

INKOM


Jurnal Informatika, Sistem Kendali dan Komputer

Volume 8, No 2, 2014

Daftar Isi:

Pendeteksian Gerakan Menggunakan Transduser Ultrasound dengan Metoda Pembandingan Pola Gema Dicky Rianto Prajitno, Ade Ramdan	61-68
Model dan Metoda Arsitektur pada Sistem Tanya Jawab Medis Wiwini Suwarningsih, Iping Supriana, Ayu Purwarianti	69-82
Model Pengukuran Proses Bisnis CRM Berbasis eTom dan IT-IL Lihardo Ranjaliba Saragih, Gendut Suprayitno, Dahmir Dahlan	83-94
Perbandingan Performa Pelacakan antara Repetitive Controller dan PI Controller Edi Kurniawan	95-102
Pengenalan Entitas User Profile Pada Twitter Titin Pramiyati, Iping Supriana, Ayu Purwarianti	103-110
Momen Warna CIELab: Deskriptor Alternatif untuk Sistem Klasifikasi Citra LANDSAT Retno Kusumaningrum, Hisar Maruli Manurung, Aniati Murni Arymurthy	111-116

2/2014

 Pusat Penelitian Informatika - LIPI					
Jurnal INKOM	Vol. 8	No. 2	Hal. 61-116	Bandung, November 2014	p-ISSN 1979-8059 e-ISSN 2302-6146

INKOM

Jurnal Informatika, Sistem Kendali, dan Komputer

p-ISSN 1979-8059, e-ISSN 2302-6146

Volume 8, No 2, November 2014

Penanggung Jawab

Kepala Pusat Penelitian Informatika - LIPI

Dewan Redaksi

<i>Ketua</i>	Dr. Didi Rosiyadi	Puslit Informatika LIPI
<i>Anggota</i>	Prof. Dr. Ir. Engkos Koswara N., M.Sc.	Puslit Informatika LIPI
	Dr. Ir. Ashwin Sasongko Sastrosubroto., M.Sc.	Puslit Informatika LIPI
	Drs. Tigor Nauli	Puslit Informatika LIPI
	Dr. Ana Hadiana, M.Eng.	Puslit Informatika LIPI
	Dr. Nasrullah Armi	PPET LIPI

Redaksi Pelaksana

<i>Ketua</i>	Arnida L. Latifah, S.Si., M.Sc.	Puslit Informatika LIPI
<i>Penyunting Ahli</i>	Suyoto, ST., MT.	Puslit Informatika LIPI
	Bambang Sugiarto, ST., MT	Puslit Informatika LIPI
	Raden Sandra Yuwana, MT.	Puslit Informatika LIPI
<i>Desain Grafis</i>	Dicky Rianto Prajitno, MT.	Puslit Informatika LIPI

Mitra Bestari

Prof. I. KG. Darma Putra (Komputer), Dr. Achmad Basuki (Informatika), Dr. Adha Imam Cahyadi (Otomasi), Dr. Bayu Hendradjaya (Informatika), Dr. Erma Suryani (Informatika), Dr. Joni Welman S. (Otomasi), Dr. M. Udin Harun Al Rasyid (Otomasi), Dr. Sugihartono (Komputer), Irwan Purnama, M.Sc.Eng (Otomasi), Leon Andretti A, M.M (Informatika)

Sekretariat

Rini Wijayanti, M.Kom	Puslit Informatika LIPI
Adi Nurhadiyatna, S.Kom., M.Kom	Puslit Informatika LIPI

Alamat Redaksi

Jurnal INKOM

Pusat Penelitian Informatika, Lembaga Ilmu Pengetahuan Indonesia

Komp. LIPI Gd. 20 Lt. 3 Jln Sangkuriang, Bandung, 40135

Telp: +62 22 2504711, Fax: +62 22 2504712

Email: jurnal@informatika.lipi.go.id, Website: <http://jurnal.informatika.lipi.go.id>

Pertama terbit: Mei 2007

Frekuensi terbit: Dua kali setahun, setiap bulan Mei dan November

Jurnal INKOM adalah jurnal yang mengkaji masalah yang berhubungan dengan Informatika, Sistem Kendali, dan Komputer dengan keberkalaan penerbitan dua kali setahun pada Mei dan November. Tulisan yang dipublikasikan berupa hasil penelitian, pemikiran atau pengembangan untuk kemajuan keilmuan atau terapan. Kelayakan pemuatan dipertimbangkan oleh penilai dengan *double blind review* berdasarkan keaslian (originalitas) dan keabsahan (validitas) ilmiah.

©2014 Hak cipta dilindungi undang-undang

Daftar Isi

Hasil Penelitian	Halaman
<i>Pendeteksian Gerakan Menggunakan Transduser Ultrasound dengan Metoda Perbandingan Pola Gema</i> Dicky Rianto Prajitno, Ade Ramdan	61-68
<i>Model dan Metoda Arsitektur pada Sistem Tanya Jawab Medis</i> Wiwin Suwarningsih, Iping Supriana, Ayu Purwarianti	69-82
<i>Model Pengukuran Proses Bisnis CRM Berbasis eTom dan IT-IL</i> Lihardo Ranjaliba Saragih, Gendut Suprayitno, Dahmir Dahlan	83-94
<i>Perbandingan Performa Pelacakan antara Repetitive Controller dan PI Controller</i> Edi Kurniawan	95-102
<i>Pengenalan Entitas User Profile pada Twitter</i> Titin Pramiyati, Iping Supriana, Ayu Purwarianti	103-110
<i>Momen Warna CIELab: Deskriptor Alternatif untuk Sistem Klasifikasi Citra LANDSAT</i> Retno Kusumaningrum, Hisar Maruli Manurung, Aniati Murni Arymurthy	111-116



Jurnal Informatika, Sistem Kendali, dan Komputer

Volume 8, No 2, 2014

Editorial

Pembaca yang terhormat, Jurnal INKOM volume 8 nomor 2 tahun 2014 berisi 6 karya tulis ilmiah. Dengan rincian 4 makalah merupakan tulisan dalam bidang informatika: Model dan Metoda Arsitektur pada Sistem Tanya Jawab Medis, Model Pengukuran Proses Bisnis CRM Berbasis eTom dan IT-IL, Pengenalan Entitas *User Profile* pada *Twitter*, *CIELab Color Moments: Alternative Descriptors for LANDSAT Images Classification System*; 1 makalah merupakan tulisan dalam bidang komputer: Pendeteksian Gerakan Menggunakan Transduser *Ultrasound* dengan Metoda Perbandingan Pola Gema. Serta satu makalah dalam bidang kendali yaitu *On Comparison between Repetitive Controller and PI Controller Tracking Performance*. Semua tulisan ilmiah yang dipublikasikan telah melalui proses seleksi dengan metoda *double blind* oleh dewan redaksi dan mitra bestari.

Rencana ke depan Jurnal INKOM adalah menjadikan Jurnal INKOM memenuhi standar sehingga dapat diindeks pada indeks global. Dalam waktu dekat, setiap naskah yang dipublikasi pada Jurnal INKOM memiliki DOI (*digital object identifier*). Besar harapan kami agar Jurnal INKOM berkembang menjadi Jurnal yang bermutu.

Pada kesempatan ini, kami mengucapkan banyak terima kasih kepada dewan pengarah, dewan redaksi, editor pelaksana, sekretariat dan para penulis yang telah melancarkan proses penerbitan Jurnal INKOM nomor ini. Mudah-mudahan Jurnal INKOM volume 8 nomor 2, 2014 ini memberikan kontribusi bagi perkembangan ilmu pengetahuan di Indonesia.

Ketua Dewan Editor

Kata kunci yang dicantumkan adalah istilah bebas. Lembar abstrak ini boleh disalin tanpa izin dan biaya

DDC 621.38

Dicky Rianto Prajitno, Ade Ramdan (Pusat Penelitian Informatika-LIPI)

Pendeteksian Gerakan Menggunakan Transduser Ultrasound dengan Metoda Perbandingan Pola Gema

INKOM, 8(2) 2014: 61-68

Pada tulisan ini kami mengusulkan metoda pendeteksi gerakan. Pendeteksian dilakukan dengan cara membandingkan dua buah pola gema dalam interval waktu tertentu dan jeda antara pola gema. Perbedaan pola gema menunjukkan terjadinya gerakan dari benda-benda di area deteksi sensor. Dengan metoda perbandingan ini, sensor menjadi sensitif terhadap gerakan atau perubahan posisi setiap objek yang berada disekitarnya. Hasil yang diperoleh memperlihatkan bahwa sensor ini secara khusus lebih sensitif dibandingkan sensor PIR dalam mendeteksi gerakan-gerakan kecil. Pada prakteknya metoda ini telah mampu mendeteksi beberapa gerakan tubuh manusia seperti: dari mulai berjalan; duduk; berbicara; hingga gerak respirasi tubuh.

(Penulis)

Kata kunci: ultrasound, deteksi gerakan, pulse echo ranging, pola gema

DDC 621.39

Wiwin Suwarningsih, Iping Supriana, Ayu Purwarianti (Sekolah Teknik Elektro dan Informatik, Institut Teknologi Bandung, Penelitian Informatika, Lembaga Ilmu Pengetahuan Indonesia)

Model dan Metoda Arsitektur pada Sistem Tanya Jawab Medis

INKOM, 8(2) 2014: 69-82

Pada makalah ini, akan dilakukan survey beberapa penelitian yang membahas mengenai sistem tanya jawab dengan domain pada bidang medis (*medical question answering* = MedQuAn). Sistem MedQuAn mengolah pertanyaan yang diajukan dalam bentuk teks bahasa alami dan kemudian sistem akan memberikan jawaban yang relevan. Makalah ini mencoba menelaah modul konseptual MedQuAn, bahwa sistem tanya jawab terdiri dari tiga komponen inti yang berbeda beserta metoda/pendekatan yang digunakan. Ketiga komponen inti tersebut adalah klasifikasi pertanyaan, pencarian dokumen, dan ekstraksi jawaban. Hasil akhir dari survey ini adalah sebuah kontribusi untuk pengembangan penelitian di masa mendatang di domain MedQuAn khususnya untuk sistem tanya jawab medis dengan menggunakan bahasa Indonesia.

(Penulis)

Kata kunci: model MedQuAn, klasifikasi pertanyaan medis, ekstraksi jawaban medis, temu balik informasi

Kata kunci yang dicantumkan adalah istilah bebas. Lembar abstrak ini boleh disalin tanpa izin dan biaya

DDC 621.39

Lihardo Ranjaliba Saragih, Gendut Suprayitno, Dahmir Dahlan (Institut Sains dan Teknologi Nasional)

Model Pengukuran Proses Bisnis CRM Berbasis eTom dan IT-IL

INKOM, 8(2) 2014: 83-94

Saat ini perusahaan Telekomunikasi melakukan transformasi model bisnis melalui *Manage service*, untuk mengurangi beban biaya perusahaan. Perusahaan telekomunikasi sangat melekat dengan teknologi mutakhir dan padat modal, maka transformasi perlu dilakukan dengan pendekatan khusus sehingga dapat mempertahankan daya saing, bukan sekedar outsourcing karyawan umumnya. Hal ini kemudian mengubah paradigma outsourcing dalam industri telekomunikasi global dalam sebuah istilah *Manage service*. Perusahaan menyadari bahwa *Manage service* ternyata rumit didefinisikan apalagi dieksekusi, sehingga perlu dilakukan pengukuran tingkat kematangan dalam skala aktivitas atau proses bisnis yang akan dialihdayakan. Penelitian dilakukan untuk mengukur tingkat kematangan proses bisnis dalam *Manage services*, menghasilkan penataan ulang model bisnis perusahaan. Peneliti menggunakan framework eTOM dan IT-IL dalam mengukur unit bisnis "*Customer Relationship Management - Customer Care*" yang akan menerapkan *Manage service*. Hasil studi adalah berbagai aspek identifikasi terkait dengan layanan potensial outsourcing yang dikelola, saran restrukturisasi tata kelola sesuai dengan referensi tingkat kematangan, dan memberikan rekomendasi kebijakan yang tepat.

(Penulis)

Kata kunci: Tingkat Kematangan Proses Bisnis, eTOM, IT-IL, CRM Customer Care

DDC 621.38

Edi Kurniawan (Pusat Penelitian Informatika, Lembaga Ilmu Pengetahuan Indonesia)

Perbandingan Performa Pelacakan antara Repetitive Controller dan PI Controller

INKOM, 8(2) 2014: 95-102

Pelacakan sinyal periodik adalah kegiatan umum dalam banyak permasalahan kontrol. Salah satu contoh adalah kontrol gerakan *pick and place* robot di industri . Kebutuhan akurasi yang tinggi menjadi sangat penting dalam banyak aplikasi kontrol. Oleh karena itu, algoritma kontrol untuk mencapai pelacakan akurasi yang tinggi khususnya untuk sinyal periodik sangat diperlukan . *Repetitive Controller* (RC) berdasarkan prinsip *internal model* adalah salah satu dari banyak algoritma kontrol yang dapat digunakan untuk pelacakan sempurna sinyal periodik. Di sisi lain, *Proportional Integral* (PI) *controller* juga dapat digunakan untuk pelacakan. Makalah ini membandingkan kinerja pelacakan PI *controller*, RC , dan PI dengan RC, di mana PI dengan RC di sini merupakan integrasi antara PI *controller* dan RC. Langkah-langkah untuk mendapatkan parameter PI, RC, dan PI dengan RC di berikan dalam makalah ini. Simulasi untuk sistem servo motor dilakukan untuk menguji kinerja RC , PI , dan PI dengan RC . Dari hasil simulasi, *Transient response* dan akurasi pelacakan dibahas secara mendalam.

