

# IC AGSIS 09

## PROCEEDINGS

International Conference on  
Advanced Computer Science and Information Systems  
December 7<sup>th</sup>- 8<sup>th</sup>, 2009  
Universitas Indonesia, Jakarta - Indonesia

Organized by



Faculty of Computer Science  
Universitas Indonesia

ISSN : 2066-1796

# ICACISIS<sup>09</sup>

**Proceedings of the International Conference on Advanced Computer Science and  
Information Systems (ICACISIS 2009)**

Universitas Indonesia  
December 7<sup>th</sup>-8<sup>th</sup>, 2009

Editors :

Dr. Mirna Adriani  
R. Yugo K. Isal, M.Sc  
Dr. Wisnu Jatmiko  
Dr. Bob Hardian  
Dr. Petrus Mursanto  
Adila Alfa Krisnadhi, M.Sc  
Herry, M.Kom

Organized by



Faculty of Computer Science  
Universitas Indonesia

## Committee

### Honorary Chairs:

- Mohammad Nuh  
Minister of National Education, Republic of Indonesia
- Tifatul Sembiring  
Minister of Communication and Information, Republic of Indonesia
- Gumilar R. Somantri  
Rector of Universitas Indonesia, INA
- T. Basaruddin  
Dean of Faculty of Computer Science Universitas Indonesia, INA
- T. Fukuda  
Nagoya University, JPN
- T. Usagawa  
Kumamoto University, JPN
- S. Hölldobler  
Technische Universität Dresden, GER
- S. Bressan  
National University of Singapore, SIN
- K. Sekiyama  
Nagoya University, JPN
- M. Ohka  
Nagoya University, JPN

### General Chairs:

- Mirna Adriani, Universitas Indonesia, INA
- Yugo K. Isal, Universitas Indonesia, INA

### Program Chairs:

- W. Jatmiko, Universitas Indonesia, INA
- B. Hardian, Universitas Indonesia, INA

### Technical Program Committee:

- A. Azurat, Universitas Indonesia, INA
- A. Buono, Bogor Agricultural University, INA
- A. Djunaidi, Sepuluh November Institute of Tech., INA
- A. Muis, Universitas Indonesia, INA
- A. Murni, Universitas Indonesia, INA,
- A. N. Hidayanto, Universitas Indonesia, INA
- A. Naba, Brawijaya University, INA
- A. Purwarianti, Bandung Institute of Tech., INA
- A. S. Baskoro, Universitas Indonesia, INA
- A. S. Nugroho, BPPT, INA
- A. Sofyan, Bandung Institute of Tech., INA
- A. Zainal A., Sepuluh November Institute of Tech., INA
- B. A. Nazief, Universitas Indonesia, INA

- B. H. Widjaja, Universitas Indonesia, INA
- B. Yuwono, Universitas Indonesia, INA
- D. I. Sensuse, Universitas Indonesia, INA
- E. K. Budiardjo, Universitas Indonesia, INA
- G. Kiswanto, Universitas Indonesia, INA
- H. Budiarto, BPPT, INA
- H. M. Manurung, Universitas Indonesia, INA
- H. Manik, Bogor Agricultural University, INA
- H. Santoso, Nagaoka University of Technology, JPN
- H. Suhartanto, Universitas Indonesia, INA
- H. Yussof, Nagoya University, JPN
- I. Budi, Universitas Indonesia, INA
- I. Fanany, Tokyo Tech. JPN
- I. Meilano, Bandung Institute of Tech, INA
- I. Yusuf, University of Tanjungpura, INA
- I. P. Gunawan, Universitas Multimedia Nusantara, INA
- K. Watanabe, Saga University, JPN
- L. Y. Stefanus, Universitas Indonesia, INA
- Marimin, Bogor Agricultural University, INA
- M. Nakajima, Nagoya University, JPN
- M. R. Widyanto, Universitas Indonesia, INA
- M. Jamal, Bandung Institute of Tech., INA
- Nurhasan, Bandung Institute of Tech., INA
- P. Mursanto, Universitas Indonesia, INA
- R. Budiarto, University Sains Malaysia, MLY
- S. Kuswadi, Sepuluh November Institute of Tech., INA
- S. Yazid, Universitas Indonesia, INA
- T. A. Masoem, Universitas Indonesia, INA
- W. C. Wibowo, Universitas Indonesia, INA
- W. S. Nugroho, Universitas Indonesia, INA
- W. Prasetya, Utrecht University, Netherlands
- Y. G. Sucahyo, Universitas Indonesia, INA
- Y. Hasegawa, Tsukuba University, JPN
- Z. A. Hasibuan, Universitas Indonesia, INA
- Z. Zyada, Tanta University, Egypt

## Table of Contents

Welcome Message from General Chairs	i
Welcome Message from Dean of Faculty of Computer Science Universitas Indonesia	ii
Welcome Message from Rector of Universitas Indonesia Committee	iii
Table of Content	v
	vii

