related testing.

Accordingly, the PI will be responsible for

- knowledge base engineering techniques
- concept learning and extraction algorithms
- design of domain ontology construct
- potential integration with FOCM
- supervision of research staff and students for the theory establishment on the design and implementation of the project.

The first co-investigator, Dr Raymond Lee – who is the expert in ontology construct and content management system development will offer assistance in the preparation of specific domain contents for our modeling and testing of those extraction algorithms.

The second co-investigator, Dr Jane You is an expert in Web mining and multi-media knowledge engineering. She will be providing expert advice and direction on fuzzy domain modeling and mining of concept associations.

In addition, the PI and both co-investigators will work together to setup the testing criteria for system evaluation in the end.

(a). Details of on-going and completed research projects funded from all (GRF and non-GRF) sources undertaken by the **PI** (in a PI or Co-I capacity) in the past **five** years.

(i). Completed GRF and non-GRF projects(no abstract is required)

(ii). On-going GRF and non-GRF projects.

[Please attach a copy of the original abstract of each listed project]

Project Title	PI/Co-I	Project Ref No.	Funding Source(s) and Amount(HK\$)	Start Date	Completion Date
Fuzzy ontology based system for content management and recommendation	PI	523708	GRF HK\$ 517,796	01-01-2009	31-12-2010

Original Abstract of the Project for On-going Project only:

With the rapid growth of Internet and the amount of digital content continues to proliferate, some form of content management process or system is becoming essential for most organizations with a significant Web presence. Despite this requirement, in the current World Wide Web (WWW), users often find it difficult to locate relevant information using search engines. The Web system also lacks recommendation features to support decision processes which can provide a more personalized and proactive retrieval experience for users. The problem of it may be due to the fact that the current World Wide Web lacks semantic markup as most web documents are not fully structured sources of information. One of the possible solutions for this problem is Semantic Web. In the latest Semantic Web technology, descriptive markup languages, such as Resource Description Framework (RDF) and Web Ontology Language (OWL), were proposed to model the web content in a machine-readable way which assists information gathering and automatic searching by software agents. Latest research attempts to extend ontology representation to model uncertain information. The approach appears to be useful for information searching though the ontology structure is somewhat ambiguous and artificial.

This project proposes an extension of the current ontology representation which is based on the integration of fuzzy logic, graph theory and Petri Net theory. By considering ontology as a graph with concept descriptors and explicit control mechanism, an ontology-based colored Petri Net can be constructed in which a concept descriptor is represented as a fuzzy relation which encodes the degree of a feature value using some fuzzy membership function. As a result, a Fuzzy Ontology Colored Map (FOCM) can be formed which is a connection matrix collecting the membership values between concept classes in the ontology graph. Thus, a fuzzy ontology could be created by using the FOCM and the

ontology document (RDF/OWL). It represents the domain knowledge in a structured and machine-interpretable form.

We will focus on Web applications to develop fuzzy ontology-based systems for content management and recommendation. In particular, we exploit ontology formulation from various domains involving imprecise or vague Web information. This is in support of modelling concept association and recommendation preferences. Moreover, the potential of ontology verification based on fuzzy logic and Petri Net theory is to be investigated. It is possible to examine the dynamic behaviour of FOCM at different transitions to determine the correctness and consistency of the ontology system.

CONTACT – a	PI	522407	GRF	HK\$ 364,320	01-01-	31-12-2009
Chaotic Oscillatory-					2008	
based Neural model						
for Turbulence						
Analysis &						
foreCasTing						

Original Abstract of the Project for On-going Project only:

Current weather forecast research based on numerical prediction models, statistical models such as the Naïve Bayesian Networks, or neural network models including Back-Propagation Neural Networks and recurrent neural networks provide challenging results on long-term weather prediction (such as temperature and relative humidity). However, for meso-scale and even micro-scale severe weather phenomena (with highly short-term chaotic perturbations), such as rainstorms, thunderstorms and even turbulence and wind shear phenomena, these tools have failed to apply. As an extension of the team's latest theoretical work on chaotic neural oscillators as a new Bifurcation Transfer Unit for neural network models, the main objective of this project (CONTACT) will focus on the applied research including the use of LEE-oscillator for the study of a meso-scale weather phenomenon – Turbulence, a challenging and complex meteorological phenomena which have vital impact on aviation safety. For the system evaluation perspective, this project will base on the related meteorological information provided by Hong Kong Observatory and Juneau International Airport. We explore the feasibility for the modelling and prediction of turbulence phenomena in HK International Airport, and to serve as a critical mass for the future applied research of turbulence modelling and forecasting research in aviation forecast.