(Penulis)

Kata kunci: *Repetitive Control*, PI *controller*, pelacakan, sinyal periodik

Kata kunci yang dicantumkan adalah istilah bebas. Lembar abstrak ini boleh disalin tanpa izin dan biaya

DDC 621.39

Titin Pramiyati, Iping Supriana, Ayu Purwarianti (STEI-Institut Teknologi Bandung)

Pengenalan Entitas User Profile Pada Twitter

INKOM, 8(2) 2014: 103-110

Atribut *trust scope* sebagai atribut untuk menentukan tingkat kepercayaan sumber informasi, akan diisi dengan data yang terdapat pada *user profile Twitter* yang dikenal sebagai *Bio Twitter*. Hanya saja, data tersebut harus sesuai dengan karakteristik dan fungsi dari masing-masing atribut *trust scope*, seperti atribut pendidikan harus diisi dengan informasi yang berkaitan dengan latar belakang pendidikan dari pemilik profil tersebut. Untuk mendapatkan data yang sesuai dengan atribut, kami melakukan *named entity recognition*, yang merupakan salah satu kegiatan pada proses ekstraksi informasi. Oleh karena itu, paper ini menjelaskan hasil proses pengenalan entitas yang dilakukan terhadap data yang terdapat pada *user profile*. Perangkat lunak yang digunakan untuk mengenali data sebagai entitas adalah *IndonesiaNetagger*. *IndonesiaNetagger*, merupakan perangkat lunak untuk mengenali entitas yang ditulis dalam bahasa Indonesia. Kami melakukan penelitian dalam empat tahap, yaitu pengenalan *entity* dengan data *Bio twitter* yang asli, identifikasi kesalahan proses pengenalan, formalisasi data dan pengujian pengenalan entitas akhir. Hasil penelitian menunjukkan keberhasilan sebagai berikut; entitas *Person* dikenali dengan benar sebesar 71% dari total data entitas yang tersedia, entitas *Organization* dikenali dengan benar sebesar 50%, entitas *Position* 20% dikenali dengan benar, dan 50% entitas *Location* dikenali dengan benar.

(Penulis)

Kata kunci: atribut *trust scope*, tingkat kepercayaan, pemilik profil, pengenalan entitas

DDC 621.32

R. Kusumaningrum (Departemen Informatika, Universitas Diponegoro), H. M. Manurung, A. M. Arymurthy (Fakultas Ilmu Komputer, Universitas Indonesia)

Momen Warna CIELab: Deskriptor Alternatif untuk Sistem Klasifikasi Citra LANDSAT

INKOM, 8(2) 2014: 111-116

Penelitian ini membandingkan 2 jenis sistem klasifikasi citra, yaitu sistem berbasis Normalized Difference Vegetation Index (NDVI) serta sistem berbasis Latent Dirichlet Allocation (LDA) menggunakan momen warna CIELab sebagai deskriptornya. Penelitian tersebut diimplementasikan untuk klasifikasi citra LANDSAT dengan mengevaluasi nilai akurasi dari kedua jenis sistem tersebut. Tujuan dari studi ini adalah untuk mengevaluasi apakah momen warna CIELab dapat digunakan sebagai deskriptor alternatif dari NDVI ketika diterapkan menggunakan model klasifikasi berbasis LDA. Hasil eksperimen menunjukkan bahwa sistem klasifikasi berbasis LDA menggunakan momen warna CIELab memberikan kinerja akurasi yang lebih baik dibandingkan sistem klasifikasi berbasis NDVI, dengan perbandingan nilai akurasi sebesar 87,43% untuk model klasifikasi berbasis LDA dan 86,25% untuk model klasifikasi berbasis NDVI. Oleh karena itu, dapat disimpulkan bahwa momen warna CIELab yang diimplementasikan pada sistem klasifikasi berbasis LDA dapat digunakan sebagai alternatif deskriptor untuk sistem klasifikasi citra penginderaan jauh pada saat data yang tersedia terbatas, khususnya ketika hanya citra *true color composite* yang tersedia.

(Penulis)

Kata kunci: *Normalized Difference Vegetation Index*, *CIELab*, momen warna, *Latent Dirichlet Allocation*, *LANDSAT*, klasifikasi citra penginderaan jauh

The descriptor given are free terms. This abstract sheet may be reproduced without permission or charge.

DDC 621.38

Dicky Rianto Prajitno, Ade Ramdan (Research Center for Informatics, Indonesian Institute of Sciences)

Motion Detection using Ultrasonic Sensor with Echo Pattern Comparison

INKOM, 8(2) 2014: 61-68

In this paper we propose a method of motion detection. The motion detection is performed by comparing two echoes pattern with a certain time interval, and the interval between certain pattern of objects echoes. The distinction of the echoes pattern are indicating that several movement occur in the surrounding sensor area. With this comparison method, the sensor becomes sensitive to any objects movement in its detection area. The result show that the sensor specifically is more sensitive in detection small movements compared to PIR sensor. In practice, this method was able to detect several human body movements such as walking, sitting, talking, and even be able to detect the movement of the body caused by human respiratory.

(Author)

Keywords: ultrasound, motion sensor, pulse echo ranging, echo pattern

DDC 621.39

Wiwin Suwarningsih, Iping Supriana, Ayu Purwarianti (School of Electrical Engineering and Informatics, Institute of Technology Bandung, Research Center for Informatics, Indonesian Institute of Sciences)

Architectural Models and Methods in Medical Question Answering Systems

INKOM, 8(2) 2014: 69-82

In this survey paper, we reviewed the current state of the art in medical question answering (MedQuAn). The MedQuAn systems are concerned with providing relevant answers in response to questions proposed in natural language. This survey tried to review the MedQuAn conceptual model, in which the question answering (QA) is therefore composed of three different components, each of which has a core component beside other supplementary components and method/approach research. These three core components are question classification, document retrieval, and answer extraction. The final result of this survey is a contribution to the development of future research in the domain of MedQuAn especially for Indonesian medical question answering system.

(Author)

Keywords: MedQuAn model, medical question classification, medical answer extraction, information retrieval

The descriptor given are free terms. This abstract sheet may be reproduced without permission or charge.

DDC 621.39

Lihardo Ranjaliba Saragih, Gendut Suprayitno, Dahmir Dahlan (Institut Sains dan Teknologi Nasional)

A Measurement Model of CRM Business Based on eTom and IT-IL

INKOM, 8(2) 2014: 83-94

Nowadays Telecom Company doing business model transformation through Manage services, to reduce cost. Telecommunications companies obviously attached to cutting-edge technology and capital-intensive, therefore needs a different approach in Order to remain competitiveness, not just a simple outsourcing. It made paradigm of outsourcing in global Telecommunications industry changed, in term of Manage service. Company realizes that Manage service is complicated defined, even more complicate executed, so its necessary to have a Maturity level measurement in activities or business Processes scale that will be outsourced. This study is conducted to examine business Processes Maturity level that build Manage services, towards company business model rearrangement. Researchers using the eTOM framework and IT-IL in measuring business unit "Customer Relationship Management - Customer Care" which will apply Manage service. Study results identification aspects that related to potential Manage services outsourced, governance restructuring advice according to reference business Processes Maturity level, and provide appropriate policy recommendations.

(Author)

Keywords: Business Process Maturity Level, eTOM, IT-IL, CRM Customer Care

DDC 621.38

Edi Kurniawan (Research Center for Informatics, Indonesian Institute of Sciences)

On Comparison Between Repetitive Controller and PI Controller Tracking Performance

INKOM, 8(2) 2014: 95-102

Tracking periodic signals are common task in many control problems. One of the examples is movement control of pick and place robot in industry. The requirement of high tracking accuracy becomes very important in many applications. Therefore, a sophisticated control algorithm that manages to achieve high accuracy tracking of periodic command is required. Repetitive Control (RC) based on internal model principle is one of control schemes that can be employed to achieve perfect tracking of periodic signal. On the other hand, Proportional Integral (PI) controller can also be used for tracking. This paper compares the tracking performance of PI controller, RC, and PI with RC, where PI with RC here is integration between PI controller and RC. Step by step design to obtain the parameters of PI, RC and PI with RC are given. A simulation on servo motor system is carried out to assess the performance of RC, PI, and PI with RC respectively. From the simulation results, the transient response and tracking accuracy are thoroughly discussed.

(Author)

Keywords: Repetitive Control, PI controller, tracking, periodic signal

The descriptor given are free terms. This abstract sheet may be reproduced without permission or charge.

DDC 621.39

Titin Pramiyati, Iping Supriana, Ayu Purwarianti (STEI-Institut Teknologi Bandung)

Entity Recognition of User Profile on Twitter

INKOM, 8(2) 2014: 103-110

Trust scope attribute as an attribute to determine the level of trust resources, will be filled with the data contained in the user profile of Twitterone of social media known as Bio Twitter. However, these data should be in accordance with the characteristics and functions of each attribute, such as education attribute must be filled in with the information relating to the educational background of the owner of the profile. To obtain the data corresponding to the trust scope attributes, we perform named entity recognition, which is one of the activities in the process of information extraction. Therefore, this paper describes the results of the entity recognition process performed on data contained in the user profile. Software used to recognize the data as an entity is Indonesia Netagger, which is to perform entity recognition that written in Indonesian language. The software recognizes only five entities namely Person, Organization, Location, Positionand Other. We carried out the research by conducting four stages namely entity recognition-with original data-Bio Twitter, error identification, formalizing data, and final test. The results show the success of entity recogniton as follow; Person entity is recognized correctly by 71% of the total data available, the entity Organization recognized correctly by 50%, 20% Position entity recognized correctly, and 50% recognized correctly as Location entity.

(Author)

Kata kunci: trust scope atributes, level trust, User profile, named entity recognition

DDC 621.32

R. Kusumaningrum (Department of Informatics, Diponegoro University), H. M. Manurung, A. M. Arymurthy (Faculty of Computer Science, University of Indonesia)

CIELab Color Moments: Alternative Descriptors for LANDSAT Images Classification System

INKOM, 8(2) 2014: 111-116

This study compares the image classification system based on normalized difference vegetation index (NDVI) and Latent Dirichlet Allocation (LDA) using CIELab color moments as image descriptors. It was implemented for LANDSAT images classification by evaluating the accuracy values of classification systems. The aim of this study is to evaluate whether the CIELab color moments can be used as an alternatif descriptor replacing NDVI when it is implemented using LDA-based classification model. The result shows that the LDA-based image classification system using CIELab color moments provides better performance accuracy than the NDVI-based image classification system, i.e 87.43% and 86.25% for LDA-based and NDVI-based respectively. Therefore, we conclude that the CIELab color moments which are implemented under the LDA-based image classification system can be assigned as alternative image descriptors for the remote sensing image classification systems with the limited data availability, especially when the data only available in true color composite images.

(Author)

Keywords: Normalized Difference Vegetation Index, Latent Dirichlet Allocation, remote sensing image classification

Pendeteksian Gerakan Menggunakan Transduser *Ultrasound* dengan Metoda Perbandingan Pola Gema

Motion Detection using Ultrasonic Sensor with Echo Pattern Comparison Method

Dicky Rianto Prajitno, Ade Ramdan

Pusat Penelitian Informatika-LIPI

Jl. Sangkuriang 21 Bandung

Email: dicky@informatika.lipi.go.id; ade@informatika.lipi.go.id

Abstract

In this paper we propose a method of motion detection. The motion detection is performed by comparing two echoes pattern with a certain time interval, and the interval between certain pattern of objects's echoes. The distinction of the echoes pattern are indicating that several movement occur in the surrounding sensor area. With this comparison method, the sensor becomes sensitive to any objects movement in it's detection area. The result show that the sensor specifically is more sensitive in detection small movements compared to PIR sensor. In practice, this method was able to detect several human body movements such as walking, sitting, talking, and even be able to detect the movement of the body caused by human respiratory.