### Plenary Lectures

Cognitive Science, Computational Logic and Connectionism <i>Prof. Steffen Hölldobler, Technische Universität Dresden, Germany</i>	1
Binaural Hearing Model and its Applications <i>Prof. Tsuyoshi Usagawa, Kumamoto University, Japan</i>	7
Self-Organizing Control for Distributed Autonomous Systems <i>Prof. Kosuke Sekiyama, Nagoya University, Japan</i>	11
A Tutorial on Sampling Data Streams: From Reservoir Sampling to Sliding Window Sampling <i>Prof. Stéphane Bressan, National University of Singapore, Singapore</i>	13
Recent Topics on Robotic Tactile Sensors <i>Prof. Masahiro Ohka, Nagoya University University, Japan</i>	17

### Image Processing

Motion Blur Identification <i>Irwan P. Gunawan, Erik Kalalembang, and Koredianto Usman</i>	23
Color Segmentation for Extracting Symbols and Characters of Road Sign Images <i>Aryuanto, Koichi Yamada, and F. Yudi Limpraptono</i>	29
Lungs Segmentation by Developing Binary Mask <i>Saleem Iqbal and Amir Hanif Dar</i>	34

Performance Evaluation of Pap Smear Cell Image Classification Using Quantitative and Qualitative Features Based on Multiple Classifiers	40
<i>Dwiza Riana and Aniati Murni</i>	
Color Image Restoration using Robust Neuro-Fuzzy Network Approach	45
<i>Rully Soelaiman, A. Dista Satria, and Yudhi Purwananto</i>	
Determining Surface Roughness Level Based on Texture Analysis	51
<i>Dedy Septiadi and Aulia MT. Nasution</i>	
Application of Particle Swarm Optimization for Edge Detection of Molten Pool in Aluminum Pipe Welding	57
<i>Ario Sunar Baskoro, Rui Masuda, and Yasuo Suga</i>	
Batik Image Reconstruction Based on Codebook and Keyblock Framework	64
<i>Wahyudi, Ade Azurat, Ruli Manurung, and Aniati Murni</i>	
<b>Image Retrieval</b>	
Segmentation Based Retrieval on General Images	69
<i>Suhendro Y. Irianto and Sri Lestari</i>	
Image Searching using Heuristic Method for Image Retrieval System	74
<i>Yeni Herdiyeni and Fitria Yumingsih</i>	
Recognition of Batik Motifs Using the Generalized Hough Transform	79
<i>Hadaiq R. Sanabila and Ruli Manurung</i>	
Clustering Batik Images Based on Log-Gabor and Colour Histogram Features	85
<i>Laksmi Rahadanti, Ruli Manurung, and Aniati Murni</i>	
<b>Information Retrieval</b>	
Random Forests Approach for Bibliographic Extraction	91
<i>Cho Cho Khaing</i>	
Implementation of Citation Analysis in Indonesian Digital Library	95
<i>Stevanus Djojokusumo, Arya Tandy, and Gunawan</i>	
Finding Structured and Unstructured Features to Improve the Search Result of Complex Question	100
<i>Dewi Wisnu Wardani</i>	

Developing Indonesian-English Speech-to-Speech Translation System	106
<i>Darwin Cuputra and Mirna Adriani</i>	
Spoken Document Retrieval Using Statistical Methods	111
<i>Armando Yonathan and Mirna Adriani</i>	
Strong's Concordance Formulation for Indonesian New Testament Bible	114
<i>Gunawan, Devi Dwi Purwanto, Herman Budianto, and Indra Maryati</i>	
Implementation of Indonesian Automated Speech Recognition for OOV Detection	120
<i>Aswin Juari and Ayu Purwarianti</i>	
Automated Essay Grading System Using SVM and LSA for Essay Answers in Indonesian	124
<i>Rama Adhitia and Ayu Purwarianti</i>	
Retrieving Opinion in Indonesian Blogs	130
<i>Clara Vania and Mirna Adriani</i>	
Pattern Based Approach in Indonesian Question-Answering System	134
<i>Hapnes Toba and Mirna Adriani</i>	
<b>Software Engineering</b>	
Modeling and Analysis of Combined Method for Software Aging	139
<i>May Tar Hla Myint</i>	
Application of Dominant-based Rough Set Approach for Defining Relative Quality of Object Oriented Software	144
<i>Petrus Mursanto, and Ardhi Tomiarfi</i>	
Service Oriented Architecture Reference Architecture Blueprint	149
<i>Edison, Virginia Tulenan, and Ford Lumban Gaol</i>	
The Analysis of UML to SystemC Model Transformation on Embedded System Design	155
<i>Maman Abdurohman, Kuspriyanto, Sarwono Sutikno, and Arif Sasongko</i>	
Basic Design of BAIK - Scripting Language with Indonesian Lexical Parser for Internet based Software Development	160
<i>Haris Hasamudin</i>	
The Structure of Software Requirement Specification Patterns: UML Based Template	166
<i>Eko K. Budiardjo</i>	