(b). Details of on-going and completed research projects funded from all (GRF and non-GRF) sources undertaken by each **Co-I** (in a PI capacity) in the past **three** years.

(i). Completed GRF and non-GRF projects (no abstract is required)

(ii). On-going GRF and non-GRF projects

[Please attach a copy of the original abstract of each listed project]

10. Major research output of previously funded projects (GRF and non- GRF sources) in descending chronological order, undertaken by the PI and each Co-I relevant to this application.

[Provide a summary (maximum 400 words in total) on the progress / publications / conferences / student-training, etc. of the projects, with the relevant project reference no.]

Ongoing 523708 GRF

Outputs: 2 project students, 1 journal, 1 book chapter, 7 conference papers, e.g.

(1)Lam, Liu and Lee (2009) MASTER: An Intelligent Ontology-based Multi-Agent System For Sightseer, International Journal of Software Engineering and Knowledge Engineering, Vol. 19, No. 2, pp. 137-157.

(2)Lim, Lee and Liu (2008) Knowledgeseeker – an Ontological Agent-based System for Retrieving and Analyzing Chinese Web Articles, IEEE FUZZ2008 of WCCI2008.

Ongoing 522407 GRF

Outputs: 1 full-time & 1 part-time MPhil students, 2 project students, 6 conference papers, e.g.

(1) Kwong, Liu, Chan and Lee (2008) Using LIDAR Data and Chaotic Oscillatory-based Neural Network for the forecast of Doppler Velocities, IEEE CEC2008 of WCCI2008.

(2) Wong, Lee and Liu (2008) Wind Shear Forecasting by Chaotic Oscillatory-based Neural Networks (CONN) with Lee Oscillator (Retrograde Signalling) Model, IEEE Proceedings of the International Joint Conference on Neural Networks

Completed CERG B-Q515

Outputs: 1 PhD, 3 journal papers, 3 book chapters, 4 conference papers, e.g.

(1)Liu et al. (2005) iBotGuard: An Internet-based intelligent robot security system using invariant face recognition against intruder, IEEE Trans. On SMC – Part C, 35(1), pp. 97-105.

(2) Lee, Liu and You (2004) iJADE WeatherMAN – A weather forecasting system using intelligent multi-agent based fuzzy-neuro network, IEEE Trans. On SMC – Part C, 34(3), pp. 369-377.

Completed ICRG Project: G-T375

Outputs: 1 MPhil, 2 journal papers, 4 conference papers, e.g.

(1)Liu et al. (2003), "Smart Shopper: An Agent Based Web-mining Approach to Internet Shopping ", Special Issue on Knowledge Discovery and Data Mining by IEEE Transactions on Fuzzy Systems, 11(2), pp. 226-237.

(2)Lee and Liu (2004), "iJADE Web-miner: An Intelligent Agent Framework for Internet Shopping", IEEE Transactions on Knowledge and Data Engineering, 16(4), pp. 461-473.

Completed CERG B-Q569

Outputs: 1 MPhil, 3 journal papers, 1 conference paper, e.g.

(1) Lee and Liu (2002), "Scene analysis using an integrated composite Neural Oscillatory elastic graph matching model", Pattern Recognition, 35, pp. 1835-1846.

(2)Lee and Liu (1999), "An Automatic Satellite Interpretation of Tropical Cyclone Patterns

Using Elastic Graph Dynamic Link Model”, International Journal of Pattern Recognition and Artificial Intelligence, 13(8), pp. 1251-1270.

Completed ICRG A-PG50

Outputs: 1 PhD, 1 MPhil, 2 books, 2 journal papers, 1 book chapter, 1 conference paper, e.g.

(1) Lee (2006), Fuzzy-Neuro Approach to Agent Applications (From the AI Perspective to Modern Ontology), Springer-Verlag, Heidelberg.

(2) Lee, "LEE-Associator - A Transient Chaotic Autoassociative Network for Progressive Memory Recalling," Neural Networks 19 644-666.

ANCILLARY INFORMATION

11. Research ethics/safety approval:

ANCILLARY INFORMATION

[It is the responsibility of the institution and the PI to ensure that the research proposal is carefully reviewed for its compliance with applicable laws, health and safety guidelines and ethical standards. Ethics clearance should be sought for research involving living animals and/or human subjects including social science research involving human subjects (e.g. potential physical or psychological harms, discomfort or stress to human participants that a research project might generate, subjects' privacy etc.). The primary responsibility of seeking the relevant approval and ethics clearance rests with the PI. The PI's institution is required to complete and sign Part III of this application form to *confirm whether the research proposal involves human subjects and certify whether the relevant approval is required and if required, the relevant approval has been given / is being sought.*]

(a) I confirm that the research proposal involves* does not involve* human subjects.