Keywords: *ultrasound, motion sensor, pulse echo ranging, echo pattern*

Abstrak

Pada tulisan ini kami mengusulkan metoda pendeteksi gerakan. Pendeteksian dilakukan dengan cara membandingkan dua buah pola gema dalam interval waktu tertentu dan jeda antara pola gema. Perbedaan pola gema menunjukkan terjadinya gerakan dari benda-benda di area deteksi sensor. Dengan metoda perbandingan ini, sensor menjadi sensitif terhadap gerakan atau perubahan posisi setiap objek yang berada disekitarnya. Hasil yang diperoleh memperlihatkan bahwa sensor ini secara khusus lebih sensitif dibandingkan sensor PIR dalam mendeteksi gerakan-gerakan kecil. Pada prakteknya metoda ini telah mampu mendeteksi beberapa gerakan tubuh manusia seperti: dari mulai berjalan; duduk; berbicara; hingga gerak respirasi tubuh.

Kata kunci: *ultrasound, deteksi gerakan, pulse echo ranging, pola gema*

1. Pendahuluan

Saat ini banyak aplikasi yang membutuhkan sebuah pendeteksi gerakan[1], seperti pada aplikasi pengamanan, pengawasan, sistem otomatis[2], game[3], dan lain-lain. Pada sistem otomatis seperti misalnya pintu otomatis, atau juga saklar lampu otomatis, dibutuhkan alat yang dapat mendeteksi keberadaan manusia. Salah satu cara untuk mendeteksi keberadaan manusia adalah dengan mendeteksi gerakan.

Satu ciri manusia sebagai makhluk hidup adalah manusia akan selalu bergerak, entah itu gerakan-gerakan besar/kasar seperti berjalan, dan berbicara, ataupun hanya sebatas gerakan-gerakan kecil/halus seperti bernafas.

Terdapat beberapa jenis sensor pendeteksi gerakan yang banyak diaplikasikan, misalnya: *Pyroelectric Infrared* (PIR)[4], kamera video, *Ultrasound*, dan *Microwave*[5]. Masing-masing sensor memiliki karakteristik sensitifitas, resolusi, dan instalasi yang berbeda.

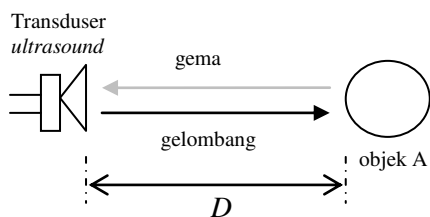
Pendeteksi gerakan kebanyakan menggunakan sensor PIR karena kepraktisannya, sayangnya resolusi deteksi yang rendah dan sifat *line of sight* menjadikan kendala pada beberapa kasus aplikasi[6]. Pada tulisan ini, diusulkan sebuah metoda untuk mendeteksi gerakan berbasis gelombang *ultrasound*. Metoda ini memiliki resolusi yang lebih baik, sehingga dapat mendeteksi gerakan-gerakan kecil yang tidak terdeteksi oleh PIR[7].

Organisasi penulisan adalah sebagai berikut: bagian pertama adalah pendahuluan, pada bagian kedua dijelaskan mengenai metoda deteksi gerakan dengan perbandingan pola gema, pada bagian

ketiga dijelaskan mengenai setup percobaan, pada bagian keempat dijelaskan mengenai hasil dan pembahasannya, dan bagian kelima adalah kesimpulan.

2. Metoda Deteksi Gerakan dengan Perbandingan Pola Gema

Metoda pengukuran jarak sebagai salah satu teknik pendeteksian gerakan dapat digambarkan seperti diperlihatkan pada Gambar 1. Metoda *pulse echo ranging* memanfaatkan laju cepat rambat gelombang akustik pada medium, dimana laju cepat rambat gelombang akustik di medium udara adalah 344m/s [8].



Gambar 1. Transduser *ultrasound* dengan sebuah objek. diam

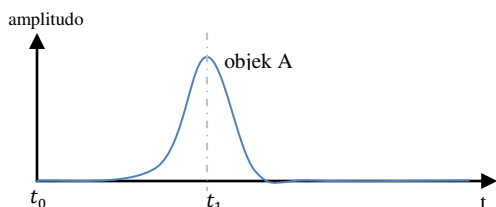
Metoda ini mengukur selang waktu yang dibutuhkan oleh gelombang dari mulai saat dipancarkan hingga gemanya diterima kembali. Persamaan untuk menentukan jarak objek terhadap sumber adalah sebagai berikut:

$$D = \frac{t \cdot v}{2}$$

dimana:

D adalah jarak dalam satuan meter,
 t waktu tempuh gelombang dalam satuan detik,
 v adalah cepat rambat gelombang akustik di medium, (udara = 343m/s).

Gambar 2 memperlihatkan pola gema dan selang waktu ($t_1 - t_0$) yang didengar oleh penerima dari objek A.



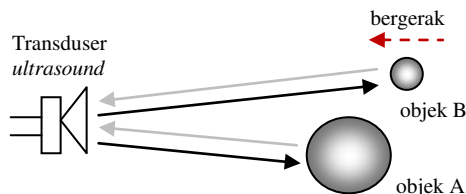
Gambar 2. Pola gema dari objek A.

Pada prakteknya jika terdapat lebih dari satu buah gema yang terdeteksi, maka yang diukur adalah gema pertama terdeteksi, sedangkan gema-gema berikutnya akan diabaikan. *Pulse echo ranging*

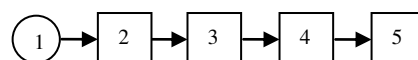
biasa digunakan pada pengukuran jarak, seperti sensor parkir pada mobil.

Hampir sama dengan metoda *pulse echo ranging*, pada perbandingan pola gema, pemancar akan mengirimkan gelombang pulsa. Objek-objek disekitar sensor akan memantulkan gema, kemudian gema ini direkam dalam interval waktu tertentu. Untuk mengetahui ada atau tidaknya pergerakan benda-benda di area deteksi, maka dua buah pola gema yang telah direkam akan dibandingkan, untuk kemudian dilihat apakah terdapat perbedaan bentuk kedua pola tersebut atau tidak. Jika ternyata terdapat perbedaan diantara keduanya, maka dapat disimpulkan bahwa telah terjadi pergerakan objek-objek di sekitar sensor [11].

Ilustrasi pada Gambar 3.a sampai dengan Gambar 3.g dibawah ini dapat menggambarkan metoda yang dimaksud. Sebuah transduser *ultrasound* memancarkan gelombang *ultrasound* ke arah dua buah objek: A dan B. Objek A dalam posisi diam, sedangkan objek B bergerak mendekati transduser seperti pada Gambar 3.a. Kemudian gema yang dihasilkan kedua objek tersebut ditangkap kembali oleh transduser dan kemudian dilakukan sampel data pola gema sebanyak dua kali dalam jeda waktu tertentu, seperti yang diperlihatkan pada Gambar 3.d dan Gambar 3.e.



Gambar 3.a. Transduser *ultrasound* dan dua buah objek yang berbeda



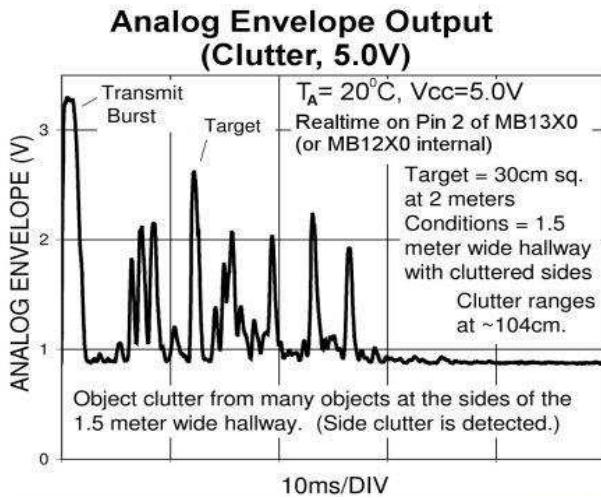
Gambar 3.b. Diagram blok sistem.

Diagram blok sistem seperti yang diperlihatkan pada Gambar 3.b., dengan urutan sebagai berikut: 1 = transduser *ultrasound*, 2 = ADC, 3 = memori, 4 = komparator.

Transduser *ultrasound* akan memancarkan gelombang ultrasonik dan kemudian menangkap gema dari benda-benda disekelilingnya, gema tersebut kemudian didijitalisasi selama selang waktu tertentu menggunakan ADC dan membentuk sebuah pola gema. Pola gema ini kemudian disimpan pada dua buah *array* di dalam

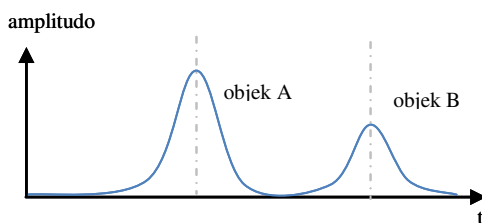
memori. Kemudian komparator akan membandingkan kedua buah pola tersebut.

Transduser yang digunakan berupa *envelope detector*, contoh keluaran transduser *envelope detector* seperti diperlihatkan pada Gambar 3.c.



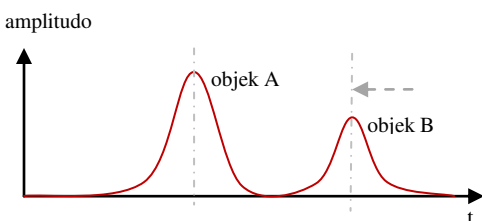
Gambar 3.c. Keluaran analog envelope dari transduser *ultrasound* MaxBotix1360[11].

Gambar 3.d memperlihatkan grafik sinyal gema dari objek A yang diam, dan sinyal gema dari objek B sebelum berpindah posisi (sebelum bergerak).



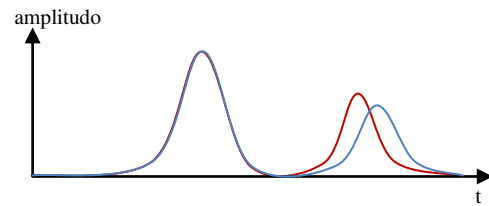
Gambar 3.d. Pola gema pertama objek A dan objek B.

Gambar 3.e memperlihatkan grafik sinyal gema dari objek A yang diam, dan sinyal gema dari objek B setelah berpindah posisi (setelah bergerak).



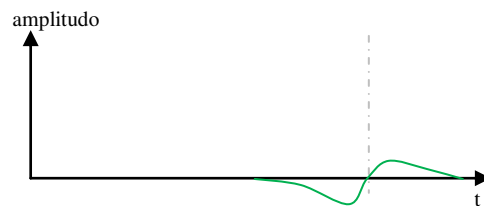
Gambar 3.e. Pola gema kedua objek A dan objek B, Objek B bergerak mendekati sensor.

Gambar 3.f memperlihatkan perbandingan grafik sinyal gema dari objek A yang diam, dan sinyal gema dari objek B setelah berpindah posisi, akan terdapat perbedaan kurva dari gema objek B yang akan merubah fasa dan amplitudonya[9].



Gambar 3.f. Perbandingan pola gema benda ke 1 terhadap benda ke 2.

Gambar 3.g. memperlihatkan selisih dari penjumlahan kedua pola gema tersebut.



Gambar 3.g. Selisih pola gema pertama terhadap pola gema kedua.

Integral dari selisih kedua sinyal ini secara relatif memperlihatkan seberapa besar pergerakan yang terjadi.

Secara sederhana jika kuantitas selisih amplitudo kedua pola gema tersebut tidak sama dengan nol maka berarti sensor mendeteksi adanya gerakan dari benda-benda di area cakupannya[9,10].

Dengan menggunakan ADC pola gema tersebut disampling dan dikuantisasi, kemudian disimpan didalam *array* data q . Terdapat dua buah *array* data gema, yang terdiri dari *array* data gema sebelumnya q' , dan *array* data gema sekarang q . Data tersebut dibandingkan untuk setiap nilai kuantisasi m_n , kemudian diakumulasi M .

Persamaan selisih absolut untuk setiap nilai kuantisasi adalah:

$$m_n = \text{abs}[q'_n - q_n] \quad (1)$$

Dan nilai total dari selisih kuantisasi adalah:

$$M = \sum_0^n m_n \quad (2)$$

Pada prakteknya digunakan *threshold* untuk mengeliminasi *noise* dari gema-gema yang tidak diinginkan, sehingga persamaannya menjadi seperti dibawah ini.

$$\therefore M = \sum_0^n f(m_n) = \begin{cases} 0, & m_n \leq \text{threshold} \\ m_n, & m_n > \text{threshold} \end{cases} \quad (3)$$

Untuk mengatur sensitifitas sistem terhadap gerakan maka dapat kita atur pada besaran berapa sistem akan bereaksi. Pada prakteknya besaran ini bisa kita dapatkan dari percobaan di lab.