Developing Indonesian-English Speech-to-Speech Translation System	106
<i>Darwin Cuputra and Mirna Adriani</i>	
Spoken Document Retrieval Using Statistical Methods	111
<i>Armando Yonathan and Mirna Adriani</i>	
Strong's Concordance Formulation for Indonesian New Testament Bible	114
<i>Gunawan, Devi Dwi Purwanto, Herman Budianto, and Indra Maryati</i>	
Implementation of Indonesian Automated Speech Recognition for OOV Detection	120
<i>Aswin Juari and Ayu Purwarianti</i>	
Automated Essay Grading System Using SVM and LSA for Essay Answers in Indonesian	124
<i>Rama Adhitia and Ayu Purwarianti</i>	
Retrieving Opinion in Indonesian Blogs	130
<i>Clara Vania and Mirna Adriani</i>	
Pattern Based Approach in Indonesian Question-Answering System	134
<i>Hapnes Toba and Mirna Adriani</i>	
<b>Software Engineering</b>	
Modeling and Analysis of Combined Method for Software Aging	139
<i>May Tar Hla Myint</i>	
Application of Dominant-based Rough Set Approach for Defining Relative Quality of Object Oriented Software	144
<i>Petrus Mursanto, and Ardhi Tomiarfi</i>	
Service Oriented Architecture Reference Architecture Blueprint	149
<i>Edison, Virginia Tulenan, and Ford Lumban Gaol</i>	
The Analysis of UML to SystemC Model Transformation on Embedded System Design	155
<i>Maman Abdurohman, Kuspriyanto, Sarwono Sutikno, and Arif Sasongko</i>	
Basic Design of BAIK - Scripting Language with Indonesian Lexical Parser for Internet based Software Development	160
<i>Haris Hasanudin</i>	
The Structure of Software Requirement Specification Patterns: UML Based Template	166
<i>Eko K. Budiardjo</i>	



Towards Risk-Oriented Security Ontology	170
<i>Basuki Rahmad, Suhono H. Supangkat, Jaka Sembiring, and Kridanto Surendro</i>	
<b>IT Governance</b>	
Interoperability Model for Public Service Information Systems	175
<i>Didi Sukyadi and Dana I. Sensuse</i>	
Alternative E-Payment B2C (Business to Consumer) Models for Indonesian Community	181
<i>Agung Firmansyah, Muhammad Ilman Akbar, Mursal Rais, Mustafa Kamal, and Putu Wuri H</i>	
Implementing a Modern Process Management System Based on the MIT Process Handbook	188
<i>Putu Wuri Handayani, Stéphane Bressan, and Omar Boucelma</i>	
Measuring the Effectiveness of a Simplified COBIT-based IT Process Maturity Assessment Method	198
<i>Budi Yuwono, Muhammad Nasri, and Rein Nusa Triputra</i>	
Information Technology Plan as an IT Governance Maturity Driver	204
<i>Budi Yuwono, Rein Nusa Triputra, and Muhammad Nasri</i>	
<b>Biomedical Applications</b>	
Multi-agent Based Medical Diagnosis System	209
<i>Ei Ei Chaw</i>	
Handling Missing Data in Patient Data Preprocessing using Information Visualization	215
<i>Hasimah Hj Mohamed, Abdul Razak Hamdan, and Azuraliza Abu Bakar</i>	
Discovering Sequential Disease Patterns in Medical Databases Using FreeSpan Mining Approach	220
<i>Oviliani Yenty Yuliana, Silvia Rostianingsih, and Gregorius Satia Budhi</i>	

ABCD Feature Extraction for Melanoma Skin Cancer Diagnosis <i>Bilqis Amalian, Chastine Fatichah, and M. Rahmat Widyanto</i>	225
Application of Non-rigid Image Registration to the Diagnosis of Herniated Nucleus Pulposus <i>Iping Supriana and Tahir Arazi</i>	230
Survey on Epitope Prediction Method on DNA Vaccine Development <i>B. Solihah, M. Rahmat Widyanto, and Asmarinah</i>	236
Microfabrication of Amperometric Glucose Sensor <i>Aminuddin Debataraaja, Nur Fauzi Soelaiman, and Latif Mawardi</i>	242
Image Thresholding on Segmentation of Teeth in Dental Panoramic Radiographs <i>Agus Zainal Arifin and Dini Adni Navastara</i>	248
 <b>Signal Processing</b>	
An Initial Energy Detection Utilizing Software Defined Radio <i>Marwanto, M. Adib Sarijari, N. Fisal, Sharifah K. Syed-Yusof, and Rozeha A. Rashid</i>	255
Blind Sound Separation Using Frequency-Domain and Time-Domain Independent Component Analysis for Machines Fault Detection <i>B.T. Atmaja and D.Arifianto</i>	259
Underwater Acoustical Remote Sensing Behaviour of Marine Biota <i>Henry M. Manik</i>	264
Controlling Mouse Cursor using Head Movement <i>Gunawan, FX. Ferdinandus, Tri Kurniawan Wijaya, Indra Maryati, and Edwin Seno Dwihapsoro</i>	267
 <b>Computer Networks, Computer Architectures, and High Performance Computing</b>	
Exploiting IPv6 Extension Header to Handle Transmission Error for IPv6 Packets over High Speed Networks <i>Supriyanto, Iznan H. Hasbullah, and Rahmat Budiarto</i>	273