[Note: All proposals involving human subjects MUST obtain ethics clearance.]

* Please tick as appropriate.

(b) Please tick the appropriate boxes to confirm if approval for the respective ethics and/or safety issues is required and has been / is being obtained from the PI's institution.

	Approval not required	Approval obtained	Approval being sought
(i) Human research ethics *	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(ii) Animal research ethics	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(iii) Biological safety	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(iv) Ionizing radiation safety	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(v) Non-ionizing radiation safety	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(vi) Chemical safety	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

(c) If approval is required by **other authorities**, please indicate *below* the names of the authorities and the prospects of obtaining such approval. If not applicable, please put down "N.A.".

N.A.

* Including ethics involving human subjects in social science research (e.g. potential physical or psychological harm, discomfort or stress to human subjects in a research project, subjects' privacy etc.) All proposal involving human subjects MUST obtain ethics approval/exemption.

12(a). List of proposed reviewers:

Points to note before completion:-

- This is NOT a compulsory section but nomination of reviewers, especially in very specialized field area is most welcome. This list serves as a reference for the RGC Panel. The named reviewer(s) may or may not be chosen to review the application.
- Applicant(s) can nominate none or a maximum of five reviewers. They should preferably be experts whom the applicant(s) has no relationship with. If however the applicant(s) i.e., the PI as well as the Co-I(s), decide to nominate reviewers with a past or present relationship, a declaration on the association must be made. It is the responsibility of the PI and the Co-I(s) to ensure that all relationships are fully and accurately declared. Failure of PI and / or Co-I(s) to disclose fully or accurately the relationship will result in disqualification of the application, which will also be taken into account by the RGC as part of the track record of the concerned PI / Co-I(s) when assessing future applications from the same PI / Co-I(s) in the capacity of PI. The RGC reserves the ultimate right to reject or disqualify future applications in serious cases.
- **PIs must ensure that their nominations of external reviewers are made known to the Co-I(s). Otherwise, the Co-I(s) will not be able to declare relationship with the nominated reviewers as required, and thus leads to disqualification of the applications.**
- Please DO NOT put down here the name(s) of any reviewer(s) whom the applicant(s) may wish to exclude from being invited for assessment.

12(b). Declaration of any past and present relationship between the investigator(s) i.e., PI and Co-Is, and the nominated reviewers [minimum one tick per reviewer]:

Nature of relationship (please elaborate in 12(c))	Reviewer				
	(i)	(ii)	(iii)	(iv)	(v)
Supervisor or Supervisee of PhD or MPhil studies	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Advisor or Advisee in research studies/research projects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Colleagues in the same organization (please specify if in the same department)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Collaborators in research projects/programmes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Co-authors of papers/patents/publications	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Partners or co-organizers of major events	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Long-time personal friends	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Teacher or student at undergraduate studies	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Members of the same editorial board	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Others (please specify)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
None	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

12(c). Elaboration on the nature of the relationship declared in Section 12(b) (e.g. when and where the relationship was / is developed, name / nature of project, publications or events involved):

13. DATA ARCHIVE POSSIBILITIES

Is the proposed project likely to generate data set(s) of retention value?

Yes

No

If yes, please describe the nature, quantity and potential use of the data set(s) in future.

(Note: If your answer to the previous question is 'No', there is no need to answer the following question.)

Nil

Are you willing to make the data set(s) available to others for reference twelve months after the publication of research results or the completion of this proposed project?

Yes No

I/We understand that the RGC only considers data archiving requests after the completion of the RGC-funded project, and the Council has full discretion in funding the archiving requests. Data sets archived with RGC funds will require users to acknowledge the originator and the RGC. The originator will also be provided with copies of all publications derived from the use of the data.

14. Supporting Documents (e.g. : Letters of collaboration, ethics/safety approval letters etc.)

Note: i) Applicants should not make use of the section to supplement the contents of other sections.

ii) Maximum 20 words for each box.

Appendix 1: Collaboration letter from a content provider

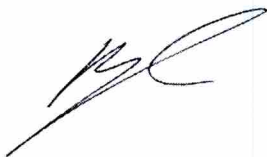
05 October 2009

Dr James Liu
Associate Professor
Department of Computing
The Hong Kong Polytechnic University

Project: Automatic Extraction of Fuzzy Domain Ontology for Content Management and Information
Retrieval in e-Publication Systems

I am pleased to be able to participate in this research project as one of the Co-investigators. I am willing to provide domain expertise, office assistance to system testing, evaluation where applicable to ensure the completion of this innovative and applied research.