3. Setup Percobaan

Pembandingan pola gema dilakukan dengan membandingkan dua buah pola gema yang sebelumnya telah direkam ke dalam *array*. Masing-masing pola gema diambil sampel sebanyak 200 buah, dengan lama pengambilan data sebesar 45ms agar didapat jarak tempuh gelombang terjauh yang dapat direkam sebesar sekitar 15 meter. Selang waktu pengambilan pola gema pertama dan kedua adalah sepertiga detik, dengan pertimbangan optimalisasi pada pengukuran gerak tubuh manusia, dari gerakan eksternal seperti gerak kaki berjalan, gerak tangan, dan juga gerakan internal yaitu gerak tubuh saat bernafas (proses respirasi eksternal).

Pengambilan sampel menggunakan ADC dengan resolusi kuantisasi sebesar 10bit untuk mendapatkan ketelitian yang cukup.

Setup pengukuran adalah sebagai berikut: ruangan berukuran 5 x 6 meter, tinggi plafon 3 meter, sensor dipasang di tengah-tengah plafon. Pengukuran dibagi menjadi 3 skenario, yaitu: ruangan kosong, ruangan diisi benda tidak bergerak, ruangan diisi benda bergerak. Pada tulisan ini ditampilkan hasil pengukuran untuk ruangan kosong, ruangan berisi orang yang sedang berjalan kaki, kemudian ruangan diisi orang yang sedang duduk diam tanpa melakukan gerakan.

4. Hasil dan Pembahasan

Hasil pengukuran ditampilkan melalui gambar-gambar pola gema yang ditangkap dan hasil pengolahannya seperti yang ditampilkan pada Gambar 4. Gambar 4.a sampai dengan Gambar 4.l memperlihatkan hasil pengukuran pola gema, dimana sumbu- x adalah besaran kuantisasi 10bit (0-1023) representasi amplitudo tegangan dari 0 sampai dengan 5volt, dan sumbu-y adalah urutan sampel (1-200) representasi waktu dari 0 sampai dengan 45ms.

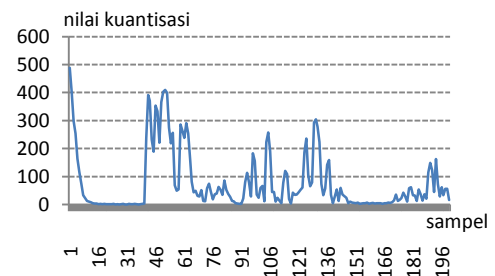
Gambar 4.a sampai dengan 4.d menampilkan hasil pengukuran dan pengolahan pola gema ruangan yang dalam keadaan kosong. Gambar 4.a dan 4.b adalah dua buah pola gema yang diambil berurutan berselang sekitar sepertiga detik, kemudian 4.c merupakan hasil perbandingan pola gema dari gambar 4.a dan 4.b. Secara ideal pola

gema pertama dan kedua tersebut adalah identik, energi yang masih terlihat pada Gambar 4.c merupakan *noise* yang tidak diinginkan dan dieliminasi dengan menambahkan *threshold* pada sistem, hasilnya seperti diperlihatkan pada gambar 4.d.

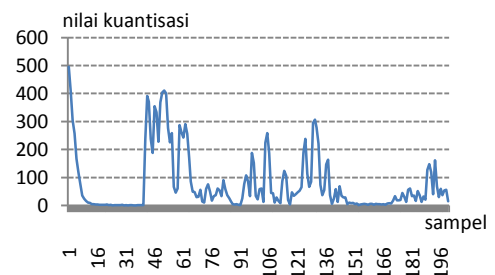
Dari pengukuran laboratorium seperti diperlihatkan pada Tabel 1 didapat *noise* yang tidak diinginkan rata-rata sekitar 11, maka jika *threshold* kita berikan di angka 15-20 sudah cukup baik.

Tabel 1. Pengukuran *noise* yang tidak diinginkan

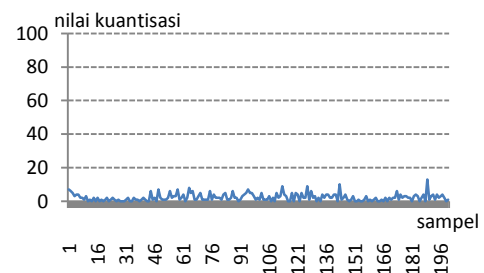
Jumlah pengukuran	Nilai terkecil	Nilai terbesar	Nilai rata-rata
30 pola, jumlah data 6000 buah	0	36	10.97 ≈ 11



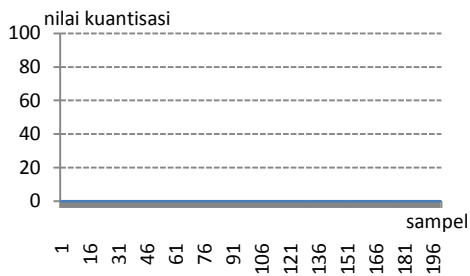
Gambar 4.a. Pola gema pertama, di ruang statis.



Gambar 4.b. Pola gema kedua, di ruang statis.



Gambar 4.c. Nilai absolut perbandingan gema pertama dan gema kedua.



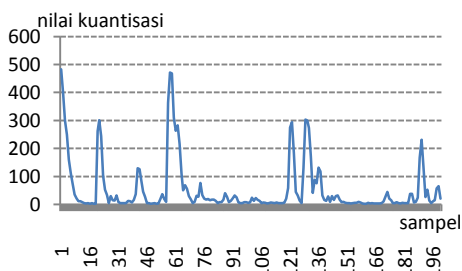
Gambar 4.d. Hasil setelah dikurangi *threshold*, jumlahnya adalah 0.

Gambar 4.e sampai dengan Gambar 4.h menampilkan hasil pengukuran dan pengolahan pola gema ruangan yang didalamnya terdapat seseorang sedang melakukan gerakan berjalan. Gambar 4.e dan 4.h adalah dua buah pola gema yang diambil berurutan, kemudian 4.c merupakan hasil perbandingan pola gema dari gambar 4.e dan 4.f, energi yang terlihat pada gambar 4.g merupakan selisih kedua pola gema tersebut, dan gambar 4.h merupakan grafik setelah *threshold*.

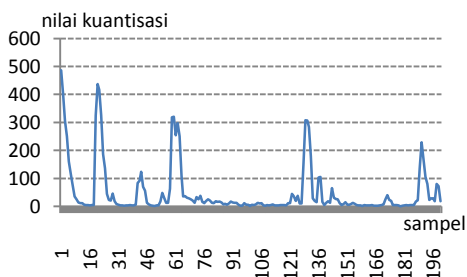
Dari pengukuran laboratorium seperti diperlihatkan pada Tabel 2 didapat nilai kuantisasi yang cukup besar dengan nilai kuantisasi terbesar adalah 668 dan nilai rata-rata sebesar 28869.

Tabel 2. Pengukuran orang berjalan

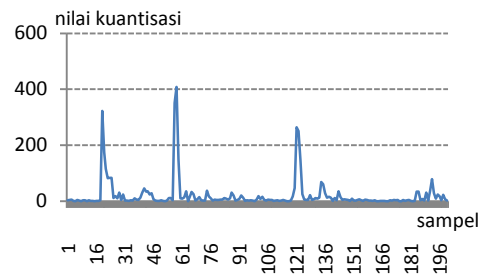
Jumlah pengukuran	Nilai terkecil	Nilai terbesar	Nilai rata-rata
24 pola, dgn jumlah data 4800 buah	0	668	28869



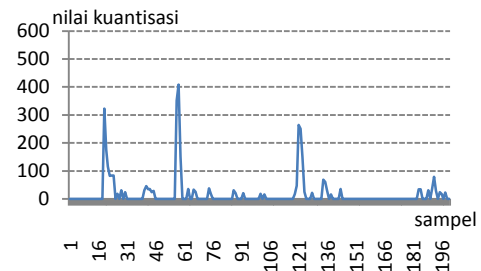
Gambar 4.e. Pola gema pertama, di ruang dinamis (orang sedang berjalan).



Gambar 4.f. Pola gema kedua, di ruang dinamis (orang sedang berjalan).



Gambar 4.g. Nilai absolut perbandingan gema pertama dan gema kedua.



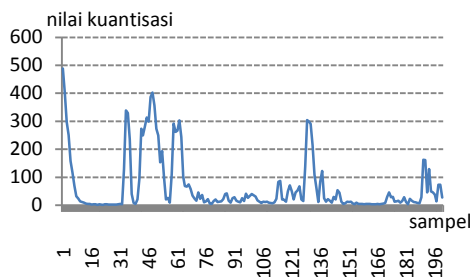
Gambar 4.h. Hasil setelah dikurangi *threshold*, jumlahnya adalah 3586.

Gambar 4.i sampai dengan 4.l menampilkan hasil pengukuran dan pengolahan pola gema ruangan yang didalamnya terdapat seseorang sedang duduk diam. Gambar 4.i dan 4.j adalah dua buah pola gema yang diambil berurutan, kemudian Gambar 4.k merupakan hasil perbandingan pola gema dari Gambar 4.i dan 4.j, energi yang terlihat pada Gambar 4.k merupakan selisih kedua pola gema tersebut, dan Gambar 4.l merupakan grafik setelah *threshold*.

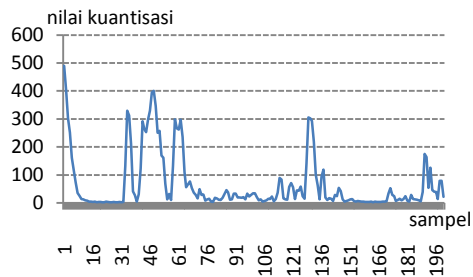
Dari pengukuran laboratorium seperti diperlihatkan pada Tabel 3 didapat nilai kuantisasi yang cukup besar dengan nilai kuantisasi terbesar adalah 242 dan nilai rata-rata sebesar 489.

Tabel 3. Pengukuran orang diam

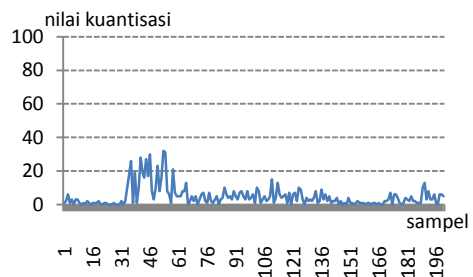
Jumlah pengukuran	Nilai terkecil	Nilai terbesar	Nilai rata-rata
30 pola, dgn jml data 6000	0	242	489



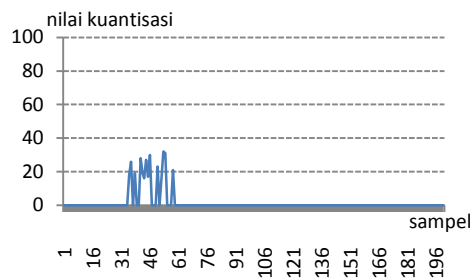
Gambar 4.i. Pola gema pertama di ruang dinamis (orang sedang duduk diam).



Gambar 4.j. Pola gema kedua di ruang dinamis (orang sedang duduk diam).



Gambar 4.k. Nilai absolut perbandingan gema pertama dan gema kedua.



Gambar 4.l. Hasil setelah *threshold* jumlahnya 323.

5. Kesimpulan

Karena metoda perbandingan pola gema secara relatif merekam seluruh gema yang dipantulkan oleh objek-objek di area sensitif sensor, berbeda dengan metoda *pulse echo ranging* yang hanya melakukan pengukuran jarak terhadap objek terdekat saja, maka metoda perbandingan pola

gema lebih efektif digunakan sebagai pendeteksi gerakan.

Dengan penekanan *noise* yang baik pada transduser dan resolusi kuantisasi ADC yang tinggi, dapat dihasilkan pendeteksi gerakan yang mempunyai sensitifitas tinggi.

Dari hasil beberapa skenario pengukuran, untuk pengukuran gerakan-gerakan besar hingga gerakan-gerakan halus, didapatkan hasil yang cukup baik, sistem dapat mendeteksi gerakan-gerakan orang yang sedang berjalan dengan baik, terlihat dari perolehan selisih nilai kuantisasi yang cukup besar, dari 24 sampel pola gema yang diambil nilai perolehan rata-ratanya adalah 28689. Kemudian sistem juga dapat mendeteksi gerakan orang bernafas dengan baik, dari 30 sampel didapat nilai perolehan rata-ratanya sebesar 489. Sedangkan untuk pengukuran ruangan kosong didapat nilai-nilai signifikan di sekitar 32 – 36, dengan nilai rata-rata sekitar 11, sehingga dapat digunakan nilai *threshold* diatas angka 15. Pada penelitian ini digunakan nilai *threshold* sebesar 20, dengan hasil yang cukup baik.

Pada pengukuran di laboratorium menggunakan transduser *ultrasound* jenis MaxBotix-1360[11] dan ADC 10bit dari *controller board* arduino uno[12], sistem cukup sensitif untuk dapat mendeteksi gerakan-gerakan tubuh manusia, dari mulai gerakan-gerakan yang besar seperti gerakan berjalan, hingga gerakan yang kecil seperti gerakan respirasi eksternal.