Seamless and Secure Mobile IPTV Binding Update using Mobile IPv6	279
<i>Armanda Caesario Cornelis, Primantara Hari Trisnawan, and Rahmat Budiarto</i>	
SF-STEP: Securing Tunnel Endpoints using Separate Filtering IPv4 and IPv6	283
<i>Abidah M. Taib and Rahmat Budiarto</i>	
The Channel Design for TLM Implementation on Embedded System Design	290
<i>Maman Abdurohman, Kuspriyanto, Sarwono Sutikno, and Arif Sasongko</i>	
Reduction of the Bandwidth Request Delay in IEEE 802.16j MMR WiMAX Networks	295
<i>Fath Elrahman Ismael, Sharifah K. Syed-Yuof, and Norsheila Faisal</i>	
Light Weight Resource-Aware Data Stream Classification	300
<i>Ary Mazharuddin Shiddiqi and Mohammed Medhat Gaber</i>	
Performance of iSCSI Protocol on Virtual Disk in Ethernet LAN	306
<i>Swee Liang Aw and Wafaa Alsalihi</i>	
Performance Evaluation of Single Bus Microprocessor Architecture	312
<i>Petrus Mursanto, Andreas Febrian, Wisnu Jatmiko, and Adhi Yuniarto</i>	
Accelerating Phase Only Correlation For Motion Estimation With Full Search Technique Using Parallel Threading in GPU	318
<i>Rosa A. Asmara, Cahya Rahmad, and Anik N. Handayani</i>	
Self PID Tuning Rule Using Grey Prediction Algorithm	323
<i>Ala Eldin Abdallah Awouda and Rosbi Bin Mamat</i>	
Simulation and Analysis of Optimum Energy Flow for an Off-Grid Hybrid Power System	329
<i>Ismail Yusuf</i>	
<b>Pattern Recognition</b>	
Feature Extraction Method on Microarray Data Classification Based on Information Gene Pairs	333
<i>Rully Soelaiman, Sheila Agustianty, Yudhi Purwananto, and I.K. Eddy Purnama</i>	
Feature Based Milling Direction of a Faceted Model Based on Intelligent Fuzzy 3D Feature Identification	341
<i>Gandjar Kiswanto, Rahmat Widyanto, and Priadhana Edi Kresnha</i>	

---

Parallel Supervised PNN Structure Determination Algorithm and its implementation using Message Passing Interface 352

*Herry and Heru Suhartanto*

Potential Use of Machine Vision in Texture Recognition of CNT/Polymer Microelectrode for Conductivity Measurement 356

*Yudan Whulanza and Gandjar Kiswanto*

Optimized Fuzzy Learning Vector Quantization for Artificial Mixture Odors Discrimination System 360

*Hadaig R. Sanabila, Rochmatullah, and Wisnu Jatmiko*

## **Robotics**

Development of Particle Swarm Fuzzy Controller for Behavior-based Mobile Robot 366

*Andi Adriansyah*

Proposition of an Adaptive Controller Dominant Type Hybrid Adaptive and Learning Controller for Robot Manipulators 373

*Munadi and Tomohide Naniwa*

Gradual Pattern Formation of Homogeneous Robot Groups Using Simple Periodic Functions: Case Study Mindstorms NXT Lego Robots 378

*Nulad W. Pambudi, Adila A. Krisnadhi, Wisnu Jatmiko, and Rizki Mardian*

An Application of Distributed Multihop Robot-based Communication Network in Restricted and Static Environment after Disaster 384

*Rizki Mardian, Wisnu Jatmiko, Adila A. Krisnadhi, and Nulad W. Pambudi*

## **Algorithms**

The Superiority of Fibonacci Sequence as Partition Sizes in Ranking Stage in Word-based Block Sorting Text Compression 391

*R. Yugo Kartono Isal*

Complexity of Odd Even Transposition Parallel Algorithm in Nonlinear Network Models 397

*Ernastuti and Ravi A. Salim*

Sequential and Parallel Fractal Compression Algorithms for Image Compression 403

*Satrya N. Ardhytia and Lely Hiryanto*

Solving University Timetabling as a Constraint Satisfaction Problem with Genetic Algorithms <i>Teddy Wijaya and Ruli Manurung</i>	409
<b>e-Learning</b>	
Applying UTAUT Theory to Reach Better Understanding on The Acceptance and Use of Learning Management System Case Study: Experiential E-Learning of Sanata Dharma University <i>I Gusti Nyoman Sedana and St. Wisnu Wijaya</i>	415
Developing Web-based Communication Media for Learning <i>Widyo Nugroho, Ichwan Suyudi, Fikri Saleh, and Taufik Hidayat</i>	421
Design and Implementation of Learning Object Ontology for e-Learning Personalization <i>Meirna A. Ramadhani, Siti Aminah, A. Nizar Hidayanto, Adila A. Krisnadhi</i>	427
Development of Student Model Ontology for Personalization in an e-Learning System based on Semantic Web <i>Leonny Pramitasari, A. Nizar Hidayanto, Siti Aminah, Adila A. Krisnadhi, Meirna A. Ramadhani</i>	434
Determining Mental States of Students Using a Soft-Computing Approach <i>Handri Santoso, Shusaku Nomura, and Kazuo Nakamura</i>	440
Author's Index	445

# Pattern Based Indonesian Question Answering System

Hapnes Toba and Mirna Adriani  
*Information Retrieval Laboratory*  
*Faculty of Computer Science, University of Indonesia*  
Email: hapnes.toba@ui.ac.id, mirna@cs.ui.ac.id

**Abstract**—This paper describes a pattern based approach to Indonesian question answering system using the Open Ephyra. In this study, we classify the factoid questions types into 8 categories, where each group is trained using specific questions. The results demonstrate the potential of the approach in an automated Indonesian question answering system.