Yours sincerely,



Dr Raymond Lee
President
IATOPIA Group Limited

PART III INSTITUTIONAL ENDORSEMENT AND DECLARATION OF RESEARCH ETHICS/SAFETY

[To be completed and signed by the appropriate authority of the institution. The institution should confirm that it has evaluated and given support to the application before submission to the RGC. It is required to verify that a PI fully meets the criteria for and is not debarred from GRF. Before giving the various confirmations required, the institution should satisfy itself that the PI is in all respects suitable to submit a project for RGC funding and complete the project, if funded. It is also required to report to the RGC immediately if a PI subsequently]

1. INSTITUTIONAL ENDORSEMENT

(* Please tick as appropriate)

I confirm that:

Staff eligibility requirement for GRF

- (a) the application has been evaluated and endorsed by the institution for submission to the RGC;
- (b) the PI*, being within the staff grade A to I, meets fully the stipulated staff eligibility requirement for GRF;

* where the PI is newly appointed, the institution has formally entered into a contract of service with him/her on or before 9 November 2009 and the contract requires him/her to report duty on or before 1 May 2010.

- (c) * the PI is/will be employed on permanent term
- * the PI is/will be employed on fixed term contract #

if the PI is/will be employed on a fixed term contract, the PI will still be eligible for an GRF grant at the time of funding award being made in June of next year and for at least the first year of the project's planned duration;

- (d) *# the PI is/will be a visiting scholar
- * the PI is NOT a visiting scholar

#where the PI is a visiting scholar, he/she has a full-time employment with the institution covering at least one year or the expected duration of the project whichever is the longer;

- (e) the institution will inform the RGC as soon as the PI ceases to be eligible to apply, receive or hold an GRF grant, and will withdraw the application, or if funded by the RGC, terminate/conclude the project as appropriate;

- (f) the institution understands that the GRF grant, if given, will be withdrawn if the project does not start within one year of the funding award or the PI leaves the UGC sector within six months of project commencement. The institution should report to the RGC as soon as possible when a PI proceeds on no-pay leave/professional leave for a continuous or cumulative period exceeding 183 days within the project period;

Longer-term research grant (only for the case where PI has selected this item at Part I 1(c)(ii))

- (g) the PI

- * meets
 * does not meet

the eligibility requirement for longer-term research grant as set out in the Supplementary Notes for Applicants of GRF for Longer Term Research;

Individual research (only for the case where PI has selected this item at Part I 1(c)(ii))

- (h) the PI

- * meets
 * does not meet

the eligibility requirement for funding support of individual research as set out in the Supplementary Notes for Applicants of GRF for Individual Research;

- (i) the institution

- * will
 * will not

facilitate arrangements for time-off for applications for individual research;

Relief support under Humanities Sub-Panel (only for the case where PI has selected this item at Part I 1(c)(ii))

- (j) the salary for the relief teacher proposed by the PI

- * exceeds

* does not exceed

the salary of Staff Grade 'G'(i.e. Lecturer(u)) or equivalent as set out in the Supplementary Notes for Applicants of GRF for Research Support under Humanities Sub-Panel;

Institutional Commitments

Allowance for undergraduate student helper (only for the case where PI has selected this item at Part I 1(c)(ii))

(k) the institution is committed to providing a monthly allowance of \$1,250 to the undergraduate student helper up to a maximum period of ten months if this proposal is funded as set out in the Supplementary Notes for Applicants of GRF for Undergraduate Student Allowance;

(l) adequate supervision, research facilities and training provisions

* will

* will not

be in place to meet the need of RPg student(s) so employed under the research grant if this application is supported by the RGC.

* No RPg student will be trained in this proposed project;

* Not applicable

(m) the research project under this GRF application

* is

* is not

in line with the role of the institution.

(n) no equipment is required

the PI's declaration and reasons at Part II Section 3(b) are true and correct and I support the PI's request for this procurement of equipment

(o) no research-related software licence/dataset is required

- the PI's declaration and reasons at Part II Section 3(c) are true and correct and I support the PI's request for this procurement of research-related software licence/dataset
- (p) this application does not include requests for purchasing (i) personal electronic devices such as cell phones, iPod, MP3 Players, digital cameras and PDAs except with sound justifications and (ii) standard equipment(e.g. desktop PCs, servers, laptop computers, printers, scanners) and standard software licence/dataset (e.g. Windows, MS-Office) known to be available or reasonably expected to be provided in the institutions concerned.

Notes on (l):

The primary duty of the Principal Investigator of the RGC grant is to complete the project according to plan and that the training of RPg students should not be used to justify any delay of project completion nor unsatisfactory project performance.