Daftar pustaka

- [1] T. Teixeira, G. Dublon, and A. Savvides, "A Survey of Human Sensing: Methods for Detecting Presence, Count, Location, Track and Identity", ACM Computing Surveys, 2010.
- [2] M.A. Troost, "Presence detection and activity recognition using low-resolution passive IR sensors", Technische Universiteit Eindhoven, 2013.
- [3] Clare Chen, Grace Li, Peter Ngo, and Connie Sun, "Motion Sensing Technology", Management of Technology – E 103, Team Bazinga, 2011.
- [4] Glolab Corporation, "Infrared Motion Detector Manual", 307 Pine Ridge Drive, Wappingers Falls, NY 12590, Glolab Corp, 2005.
- [5] Sukhivinder Singh, Qilian Liang, and Dechang Chen, "Sense through Wall Human Detection Using UWB Radar", Dept. of Electrical Engineering, University of Texas at Arlington, Arlington, 2011.
- [6] Leslie Hodges, "Ultrasonic and Passive Infrared Sensor Integration for Dual Technology User Detection Sensors", ECE 480 - Application Note, Michigan State University, 2009.

- [7] Don B Munroe, "Comparison of Watt Stopper's Patented Dual Technology Sensor vs. Microphonics", <http://www.wattstopper.com/>, 2013, diakses tanggal 10-03-2013.
- [8] Suzanne Amador Kane, "Ultrasound Imaging Physics", Physics Department, 370 Lancaster Avenue, Haverford College, Haverford PA 19041, 2008.
- [9] Victor Kremin, "Ultrasound Motion Sensor", Application Note – AN2047, Cypress MicroSystems, 2002.
- [10] Kurt Inman, "What Is an Ultrasonic Motion Detector?", <http://www.wisegeek.com/what-is-an-ultrasonic-motion-detector.htm>, diakses tanggal 10-03-2013.
- [11] XL-MaxSonar®- EZ™ Series High Performance Sonar Range Finder, MaxBotix Inc. USA, 2013.
- [12] Masimo Banzi, "Getting Started with Arduino" Second Edition, O'Reilly, 2013.

Model dan Metoda Arsitektur pada Sistem Tanya Jawab Medis

Architectural Models and Methods in Medical Question Answering Systems

Wiwin Suwarningsih^{1,2}, Iping Supriana¹, Ayu Purwarianti¹

¹*Sekolah Teknik Elektro dan Informatik, Institut Teknologi Bandung
Jl. Ganesha 10 Bandung*

²*Pusat Penelitian Informatika-LIPI*

Jl. Sangkuriang 21 Bandung

Email: wiwin.suwarningsih@lipi.go.id; iping@informatika.org; ayu@informatika.org

Abstract

In this survey paper, we reviewed the current state of the art in medical question answering (MedQuAn). The MedQuAn systems are concerned with providing relevant answers in response to questions proposed in natural language. This survey tried to review the MedQuAn conceptual model, in which the question answering (QA) is therefore composed of three different components, each of which has a core component beside other supplementary components and method/approach research. These three core components are question classification, document retrieval, and answer extraction. The final result of this survey is a contribution to the development of future research in the domain of MedQuAn especially for Indonesian medical question answering system.

Keywords: *MedQuAn model, medical question classification, medical answer extraction, information retrieval*

Abstrak

Pada makalah ini, akan dilakukan survey beberapa penelitian yang membahas mengenai sistem tanya jawab dengan domain pada bidang medis (*medical question answering* = MedQuAn). Sistem MedQuAn mengolah pertanyaan yang diajukan dalam bentuk teks bahasa alami dan kemudian sistem akan memberikan jawaban yang relevan. Makalah ini mencoba menelaah modul konseptual MedQuAn, bahwa sistem tanya jawab terdiri dari tiga komponen inti yang berbeda beserta metoda/pendekatan yang digunakan. Ketiga komponen inti tersebut adalah klasifikasi pertanyaan, pencarian dokumen, dan ekstraksi jawaban. Hasil akhir dari survey ini adalah sebuah kontribusi untuk pengembangan penelitian di masa mendatang di domain MedQuAn khususnya untuk sistem tanya jawab medis dengan menggunakan bahasa Indonesia.

Kata kunci: *model MedQuAn, klasifikasi pertanyaan medis, ekstraksi jawaban medis, temu balik informasi*

1. Pendahuluan

Sistem tanya jawab (*Question Answering System=QAS*) adalah sistem yang menerima masukan berupa pertanyaan dalam bahasa alami dan kemudian memberikan keluaran berupa jawaban dalam bahasa alami pula. Dilihat dari ruang lingkup permasalahan, QAS dapat dibagi menjadi dua jenis yaitu QAS dengan domain terbuka (pada domain ini semua permasalahan dapat dibahas melalui proses tanya jawab pada sistem) dan QAS dengan domain terbatas/tertutup (pada domain ini hanya membahas satu topik saja misalnya kesehatan, olahraga, politik dan sebagainya).

Pada makalah ini akan dibahas penelitian sistem tanya jawab dengan domain terbatas seperti domain medis, klinik, biomedis bahkan berbagai konsultasi kesehatan. Diawali dengan tiga QAS biomedis terkemuka yang merupakan bentuk aplikasi QAS biomedis *online* diantaranya adalah AskHermes ([2],[3]), EAGli [8] dan HONQA ([9],[10]). AskHermes diprakarsai oleh Cao *et al.* ([2],[3]), sistem ini memungkinkan dokter untuk memasukkan pertanyaan dengan cara alami dengan formulasi permintaan minimal dan memungkinkan dokter untuk secara efisien memandu semua kalimat jawaban untuk memberikan informasi secara cepat kepada pengguna. Gobeil *et al.* [8] menyuguhkan aplikasi dengan nama EAGli, aplikasi ini menggunakan *medical subject headings* untuk membantu menjawab pertanyaan

dengan jenis definisi. Bentuk jawaban yang dihasilkan berupa *multi-frase* dan daftar dari *entitas* tunggal. Sedangkan Cruchet *et al.* ([9],[10]) menyuguhkan aplikasi dengan nama HONQA, dimana antarmuka yang dibangun pada aplikasi ini sangat sederhana dan mudah untuk dimengerti. Hasil jawaban akan diinterpretasikan dalam bentuk memilih bahasa yang digunakan (Perancis atau Inggris), menentukan jenis jawaban dan medis yang diharapkan, menentukan jumlah dari jawaban yang akan ditampilkan.

Penelitian lain yang mengidentifikasi tentang penggunaan dan efektivitas/ efisiensi sumber biomedis *online* dalam menjawab pertanyaan medis/ klinis telah dilakukan oleh para peneliti medis dan praktisi. Seperti penelitian yang dilakukan oleh Wren [1] yaitu mencoba untuk meminimalkan waktu pencarian dan *browsing* serta memaksimalkan penggunaan pengetahuan dan data. Ely *et al.* [11] telah menemukan bahwa dokter menghabiskan waktu rata-rata 2 menit atau kurang dalam mencari jawaban, sementara Hersh *et al.* [13] telah menemukan bahwa waktu yang dibutuhkan rata-rata lebih dari 30 menit untuk mencari jawaban dalam hal perawatan kesehatan. Akibatnya, banyak pertanyaan klinis yang tidak terjawab. Studi mengenai hambatan untuk menemukan jawaban atas pertanyaan medis/ klinis ([11],[12]) ditemukan bahwa dokter memiliki 'keraguan untuk menjawab', waktu yang diperlukan berlebihan untuk proses pencarian, kesulitan merumuskan jawaban, ketidakpastian tentang pencarian strategi yang optimal, dan faktor kegagalan yang utama adalah sumber daya yang dipilih tidak memberikan solusi yang optimal.

Permasalahan mengenai kondisi waktu yang diperlukan berlebihan untuk proses pencarian, kesulitan merumuskan jawaban dan ketidakpastian tentang pencarian strategi yang optimal, maka diperlukan pemrosesan tahap awal yang optimal yaitu pada tahap pemrosesan pertanyaan ([11],[12]). Tahap pemrosesan pertanyaan yang optimal dapat dilakukan dengan cara klasifikasi pertanyaan seperti yang dilakukan oleh Ely *et al.* ([11],[12]), telah mengembangkan taksonomi generik umum untuk jenis pertanyaan klinis dan "*Evidence Taxonomy*" dari pertanyaan klinis berdasarkan pada studi yang mereka lakukan dalam bentuk perawatan primer. Pada tingkat atas *Evidence Taxonomy*, pertanyaan di klasifikasikan ke dalam klinis dan non-klinis. Pertanyaan klinis dibagi menjadi umum dan tertentu. Pertanyaan umum di klasifikasikan ke dalam *evidence* dan *Non-Evidence*. Pertanyaan *evidence* di klasifikasikan lebih lanjut ke kategori intervensi dan non-intervensi. Ely *et al.*, telah menyimpulkan

bahwa hanya pertanyaan *Evidence Taxonomy* berpotensi memiliki jawaban yang tepat.

Penelitian lain melakukan klasifikasi pertanyaan berdasarkan evaluasi manual dan evaluasi berdasarkan kerangka kerja PICO, seperti yang dilakukan oleh Huang *et al.* [7]. PICO frame awalnya dikembangkan untuk pertanyaan masalah terapi kemudian diperluas ke semua jenis pertanyaan klinis. Studi empiris telah menunjukkan bahwa penggunaan PICO frame meningkatkan kekhususan dan kejelasan konseptual masalah klinis, dan memberikan hasil pencarian yang lebih tepat. Sedangkan Kobayashi dan Shyu [4] berkonsentrasi pada klasifikasi pertanyaan menggunakan metode parsing yang berbeda dalam bentuk informasi semantik w/h dan w/o. Menurut Kobayashi dan Shyu klasifikasi taksonomi pertanyaan domain klinis merupakan langkah awal dalam mengembangkan sistem pengambilan jawaban waktu nyata. Mereka menunjukkan bahwa menggunakan jenis semantik UMLS dan algoritma mesin pembelajaran dapat meningkatkan kinerja klasifikasi pertanyaan klinis untuk kategori taksonomi generik.

Liu *et al.* [6] juga mengembangkan model mesin pembelajaran untuk klasifikasi otomatis antara pertanyaan konsumen dan pertanyaan profesional. Untuk mengevaluasi ketahanan model, Liu *et al.* menguji model yang di gunakan langsung oleh konsumen pada *PointCare* dataset untuk konsumen dan praktek *online* dataset. Sedangkan Patrick dan Li [5] menyatakan bahwa untuk melakukan klasifikasi harus mengikuti empat langkah utama yaitu pertama, satu set besar pertanyaan klinis dikumpulkan dari staf di Unit Perawatan Intensif; kedua, taksonomi pertanyaan klinis dirancang untuk tujuan tanya jawab; ketiga pedoman penjelasan dibuat dan digunakan untuk membubuhi keterangan dari set pertanyaan dan keempat, model klasifikasi multilayer dibangun untuk mengklasifikasikan pertanyaan-pertanyaan klinis.

Berdasarkan hal tersebut diatas, pada makalah ini akan dilakukan analisa berupa pengembangan penelitian dimasa mendatang dalam domain medis (*medical question answering* = MedQuAn). Kegunaan *survey* ini adalah untuk melihat peluang dan potensi yang akan digunakan untuk membangun sistem tanya jawab medis dengan menggunakan Bahasa Indonesia. Sehingga diharapkan hasil akhir dari *survey* makalah ini dapat memberikan sebuah kontribusi berupa metoda atau pendekatan yang paling tepat untuk sistem tanya jawab medis Bahasa Indonesia. Oleh karena itu analisa yang akan dilakukan menggunakan tiga komponen inti sistem tanya

jawab yaitu klasifikasi pertanyaan, temu balik informasi dan ekstraksi jawaban. Penelaahan model konseptual MedQuAn meliputi pengembangan metode yang efektif untuk ekstraksi jawaban, konstruksi taksonomi pertanyaan dan diskusi analitis model MedQuAn beserta kontribusi utama dari penelitian yang telah dilakukan, hasil eksperimen, dan keterbatasan penelitian.

Secara keseluruhan, makalah ini diorganisasikan kedalam bagian-bagian sebagai berikut, bagian 2 akan dijelaskan metoda yang digunakan untuk membangun sistem MedQuAn. Bagian 3 membahas arsitektur generik pada sistem MedQuAn. Bagian 4 berisi analisa komponen inti sistem MedQuAn (klasifikasi pertanyaan, temu balik informasi dan ekstraksi jawaban) yang dilakukan pada penelitian MedQuAn. Bagian 5, menganalisa peluang untuk pengembangan penelitian pada sistem tanya jawab medis menggunakan Bahasa Indonesia. Bagian 6 menyajikan sebuah ringkasan dari area riset MedQuAn.