## I. INTRODUCTION

Question answering is a form of information retrieval that concern about an exact answer from a given natural language question rather than a query string. An automated question answering system (QAS) tries to retrieve explicit answers in the form of a single answer or snippets of text rather than a whole document or set of documents. One of the biggest challenges in QAS is how to categorize a question into a particular category that further will be used to find exact answer(s) within a large collection of documents. Research in QAS is initialized during the 70's when [25] developed a system that can recognize natural language in SHRDLU. Such system is further developed by [15] in QUALM, that can recognized stories. The re-initiation of QAS is started in the 90's when the Internet has become a redundant source of information [3].

The results of two main evaluation forum in QAS, i.e. TREC (Text Retrieval Conference, <http://trec.nist.gov>) from 1999-2007, and CLEF (Cross Language Evaluation Forum, <http://www.clef-campaign.org/>) from 2002-2008, have shown that question-answering is still need further enhancements. The question-answering techniques that exists from the field of natural language (NLP), information retrieval (IR) and combination of both, are promising, but there is still lack of standardization in methods, techniques and evaluation [8]. According to [3], there is no ultimate QAS. Each approach has its own niche, application environment, and tasks. If the quality of the answers is crucial, NLP should be applied. If facts need to be extracted from text, IR techniques should be used. The main techniques that have been mostly used in QAS research are: semantic analysis using semantic role labeling, name entity recognizer, path dependency [21], semantic markup [16], n-gram

passages [4, 5], statistical [11][17], and the combination of semantic structures and probabilistic approach [18].

A common feature of NLP-based QAS is the ability to convert text input into formal representation of meaning such as logic (first order predicate calculus), semantic networks, conceptual dependency diagrams, or frame-based representations [13]. IR-based QAS are usually completed with shallow or deep NLP techniques, focuses on fact retrieval from a large text corpus. Document redundancy, i.e. a number of similar statements that contain the answer, in a large corpus and the use of shallow NLP techniques increase the chance of finding the right answer without any guarantee that the answer is correct. Shallow NLP techniques, combined with statistical methods, pattern learning, and passage retrieval, have been largely used in the extraction of definition answers in TREC and CLEF, as described in [20], [23], [6], and [22]. Future research trend should involve the combination of the approaches into one single system and adapt the techniques to the application domain.

In the Information Retrieval Lab at the University of Indonesia there have been some researches done to initiate the Indonesian QAS and make some contribution in CLEF [2, 24]. Further research approaches can be grouped in three mainstreams, i.e. the semantic analysis [14, 12], statistical approach [1], and the combination of both [9].

In this experiment, we try to adapt the categorization approach within the OE pipeline that used pattern learning approach for Indonesian language. Pattern learning approach is a form of rule-based approach for question categorization. Another approaches that mainly use are language modeling and machine learning based [19]. The final goal of this experiment is to investigate how to fit Indonesian interrogative sentences in an English based QAS, as a preliminary study to develop a QAS that proper to Indonesian language that combines the NLP and statistical approaches.

The rest of the paper will be organized as follow: section 2 presents the state-of-the art of Indonesian question categorization and the pattern learning approach that is used in Open Ephyra. Section 3 presents the strategy that is used to develop question patterns from Indonesian interrogative sentences. Section 4 gives the experimental setup and results. Section 5 gives the evaluation of the results, and section 6 is a conclusion and plans for future works.

## II. QUESTION CATEGORIZATION

The study of Purwarianti, et. al. 2006 [19], has shown that question categorization for Indonesian interrogative sentence using shallow parser and machine learning can achieved 95% accuracy. However, such learning categorization requires deep analysis in sentence structures to be able to develop a robust parser and extraction of learning features. The problem in such deep analysis for Indonesian language is the limitation of the resources of the language itself, that lack of robust parser.

Another possibility for simpler approach to question categorization is by matching the pattern of each question type, and tries to categorize question based on the position of question words, and various question keywords. To make sure that the categorization can be done, we need thus to develop a number of pattern rules that reflect the structure of each question type.

OE<sup>1</sup> is an example of QAS that use pattern learning approach to categorize questions. It learns text patterns that can be applied to text passages for answer extraction. OE can learn question-answer pairs and use common retrieval system, i.e. web search engine or various IR system, to fetch document text. OE is an open domain QAS that has modular and extensible framework. It consists of four main modules (see Figure 1): 1) Question analyzer, 2) Query generator; 3) Search engine and 4) Answer extractor. Each of the modules can be used independently and thus suitable to experiment multiple approaches to question-answering in one system.

There are two main steps for the pattern-learning approach in OE. The first is to learn the question patterns from question templates according to each question types. The aim of this step is to interpret the questions and transform them into queries. The second step is to learn the answer pattern from question-answer pairs. The aim of this second step is to extract answers candidate from relevant document snippets and to rank them. The question templates need to be manually developed according to various interrogative sentences that are independent for each natural language.

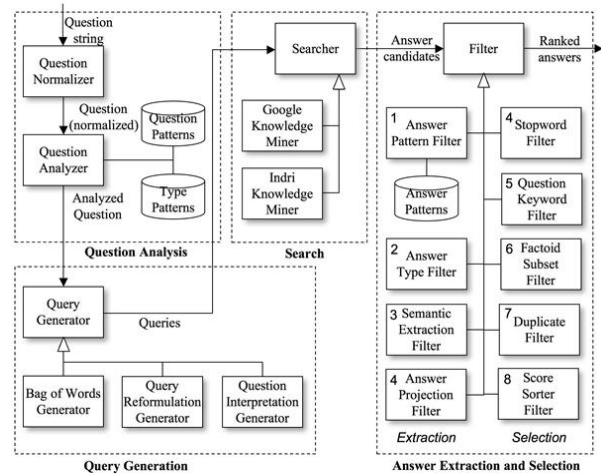


Fig. 1. Open Ephyra Framework (Schlaefter et.al., 2006)

An important strategy in this approach is to determine the objects interpretation from each question, i.e.: the property (PO), target (TO) and context (CO). For example, a question “*Dimana letak kampus UI?*” (*Where lies the campus of UI?*). It has the interpretation of:

- Property: TEMPAT (location)
- Target: kampus (campus)
- Context: UI (University of Indonesia)

The question asks for a property of the target object “kampus”, which is a location. The context object “UI”, narrows down the search to a particular place. Together, all of this objects form the interpretation of the question that further will also be used in developing a query for the search engine.

OE interprets a question by sequentially matching all the question patterns for each property. If it matches a pattern, OE will extract the target object and the context object. Therefore, it is possible that a question has more than one interpretation, if the question type is not clear enough.

Each of the properties is associated with a set of answer pattern that is assessed and extracted during the second learning phase. The format of an answer pattern is similar to the question patterns, it consists of:

- a target tag <T>
- a number of random number of context tag <C>
- a property tag <P>.

During the answer extraction phase, OE replaces all occurrences of target or context objects in a text snippet. Every time a snippet matches a pattern, the part of the text that associated to the property tag is extracted. In the example above, the following snippet from a relevant document:

<sup>1</sup> <http://www.ephyra.info>

“Bus kampus disediakan untuk melayani kebutuhan transportasi mahasiswa di dalam kampus UI Depok”

will be transform into

“Bus kampus disediakan untuk melayani kebutuhan transportasi mahasiswa di dalam <TO> <CO> Depok”.

The pattern “di dalam <TO> <CO> Depok”, will be used to extract the property *Depok*. The pattern has been learned from the question-answer pairs which is an answer that needs to be found.

To extract the answer, OE apply two kinds of regular expression (Zhang, 2002.):

- 1) that covers a target tag <T>, a property <P> and any characters in between them, i.e. the context; or or
- 2) covers one word or any characters preceding or following the <P> tag.

Finally OE will assess the answer patterns and assign a *confidence* score to the patterns. A confidence score is calculated by taking the proportion on how often a pattern could be used to extract a correct answer and the sum of how often it extracts a wrong answer and a right answer. Further, for each extracted property, the total number of snippets that has been assessed is also recorded, to compute the *support*, i.e. the ratio between correct answer and number of snippets. The answer that has the highest confidence and its value is above the support threshold will be considered as the best answer candidate.

### III. QUESTION PATTERN DEVELOPMENT

The properties and the question patterns is determined by analyzing 8 types of factoid question, each consists of 5 training questions and 25 testing questions. The question types are coarse grain according to CLEF (Forner, et. al. 2008.), i.e.:

- *ORANG* (people)
- *WAKTU* (time)
- *TEMPAT* (location)
- *ORGANISASI* (organization)
- *UKURAN* (measure)
- *ANGKA* (count)
- *OBJEK* (object)
- *LAIN-LAIN* (others).

For each question type, a question pattern is developed which reflects the variations arise in natural language writings or conversations that occur in the training question. The testing question is used to evaluate the accuracy of the pattern recognizer.

During the design of each question pattern, the following steps are taking into account:

- 1.The main question word(s) indicated for each type.
- 2.The position in which keywords of question occurs in an interrogative sentence, that indicates the context and target of a question.
- 3.The alternative of question words or phrases that indicate a special meaning to a question type.

As an example, these are some question patterns for the property location (= where) in Indonesian:

1. Main question word(s):  
(dimana|dimanakah) letak <TO> <CO>  
(dimana|dimanakah) <CO> <TO>  
(dimana|dimanakah) <TO> berada
2. Position of keywords (context and target):  
<TO> ada dimana (saja)?  
<TO> <PO> ada dimana  
dimana (saja)? (kah|sih)? <TO>  
berada  
<TO> terletak <CO> (apa)?  
<TO> <CO> (ada)? dimana (saja)?  
<CO> <TO> (ada)? dimana (saja)?  
<TO> <CO> ada dimana?
3. Alternative question words or phrases:  
(apa|apakah) nama <TO> <CO>  
(apa|apakah) nama  
(daerah|tempat|lokasi) <TO> <CO>  
(apa|apakah) nama  
(daerah|tempat|lokasi) <CO> <TO>  
di (lokasi|daerah|tempat) mana  
(saja)? <TO> <CO>  
di (lokasi|daerah|tempat) mana  
(saja)? <CO> <TO>  
terletak dimana (saja)? <TO> <CO>

During the answer patterns extraction, i.e. after the interpretation has been done, a tuple that consists of: a target, an arbitrary number context objects, and the answer or the property, is generated as query string. The query string will be used by IR system to retrieve the passage that match the pattern. It is possible to selectively add tuples for properties that are not sufficiently covered by the training question in the future to raise the accuracy of the system.