2. Metoda yang digunakan pada MedQuAn

2.1 Pendekatan Awal untuk MedQuAn

Di antara peneliti sistem tanya jawab medis seperti Huang *et al.* [7], Yu *et al.* ([18],[19]), dan Kobayashi dan Shyu [4] telah melakukan penelitian bahwa klasifikasi pertanyaan sebagai langkah pertama menuju pengembangan sistem tanya jawab medis.

Huang *et al.* [7] meneliti kecukupan dan kesesuaian PICO sebagai kerangka representasi untuk pertanyaan klinis yang diajukan dalam

bahasa alami, dengan cara klasifikasi pertanyaan klinis untuk perawatan primer secara manual. Penelitian ini menegaskan kembali keseluruhan nilai kerangka PICO, tetapi juga menemukan bahwa PICO berpusat pada jenis pertanyaan terapi dan kurang cocok untuk jenis-jenis pertanyaan lainnya. Hal ini juga mencatat bahwa banyak UMLS jenis semantik menunjukkan asosiasi yang kuat dengan unsur-unsur PICO tertentu, sementara jenis semantik lainnya dapat dipetakan ke lebih dari satu slot PICO (Contoh antarmuka “Search MEDLINE/PubMed via PICO” dapat dilihat pada Gambar 1).

Yu dan Sable [20] mengembangkan komponen *filtering* pertanyaan yang secara otomatis menentukan apakah pertanyaan memiliki atau tidak jawaban, berdasarkan bukti taksonomi Ely *et al.* Mereka menggunakan berbagai algoritma *supervised machine learning*, dengan fitur *bag-of-word* dan fitur semantik yang terdiri dari konsep UMLS dan jenis semantik. Hasil telah menunjukkan bahwa menggabungkan fitur semantik secara umum cukup meningkatkan kinerja klasifikasi pertanyaan. Serta telah mengidentifikasi bahwa algoritma probabilistik pengindeksan menjadi algoritma terbaik dengan tingkat akurasi 80,5%.

Dalam sebuah studi tindak lanjut, Yu *et al.* [18] memfokuskan pada otomatisasi klasifikasi pertanyaan ke dalam kategori tertentu pada bukti taksonomi. Hasil evaluasi menunjukkan bahwa *Support Vector Machine* (SVM) melebihi semua sistem lain dalam banyak kasus. Hasil lain dari evaluasi mengungkapkan bahwa konsep UMLS dan jenis semantik sebagai fitur tambahan dapat meningkatkan hasil dalam banyak kasus.

Penelitian terbaru Yu dan Cao [21]

Protocol A:

Search MEDLINE/PubMed via PICO

Patient/Problem:

Intervention:

Compare to (leave blank if none):

Outcome (optional):

Age Group:
Not specified

Gender:
Not specified

Select Publication type:
Publication type

Protocol B:

Search MEDLINE/PubMed via PICO

Patient/Problem:

Intervention:

Compare to (leave blank if none):

Outcome (optional):

Age Group:
Not specified

Gender:
Not specified

Type of question:

therapy diagnosis etiology prognosis

specific search (narrow) sensitive search (broad)

OR

Systematic Reviews

OR

NO filters

Gambar 1. Contoh antarmuka Search MEDLINE/ PubMed via PICO (sumber : <http://pubmedhh.nlm.nih.gov.nlm/picostudy/>)

mengeksplorasi pendekatan *supervised machine learning* menggunakan SVM untuk otomatisasi proses mengelompokkan pertanyaan klinis menjadi topik yang umum, dan kedua pendekatan *supervised* seperti *Logistic Regression* dan *Conditional Random Fields (CRF)* dan pendekatan *unsupervised* seperti *Inverse Document Frequency (IDF) model* dan *Domain Filtering Inverse Document Frequency (IDF)* – dapat secara otomatis melakukan penggalian pertanyaan klinis berdasarkan kata kunci. Hasil evaluasi menggunakan koleksi pertanyaan klinis dengan menggunakan bahasa alami, telah menunjukkan bahwa pencocokan istilah pertanyaan pada konsep UMLS dan jenis semantik mengakibatkan peningkatan kinerja. Sedangkan untuk klasifikasi pertanyaan peningkatan kinerja terbukti dengan menggunakan SVM dan untuk domain penyaringan dengan pendekatan *unsupervised* untuk ekstraksi kata kunci. Hasil lainnya dari penelitian juga menunjukkan bahwa kedua pendekatan *supervised* mengungguli *unsupervised* untuk ekstraksi kata kunci tersebut, dan antara dua pendekatan *supervised* dihasilkan bahwa CRF mengungguli *Logistic Regression*.

Kobayashi dan Shyu [4] meneliti kinerja klasifikasi pertanyaan klinis dengan menggunakan representasi alternatif pertanyaan yang dihasilkan dengan menggunakan metode parsing yang berbeda dan ditambah dengan informasi mengenai konsep UMLS dan jenis semantik. Mereka menggunakan pertanyaan berlabel dari penelitian Ely *et al.*, dimana taksonomi dengan kategori informasi juga sebagai bentuk pertanyaan. Hasil penelitian menunjukkan bahwa menggunakan UMLS jenis semantik meningkatkan kinerja klasifikasi.

2.2 Pendekatan Berbasis Pengetahuan Non-Semantik QA Medis

Yu *et al.* ([19],[20],[21]) menggambarkan proses penerapan sistem QA medis mereka yaitu MedQA, yang menghasilkan jawaban berdasarkan level paragraf dari kedua koleksi MEDLINE dan web. MedQA memuat peringkasan teks dalam tahap pengolahan jawaban dari proses QA.

Klasifikasi pertanyaan dalam sistem MedQA dilakukan dengan menggunakan pendekatan *shallow syntactic parser* dan *standard IR engine*. Untuk ekstraksi jawaban, mereka menggunakan beberapa strategi untuk mengidentifikasi kalimat yang relevan, termasuk metode deteksi zona dokumen untuk artikel biomedis ([19],[20][21]). Dimana kalimat dikategorisasi dengan menggunakan frase isyarat [22] dan identifikasi

pola Lexico-sintaksis yang terdiri dari definisi kalimat. Untuk peringkasan teks, MedQA menggunakan *clustering* hirarkis [20] dan teknik *centroid-based summarization* [21].

Yu *et al.* menyadari perlunya menggunakan parser yang kuat dan akurat untuk domain spesifik. Mereka juga mencatat bahwa saat implementasi MedQA tidak dapat menangkap informasi semantik yang memainkan peran penting untuk kedua ekstraksi jawaban dan peringkasan teks.

Sang *et al.* [22] menggambarkan pekerjaan yang sedang berlangsung dalam mengembangkan sistem QA medis berbahasa Belanda. Mereka menggunakan dua strategi yang berbeda untuk ekstraksi informasi secara *offline*, salah satunya adalah pemanfaatan ensiklopedia medis berbahasa Belanda, dan yang lainnya menggunakan pola sintaksis di dasarkan pada hubungan ketergantungan untuk mengekstraksi *tuple* semantik. Analisis mereka dari hasil evaluasi menunjukkan bahwa kurangnya cakupan adalah sumber utama dari kesalahan dan bahwa pengetahuan ontologis dari domain akan sangat berguna dalam meningkatkan kinerja sistem QA.

2.3 Pendekatan Berbasis Semantik QA Medis .

Jacquemart *et al.* [23] mengembangkan QA medis bahasa Perancis dengan pendekatan berbasis semantik. Dalam studi kelayakan sistem QA medis, Jacquemart dan Zweigenbaum [24] meneliti masalah apakah dokumen yang relevan dengan pertanyaan medis dapat di temukan melalui pencarian Web dan apakah pertanyaan medis dapat di modelkan semantik dan dikategorikan dalam konseptual kerangka prototipe sistem QA mereka. Untuk tujuan penelitian, mereka menggunakan 100 pertanyaan klinis pada bedah mulut, yang masing-masing diubah menjadi bentuk kanonik dengan menyederhanakan pertanyaan kompleks menjadi pertanyaan yang lebih langsung atau dengan instansiasi pertanyaan yang tergantung pada konteks pertanyaan.

Mengenai isu pertama dari fokus mereka, Jacquemart dan Zweigenbaum telah menemukan Google adalah mesin pencari terbaik di Web. Namun, mengingat bahwa dari 100 pertanyaan hanya 60% diperoleh hasil yang relevan, mereka mencatat bahwa spesialisasi tinggi diperlukan untuk domain medis dan pertanyaan berorientasi klinis, digabungkan dengan sumber daya *online* bahasa Perancis yang lebih terbatas, dapat membatasi jumlah ketersediaan untuk menjawab pertanyaan.

Mengenai isu kedua, Jacquemart dan Zweigenbaum memodelkan bentuk untuk 100

pertanyaan medis sebagai pola *syntactico-semantic*, dalam rangka mengidentifikasi keteraturan dan menangkap konten semantik mereka. Pola-pola ini diperoleh dengan generalisasi bentuk kanonik untuk kategori domain spesifik yang generik. Mereka kemudian membangun model semantik pertanyaan, dalam bentuk triple semantik [*Concept*]-(*Relation*)-[*Concept*], dengan mengidentifikasi hubungan semantik yang relevan dalam UMLS *Semantic Network*. Mereka memperoleh 66 pola *syntactico-semantic* yang berbeda dan di kelompokkan menjadi delapan model semantik generik. Tiga dari model semantik, yang menyumbang 90 dari 100 pertanyaan dalam koleksi, sesuai dengan representasi tiga semantik [A]-(*R*)-[B] dengan modalitas “*which*”, “*does*”, atau “*why*” (lihat Tabel 1).

Tabel 1. Model semantik untuk pertanyaan medis [24]

[Which X]-(<i>R</i>)-[B]
[A]-(<i>R</i>)-[which Y]
Does [A]-(<i>R</i>)-[B]
Why [A]-(<i>R</i>)-[B]
[Which X,Y]-(<i>R</i>)-[B]
[Which X]-(<i>R</i>)-[B,C]
Duration [A]-(<i>precedes</i>)-[B]
Define [A]
Which specific precaution if [A]-(<i>R</i>)-[B]

Jacquemart dan Zweigenbaum mencatat bahwa mengotomatisasi konversi pertanyaan ke dalam bentuk kanonik membutuhkan penelitian lebih lanjut. Pemanfaatan hubungan semantik UMLS untuk tugas ini memerlukan kesesuaian antara persyaratan bahasa alami dan hubungan tersebut.

Niu *et al.* ([25],[26],[27]), mengusulkan pendekatan QA yang menempatkan jawaban dengan cara identifikasi, dengan menentukan peran semantik yang sesuai dengan empat bidang dalam frame PICO. Pendekatan ini didasarkan pada identifikasi empat peran yang diwakili oleh PICO dalam kandidat tanya jawab bahasa alami dan kemudian membandingkan peran pertanyaan dan kandidat jawaban yang sesuai dalam rangka untuk menentukan apakah ada calon jawaban yang benar. Dalam rangka menerapkan metode *role-based* dalam QA, Niu *et al.* mempertimbangkan masalah pendeteksian peran PICO diteks, menentukan batas tekstual masing-masing peran, dan mengidentifikasi hubungan antara peran yang berbeda, dengan fokus pertanyaan-pertanyaan yang berhubungan dengan terapi. Mereka mencatat bahwa dampaknya paling sulit untuk mendeteksi peran non-*Name Entity*.

Demner-Fushman *et al.* ([28], [29], [30], [31], [32], [33]) telah melakukan penelitian yang

sewarna dengan Niu *et al.*. Dalam pandangan mereka frame PICO sebagai pengorganisasian struktur inti pengetahuan untuk sistem QA medis dan QA klinis sebagai masalah penyatuan semantik antara PICO dengan kerangka permintaan dan jawaban. Lin dan Demner [34] menggambarkan ekstraktor pengetahuan semantik yang dapat digunakan sebagai komponen sistem QA klinis untuk mengidentifikasi elemen frame PICO dari MEDLINE untuk mengklasifikasikan tingkat *evidence*. Sebagai dasar untuk menentukan kualitas *evidence*, mereka menggunakan *Strength of Recommendations Taxonomy* (SORT) yang dikembangkan oleh Ebell *et al.* [60] dengan tingkat *evidence* A-, B-, dan C- sesuai dengan validitas tujuan dan kekuatan mereka.

Weiming *et al.* [35] mengusulkan pendekatan QA klinis, yang mencakup pengelompokan semantik, berdasarkan representasi semantik pertanyaan dan dokumen menggunakan konsep UMLS, jenis semantik, dan hubungan semantik. Pada tahap analisis pertanyaan, sistem Weiming *et al.* mengurai pertanyaan menggunakan MetaMap transfer (MMTx) dan Sem-Rep untuk mengidentifikasi konsep medis. Sistem ini menggunakan kata kunci berupa kata benda dan konsep aturan pemetaan untuk menafsirkan hubungan semantik dalam pertanyaan dan dokumen.

Konsep medis yang dihasilkan dalam tahap analisis pertanyaan, adalah sinonim, akronim, dan singkatan yang digunakan untuk mengambil dokumen yang relevan dan untuk memilih kandidat kalimat. Pada tahap ekstraksi jawaban, level jawaban yang dihasilkan dari calon kalimat oleh pemetaan jenis semantik dan hubungan dalam kandidat pertanyaan. Pada tahap pengelompokan semantik, jawaban di kelompokkan berdasarkan hubungan hirarkis di UMLS. Sistem ini berisi tiga jenis informasi untuk setiap jawaban: tipe semantik, konsep terkait, dan kalimat dari mana jawabannya berasal.

2.4 Pendekatan Berbasis Logika QA medis

Terol *et al.* [17] telah meneliti pendekatan berbasis logika, dalam mengadaptasi generik sistem QA medis. Sistem QA medis di rancang untuk menjawab pertanyaan bahasa alami dengan basis taksonomi generik pertanyaan klinis. Pengolahan QA dalam sistem ini didasarkan pada penurunan dari bentuk logika (*Logic Form*=LF) teks melalui penerapan teknik pemrosesan bahasa alami (*Natural Language Processing*=NLP).

Sistem QA medis Terol *et al.* terdiri dari empat modul pengolahan utama. Keempat tahapan proses

QA mengandalkan kalimat *preprocessing* dan penurunan bentuk logika serta NER medis dan pola pertanyaan.

Bentuk logika kalimat diperoleh dengan menerapkan aturan NLP dengan hubungan ketergantungan dari kata-kata dalam kalimat. Terol *et al.* menggunakan *broad-coverage parser* untuk mendapatkan hubungan ketergantungan. Setelah hubungan ketergantungan diperoleh, bentuk logika yang diperoleh dengan menerapkan dua macam aturan NLP ke pohon ketergantungan (*dependency tree*), penelusuran dimulai dari daun berlanjut untuk membentuk pohon dan berakhir di akar.

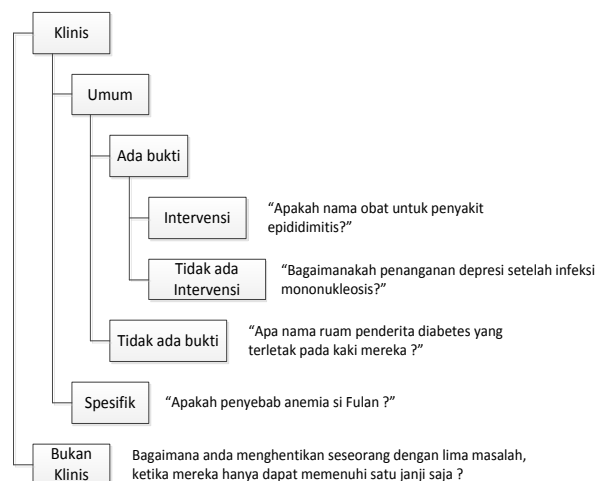
Inti dari sistem QA medis Terol *et al.* adalah modul untuk analisis pertanyaan. Tahap analisis pertanyaan terdiri dari klasifikasi pertanyaan dan analisis pertanyaan. Tahap klasifikasi pertanyaan terdiri dari penurunan bentuk logika pertanyaan, ekstraksi dari bentuk logika kata kerja, NER (*named entity recognition*) medis, perhitungan *medical entities score in question* (MESQ) melalui analisis bentuk pertanyaan, pencocokan kata kerja utama dan entitas medis pada bentuk logika pertanyaan dengan kata kerja dan entitas medis pada pola pertanyaan generik, dan akhirnya pemilihan pola yang paling memenuhi kriteria akan dipilih. Tahap analisis pertanyaan terdiri dari menentukan semantik dari pertanyaan dengan menggunakan WordNet dan UMLS metathesaurus, pengenalan jenis jawaban yang diharapkan disesuaikan dengan klasifikasi jenis jawaban pada 10 jenis pertanyaan generik, dan identifikasi kata kunci melalui penerapan heuristik untuk predikat dan hubungan antara predikat dalam bentuk logika pertanyaan.

Akhirnya, proses ekstraksi jawaban terdiri dari penurunan bentuk logika calon jawaban, identifikasi kata kerja utama, perbandingan dari kata kerja utama dengan himpunan kata kerja yang sesuai dengan pertanyaan generik, pengenalan elemen jaringan medis, verifikasi apakah elemen jaringan medis sesuai dengan yang diharapkan oleh pertanyaan, dan analisis predikat yang berkaitan jawaban kandidat, kata kerja utama, dan sisa elemen jaringan medis dalam jawaban LF. Terol *et al.* mengevaluasi modul analisis pertanyaan memperoleh 94,4% presisi pada 250 pertanyaan.

3. Arsitektur Sistem MedQuAn

Pada bagian ini, akan dibahas mengenai arsitektur pada sistem tanya jawab medis. Teknologi untuk sistem tanya jawab medis bergantung pada penyesuaian yang relevan sehingga akan hanya memberikan sebuah jawaban yang terukur sesuai dengan interpretasi semantik dari pertanyaan.

Sebuah paradigma yang dominan di bidang medis adalah *Evidence-Based Medicine* (EBM) yang dikembangkan oleh Sacket *et al.*[14]. EBM mengacu pada penggunaan bukti terbaik yang diperoleh dari penelitian ilmiah dalam membuat keputusan klinis. Dalam kerangka EBM, dokter didesak untuk mengajukan pertanyaan dan menemukan bukti yang terbaik. Salah satu metoda yang digunakan adalah taksonomi pertanyaan, seperti yang dilakukan oleh Bergus *et al.* [15]. Bergus telah mengembangkan taksonomi pertanyaan medis sesuai dengan unsur-unsur pertanyaan PICO dan pertanyaan dengan kategori klinis. Sedangkan Ely *et al.* ([16],[11],[12]) mengembangkan taksonomi generik umum untuk jenis pertanyaan klinis dan "*Evidence Taxonomy*" dari pertanyaan klinis berdasarkan pada studi yang mereka lakukan dalam bentuk perawatan primer. *Evidence Taxonomy* yang dibuat Ely *et al.* dapat dilihat pada Gambar 2. Ely *et al.* ([11],[12]) mengidentifikasi 10 jenis yang paling umum dari pertanyaan klinis generik seperti pada Tabel 2.



Gambar 2. *Evidence Taxonomy* untuk pertanyaan klinis [16]

Penelitian lain melakukan analisis pertanyaan dengan cara mengekstraksi informasi yang dibutuhkan dari pertanyaan. Seperti yang dilakukan oleh Cao *et al.* ([2],[3]) menyajikan arsitektur AskHERMES (lihat Gambar 3). Pada sistem ini analisis pertanyaan secara otomatis mengekstrak informasi yang dibutuhkan dari pertanyaan dan jawaban berupa daftar *query*. Modul Pertanyaanyang terkait akan di ekstraksi dengan mengembalikan daftar pertanyaan serupa. Temu balik informasi (*information retrieval*) mengembalikan dokumen yang relevan yang telah di indeks secara lokal. Sedangkan ekstraksi informasi bertugas mengidentifikasi bagian-bagian yang relevan dari sebuah dokumen.

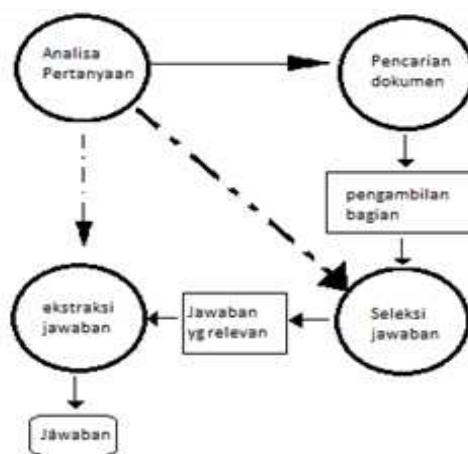
Presentasi jawaban agregat menghilangkan informasi yang berlebihan, secara otomatis untuk menghasilkan ringkasan terstruktur dan menyajikan ringkasan kepada pengguna.

Tabel 2. Sepuluh jenis pertanyaan medis yang paling umum ([11],[12])

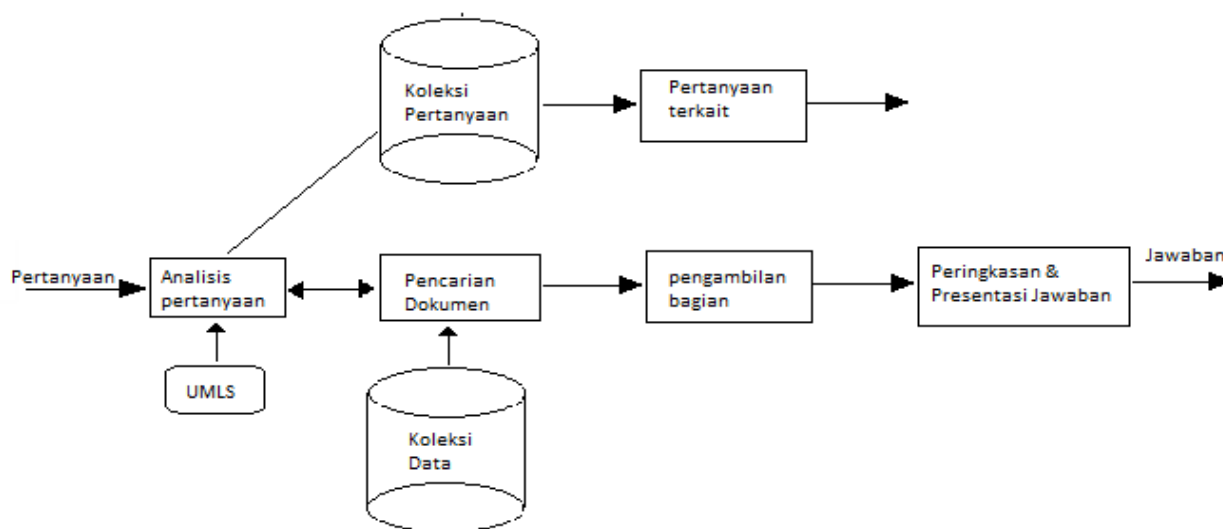
Apakah obat yang dapat diminum untuk kondisi X ? Apakah penyebab dari gejala X ? Apakah indikasi hasil tes menunjukkan situasi X ? Apakah dosis obat X? Bagaimana seharusnya saya mengobati kondisi X ? Bagaimana seharusnya saya mengelola kondisi X ? Apakah penyebab dari temuan fisik X? Apakah penyebab dari temuan hasil tes X? Dapatkan obat X menyebabkan Y? Dapatkan pasien berada pada kondisi X ?

Penelitian lain melakukan analisis pertanyaan dengan cara mengekstraksi informasi yang dibutuhkan dari pertanyaan. Seperti yang dilakukan oleh Cao et al. ([2],[3]) menyajikan arsitektur *AskHERMES* (lihat Gambar 3). Pada sistem ini analisis pertanyaan secara otomatis mengekstrak informasi yang dibutuhkan dari pertanyaan dan jawaban berupa daftar query. Modul pertanyaan yang terkait akan diekstraksi dengan mengembalikan daftar pertanyaan serupa. Temu balik informasi (information retrieval) mengembalikan dokumen yang relevan yang telah di indeks secara lokal. Sedangkan ekstraksi informasi bertugas mengidentifikasi bagian-bagian yang relevan dari sebuah dokumen. Presentasi jawaban agregat menghilangkan informasi yang berlebihan, secara otomatis untuk menghasilkan ringkasan terstruktur dan menyajikan ringkasan kepada pengguna.

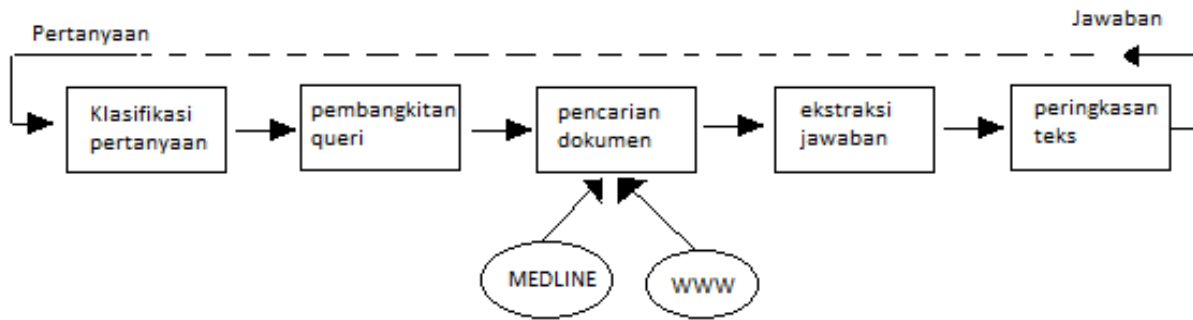
Terol et al. [17] melakukan penelitian dengan menggunakan pendekatan berbasis logika untuk mengadaptasi generik sistem tanya jawab medis. Sistem tanya jawab medis yang dirancang adalah untuk menjawab pertanyaan bahasa alami, dimana pertanyaan yang paling sering berupa pertanyaan taksonomi generik klinis. Pengolahan tanya jawab dalam sistem ini didasarkan pada derivasi bentuk logika dari teks melalui aplikasi NLP (Natural Language Processing). Sistem tanya jawab medis Terol et al. terdiri dari empat modul proses utama seperti yang ditunjukkan pada Gambar 4. Keempat tahapan utama proses tanya jawab pada penelitian Terol ini mengandalkan kalimat preprocessing dan derivasi bentuk logika serta pada pengenalan nama entitas (named entity recognition=NER) medis dan pola pertanyaan.