For our example in the previous section, the query string that will be generated is: “#1(kampus) #1(UI) #1(depok)”. The answer is included in the query string, this is to ensure that the snippet contain the target and the property.

### IV. EXPERIMENTAL SETUP DAN RESULTS

The document collection used in the experiment consists of 5 web documents (news articles, wiki's, and blog) on special issues from different domains. The issues ranges from historical occasions, biography, political news, myths and poetry. Every document that downloaded has positive relevant judgment to a specific question type, i.e. each document has snippets that contains answer(s) for the question given for each type. There are in total 190



documents in 6 MB text. The documents are processed by Perl programming language to form TREC format document, that further indexed using Indri. The index file is 2 MB big, and used as the local corpus for the system.

For the training phase, we prepare 40 questions, i.e. 5 questions for a question type and 1 question per subject. Each training question is has its answer pair in TREC format, that will be used in the pattern extraction phase. An example of the an answer extraction result for location type, is as follow:

```
<TO> [^<]*?<PO_NEtempat>
#correct: 5
#incorrect: 6
Confidence = 0.45454547
Support = 0.035311
```

This mean that this pattern can be found in 11 passages (not necessary in 11 documents), with 5 passages contains correct answer and 6 incorrect answer, and the total number of passages extracted from the all 8 properties is 142 = (5/0.035311).

After the training is done, we run the OE to test how accurate the result of training session. First we run the same question as the training phase, and obtain the following results:

Property	Interpretation				Answers			
	Single		More		Incorrect (W)	Unsupported (U)	Inexact (X)	Correct (R)
	Correct	Wrong	Correct	Wrong				
People	0	0	5	0	4	0	0	1
Time	5	0	0	0	3	0	0	2
Location	5	0	0	0	2	0	0	3
Organization	2	1	1	1	2	0	0	3
Measure	4	0	1	0	3	0	0	2
Number	3	0	2	0	4	0	0	1
Object	2	1	2	0	2	1	0	2
Other	4	0	0	1	5	0	0	0
<b>Total</b>	<b>25</b>	<b>2</b>	<b>11</b>	<b>2</b>	<b>25</b>	<b>1</b>	<b>0</b>	<b>14</b>

We also test the learning result using the testing questions, that consists of 5 variations for each issues, except for the ‘other’ type, thus in total 180 questions. The results for the testing are:

Property	Interpretation				Answers			
	Single		More		Incorrect (W)	Unsupported (U)	Inexact (X)	Correct (R)
	Correct	Wrong	Correct	Wrong				
People	6	6	11	2	23	0	0	2
Time	14	4	7	0	21	0	0	4
Location	17	7	0	1	14	0	1	10
Organization	5	13	2	5	21	0	0	4
Measure	7	7	10	1	20	0	2	3
Number	10	3	12	0	21	0	2	2
Object	5	18	1	1	21	1	1	2
Other	1	1	3	0	5	0	0	0
<b>Total</b>	<b>65</b>	<b>59</b>	<b>46</b>	<b>10</b>	<b>146</b>	<b>1</b>	<b>6</b>	<b>27</b>

In the interpretation column, we separate the interpretation for “single” interpretation, i.e.: exact only one interpretation for a question; and a “more” interpretation, means that a question can be interpreted into one or more category, according to the patterns.

The accuracy is calculated for the interpretation results and the answers. We use the accuracy definition from CLEF. 2008., i.e. the average of SCORE(q) over all 200 questions q, where SCORE(q) is 1, if the answer to q assessed as correct, and 0 otherwise.

The accuracy for the tested-training questions, and the testing questions is shown in the following table:

Tested-training questions	Interpretation Accuracy	90.00
	Answers Accuracy	35.00
Testing questions	Interpretation Accuracy	61.67
	Answers Accuracy	15.00

## V. EVALUATION

The accuracy for the tested-training questions is much higher than the testing questions. This result suggests that the patterns used during the training session are not good enough to cover the question variations. Further we have to give extra attention to informality of natural language that is used to write document, especially for blogs and wiki’s documents. This experiment is more concern in the interpretation of questions. If we see only the interpretation, than the results is promising. Although it seems that we need to cover more question variations.

The most difficult part to interpret question patterns is for people and organization types. In Indonesian, both can be asked using the “siapa (= who)” question word, and thus gives double interpretation to a specific question, that effects on the confidence score, i.e. give lower score, because less correct answer is extracted.

The object type is also hard to interpret, because the similar question pattern for this type can be occur in another type(s).

## VI. CONCLUSION AND FUTURE WORKS

In this experiment, we have adapted the question pattern approach in OE. The result shows that OE is promising to be adapted for Indonesian language. The main shortcoming of the pattern learning approach is that question patterns need to be developed specifically and the answer extraction phase needs redundant sources, i.e. large search space.

There are a number of things need to be further investigated, such as:

1. How to generate a more generic pattern that can be used to interpret question accurately.
2. How to decrease the runtime during the learning steps, i.e. how to prune the unnecessary pattern to be learned more than once.

3. How to deal with ambiguity in question words and keyword phrases that can be occurred in more than one question types.
4. How to give sense of contextuality during the answer selection phase. For example how can we deal with a time frame, and thus if the question is about the president in present time, it returns the correct answer, and not the president in the past time, although both answers are can be found as patterns in the relevant documents.

Based on the shortcoming, adaptation of OE into Indonesian language need the following course:

1. Develop more fine grained question types that each represent special name-entity (NE) type. For example, for type "location", can be more precisely defined as: university, country, etc.
2. Modify or change the natural language specific components, such as: NE tagger, stemmer, phrase chunker, part-of-speech (POS) tagger, and tokenizer.
3. Develop a statistical/machine learning question classifier that can be used to categorize questions based on their features, such as: unigram, bigram, or the question word.

#### REFERENCES

- [1] Adiwibowo, Septian. 2008. Answer Finding in Indonesian-English Question Answering System Using Word Weighting and Information from the Internet. Bachelor Thesis Faculty of Computer Science University of Indonesia.
- [2] Adriani, Mirna. et. al. 2005. University of Indonesia Participation at CLIR - CLEF 2005.
- [3] Andrenucci, Andrea & Sneider, Eriks. 2005. Automated Question Answering: Review of the Main Approaches. Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05) IEEE Computer Society.
- [4] Buscaldi, Davide, et. al. 2006. N-gram vs. Keyword-based Passage Retrieval for Question Answering. Proceedings of Cross Language Evaluation Forum (QA@CLEF 2006).
- [5] Buscaldi, Davide, et. al. 2009. Answering Questions with an n-Gram Based Passage Retrieval Engine. Journal of Intelligent Information System. Springer Netherlands. Published Online 13 March 2009. DOI 10.1007/s10844-009-0082-y.
- [6] Clarke, C. et al., 2001. Web reinforced Question Answering, in D. Harman and E. Voorhees (eds): Proc of TREC 2001, NIST, Gaithersburg, USA.
- [7] CLEF. 2008. GUIDELINES for the PARTICIPANTS in QA@CLEF 2008.
- [8] Ferrucci, et. al. 2009. IBM Research Report: Towards the Open Advancement of Question Answering System. RC24789 (W0904-093) April 22nd.
- [9] Fitria, Lily. 2007. Answer Finding using Sentence Dependency Structure in Question Answering System. Master Thesis Faculty of Computer Science University of Indonesia.
- [10] Forner, Pamela, et. al. 2008. OVERVIEW OF THE CLEF 2008 MULTILINGUAL QUESTION ANSWERING TRACK.
- [11] Ittycheriah, Abraham, et. al. 2001. IBM's Statistical Question Answering System. Proceedings of the 10th Text Retrieval Conference (TREC 2001).
- [12] Jati, Ranga. M. 2009. The Development of Natural Language Generation with Chart Generation Approach and Its Application in Indonesian Question Answering. Bachelor Thesis Faculty of Computer Science University of Indonesia.
- [13] Jurafsky, D and Martin, J.H. 2000. Speech and language processing, Prentice Hall, NJ, USA.
- [14] Larasati, Dian S. and Manurung, Ruli. 2007. Towards a Semantic Analysis of Bahasa Indonesia for Question Answering. Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics (PACLIC).
- [15] Lehnert, Wendy G. 1977. The Process of Question Answering. Dissertation Thesis. Yale University
- [16] Lopez, Vanessa, et. al. 2005. AquaLog: An Ontology-Portable Question Answering System for the Semantic Web. ESWC (European Semantic Web Conference 2005), LNCS (Lecture Notes on Computer Science) 3532, pp. 546 – 562. Springer-Verlag Berlin Heidelberg.
- [17] Metzler, Donald dan Croft. W. Bruce. 2004. Analysis of Statistical Question Classification for Fact-based Questions. Kluwer Academic Publisher.
- [18] Narayanan, Srindi & Harbagiu, Sanda. 2004. Question Answering Based on Semantic Structures.
- [19] Purwarianti, et. al. 2006. Estimation of Question Types for Indonesian Question Sentence. Department of Information and Computer Sciences, Toyohashi University of Technology.
- [20] Ravichandran, D. and Hovy, E.H. 2002. Learning Surface Text Patterns for a Question Answering System", Proc. of ACL-2002, ACL press, USA.
- [21] Schlaefler, Nico. 2007. Deploying Semantic Resources for Open Domain Question Answering. Diploma Thesis. Language Technologies Institute School of Computer Science Carnegie Mellon University.
- [22] Schlaefler, Nico, et. al. 2006. A Pattern Learning Approach to Question Answering within the Ephyra Framework. LNAI 4188, pp. 687–694, Springer-Verlag Berlin Heidelberg.
- [23] Soubbotin, et.al. 2002. Use of Patterns for Detection of Answer Strings: A Systematic Approach, in L.P. Buckland and E.Voorhees (eds): Proc. of TREC 2002, NIST, Gaithersburg, USA.
- [24] Wijono, Sri Hartati, et. al. 2006. Finding Answers to Indonesian Questions from English Documents. Proceedings of Cross Language Evaluation Forum (QA@CLEF 2006).
- [25] Winograd, Terry. 1971. Procedures as a Representation for Data in a Computer Program form Understanding Natural Language. Dissertation Thesis. Massachusetts Institute of Technology.
- [26] Zhang, D., Lee, W. 2002. Web Based Pattern Mining and Matching Approach to Question Answering. Proceedings of the 11th Text REtrieval Conference (TREC).