Gambar 4. Sistem Arsitektur QA medis[17].



Gambar 3. Arsitektur AskHERMES ([2],[3])



Gambar 5. Arsitektur MedQA ([18],[19])

Studi lain yang seperti yang dilakukan oleh Yu *et al.* ([18],[19]) menggambarkan implementasi sistem tanya jawab medis dengan nama MedQA, aplikasi ini menghasilkan tingkat jawaban paragraf dari koleksi MEDLINE dan web. Sistem dalam implementasi saat ini dengan pendefinisian pertanyaan (misalnya, "Apakah X?"). MedQA memuat ringkasan teks yang diolah pada tahap pengolahan jawaban. Sistem arsitektur MedQA dapat dilihat pada Gambar 5.

Berdasarkan paparan dari beberapa arsitektur MedQuAn di atas maka dapat disimpulkan bahwa komponen inti untuk membangun sebuah sistem tanya jawab medis harus terdiri dari analisa dan klasifikasi pertanyaan, pencarian dokumen yang sesuai dengan kata kunci yang diberikan oleh proses klasifikasi pertanyaan dan tahap akhir adalah ekstraksi jawaban yang dibutuhkan untuk memilih jawaban yang sesuai dengan pertanyaan. Rincian tiga fase inti pada sistem tanya jawab MedQuAn akan dijelaskan pada bagian 4.

4. Tiga Fase Inti pada MedQuAn

Pada bagian ini kami akan membahas secara singkat tiga fase inti dan tren saat ini secara keseluruhan dalam penelitian MedQuAn, dimana fase utama ini terdiri dari klasifikasi pertanyaan, pencarian dokumen dan ekstraksi jawaban.

4.1 Klasifikasi Pertanyaan MedQuAn

Banyak penelitian dalam klasifikasi pertanyaan lebih memfokuskan pada semantik pertanyaan dan potensi jawabannya. Pertanyaan umumnya sesuai dengan pola bahasa yang dapat di prediksi, dan karena itu diklasifikasikan berdasarkan taksonomi Patric dan Li [5] menggunakan klasifikasi pertanyaan multilayer dengan pendekatan pembelajaran mesin (SVM-Support Vector Machine). Mereka mengadopsi sebagai pendekatan klasifikasi standar untuk secara otomatis mengelompokkan pertanyaan menurut taksonomi pertanyaan dan template pertanyaan generik.

Selain itu, validasi sebanyak dua kali di pilih sebagai mekanisme evaluasi dengan menghitung akurasi keseluruhan dan nilai individual F-score. Belajar dari seleksi fitur adalah salah satu masalah yang paling penting untuk mempengaruhi kinerja pembelajaran mesin. Fitur set yang dipilih di sini melibatkan lima set utama fitur yang dapat diperpanjang untuk menciptakan sebelas fitur set jika diperlukan yaitu (i) Unigram: setiap token dalam sebuah pertanyaan, (ii) Bigram kelompok dua token dalam pertanyaan; (iii) Kata interogatif: tanda pertama atau dua token pertama dalam sebuah pertanyaan yang biasanya merupakan jenis jawaban; (iv). Kategori SNOMED: kategori top SNOMED setiap istilah medis dalam pertanyaan, (v). Struktur Argumen

Predikat (PAS). Liu *et al.* [6], mengembangkan pendekatan pembelajaran mesin secara otomatis dapat mengklasifikasikan pertanyaan medis ke jenis pertanyaan yang ditentukan oleh 'evidency taksonomi'. Dengan menggunakan total 200 pertanyaan bernotasi, sepuluh kinerja cross-validation menunjukkan akurasi lebih dari 80% untuk menangkap pertanyaan dengan jawaban yang dapat dibuktikan.

Beberapa sistem telah meneliti penggunaan fitur sintaks untuk klasifikasi tetapi umumnya telah melakukannya hanya sebagai suplemen semantik bukan sebagai pengganti. Cao *et al.* ([2],[3]) dan Lee *et al.* [36], mengeksplorasi penggunaan pembelajaran mesin dengan pendekatan secara otomatis dapat mengklasifikasikan pertanyaan klinis. Cao *et al.* ([2],[3]) bereksperimen dengan beberapa algoritma pembelajaran mesin yang biasa digunakan untuk klasifikasi pertanyaan, termasuk Naive Bayes, pohon keputusan, dan *support vector machine* (SVM), dan hasil dari 10 kali lipat *cross-validation* menunjukkan bahwa SVM melakukan klasifikasi yang terbaik.

Sementara, Lee *et al.*[36] mengeksplorasi pendekatan SVM secara otomatis mengklasifikasikan pertanyaan klinis dalam kategori taksonomi yang di buat oleh Ely *et al.*

[11]. Dengan menggunakan 200 pertanyaan, kinerja Lee menunjukkan akurasi lebih dari 80% dalam 10 kali lipat *cross-validation* untuk mengklasifikasikan pertanyaan dalam kategori yang ditentukan oleh taksonomi.

Studi-studi lain telah mengidentifikasi dimensi tambahan yang dapat berguna untuk klasifikasi pertanyaan, misalnya perbedaan antara pertanyaan-pertanyaan faktual dan analitis. Cao et al. [2], mengklasifikasikan pertanyaan menjadi 12 topik umum untuk memudahkan pencarian informasi. Topik tersebut meliputi perangkat, diagnosis, epidemiologi, etiologi, sejarah, manajemen, farmakologi, temuan fisik, prosedur, prognosis, tes dan pengobatan & pencegahan, yang telah digunakan untuk membubuhi keterangan 4654 pertanyaan klinis yang dicatat oleh dokter. Gobeil et al. [8] membuat kategorisasi pertanyaan dalam menganalisis pertanyaan untuk menemukan

pertanyaan sesuai dengan target yang ditetapkan dan untuk menemukan query yang akan digunakan untuk temu balik informasi. Kategorisasi pertanyaan menggunakan Government and Binding parser, FIPS, yang menghasilkan representasi sintak yang mendalam dari pertanyaan.

Kategorisasi pertanyaan berlaku pada pola yang spesifik ke pola yang lebih generik untuk mengidentifikasi target yang telah ditetapkan. Pola paling umum untuk memilih frase kata benda (NP) ditemukan setelah wh (misalnya What) dan sebelum kata kerja bantu. NP-parsing untuk mengidentifikasi konstituennya (misalnya head vs adjectives). Kemudian, normalisasi kategori mencoba untuk mengasosiasikan NP terpanjang ke daftar 450 jenis target yang telah ditetapkan.

Sebuah ringkasan singkat dari pendekatan klasifikasi pertanyaan disajikan dalam Tabel 3.

Tabel 3. Matriks Metoda Pada Tahap Klasifikasi Pertanyaan

Metoda Yang Digunakan		Spesifikasi Hasil :
		Feature/ category/ template/taxonomy/ subclass.
Patric and Li[5]	Support Vector Machines (SVM)	Unigram; Bigram; Interrogative word; SNOMED category; Predicate argument structure (PAS).
Cao et al.[2]	Supervised machine-learning : Naïve Bayes, Decision tree, and Support vector machines (SVMs),	n-grams, part-of-speech (POS) and stemming
Lee et al.[36]	Supervised machine-learning	Five hierarchical categories : <i>Clinical or Nonclinical; General versus Specific; Evidence and No-evidence; Intervention versus Nointervention.</i>
Cao et al.[3]	Machine learning	Question template (e.g., “What-type” and “How-type”)
Yu et al.[14]	Supervised machine-learning	Evidence taxonomy
Gobeil et al.[8]	Government and Binding parser, FIPS	<ul style="list-style-type: none"> • Wh-type questions • Pattern : NP-Wh- auxiliary verb

4.2 Pencarian Dokumen MedQuAn

Tujuan dari pencarian dokumen adalah untuk mengambil hasil yang akurat sebagai tanggapan atas permintaan yang diajukan oleh pengguna, dan untuk peringkat hasil ini disesuaikan dengan relevansi yang dibutuhkan oleh pengguna. Sistem IR berdasarkan *cosine similarity* dengan cara mengembalikan dokumen berdasarkan semua kata kunci yang diperoleh pada saat melakukan klasifikasi pertanyaan.

Gobeil et al. [8] menyatakan bahwa saat ini pengambilan dokumen dapat dilakukan melalui dua mesin pencari PubMed, adalah mesin pencari NLM yang diakses melalui e-utilisasi, dan EasyIR adalah mesin pencari lokal di MEDLINE. EasyIR didasarkan pada versi pengindeksan secara lokal

dari MEDLINE dengan menggunakan kombinasi linear dari vektor ruang model pencarian yang menunjukkan efektivitas kompetitif di TREC-Genomic. PubMed mengembalikan daftar dokumen dengan urutan kronologis berdasarkan peringkat, sementara EasyIR mengembalikan daftar dokumen dengan nilai kesamaan peringkat. Untuk versi online kedua mesin pencari, jumlah K dokumen diambil diatur ke dalam urutan 50 dokumen.

Sedangkan Lee et al. [36] menggambarkan pertanyaan Term Generation dan dokumen Retrieval diberlakukan dengan menggunakan LT CHUNK untuk mengidentifikasi frase kata benda dari pertanyaan medis dan kemudian menerapkan frase nominal sebagai istilah permintaan untuk

mengambil dokumen yang relevan. Lee menerapkan alat Lucene untuk mengindeks koleksi MEDLINE, dari mana kita mengambil dokumen yang relevan dengan menggunakan query. Yu et al. [14], menerapkan Google sebagai mesin pencari untuk mengambil dokumen Web. Mereka menerapkan alat pengindeksan Lucene untuk mengindeks dokumen dari koleksi teks lokal (yaitu abstrak MEDLINE).

Studi-studi lain telah mengidentifikasi proses temu balik untuk menentukan pendefinisian pertanyaan dan menentukan nilai dari jawaban secara praktis. Secara khusus, Cao et al. [2] mendefinisikan bagian dalam AskHERMES sebagai satu atau lebih kalimat yang berdekatan, di mana setiap kalimat menggabungkan satu atau lebih istilah permintaan dari pertanyaan. Pendekatan mereka berbeda dari *TextTiling*, metode populer untuk segmentasi multi-level, bahwa pertanyaan yang diajukan memainkan peranan penting untuk pengenalan bagian dalam sistem mereka. Sebuah ringkasan singkat dari temu balik informasi disajikan pada Tabel 4.

Tabel 4. Matriks Pencarian Dokumen

Search Engine	Sumber Corpus
Gobeil et al. [8]	Pub Med and Easy IR
Yu et al. [14]	Google
Cao et al. [2]	Pub Med
Lee et al. [32]	Lucene

4.3 Ekstraksi Jawaban MedQuAn

Ekstraksi jawaban mengidentifikasi dokumen melalui kalimat yang relevan untuk menjawab pertanyaan. Beberapa sistem telah dieksplorasi untuk melakukan ekstraksi informasi, seperti yang dilakukan oleh Lee et al. [36] yaitu menerapkan sistem ekstraksi informasi yang kuat dengan nama Autoslog-TS untuk menghasilkan secara otomatis pola Lexico-sintaksis. Mereka secara otomatis mengidentifikasi pola Lexico-sintaksis, yaitu pola yang menggabungkan leksikon dan informasi sintaks untuk mengidentifikasi definisi kalimat. Strategi Lee et al. adalah untuk mendapatkan daftar lengkap dari pola Lexico-sintaksis yang telah dihasilkan dari satu set definisi kalimat. Yu et al. [4], mengembangkan beberapa strategi untuk mengidentifikasi kalimat yang relevan. MedQA merupakan aplikasi pertama yang mengklasifikasikan kalimat ke jenis tertentu. Selain itu, MedQA membuat kategori kalimat berdasarkan fitur linguistik dan sintaksis.

Penelitian lain seperti Cao et al. [3], mengembangkan peringkasan dan pendekatan

presentasi tanya jawab yang dirancang berdasarkan teknik *clustering*. Model ekstraksi kata kunci otomatis mereka secara efektif mengekstrak kata kunci dari pertanyaan dan kata kunci tersebut kemudian dapat digunakan untuk secara hirarkis sebagai struktur jawaban ringkas. Patric dan Li [5], melakukan klasifikasi taksonomi pada sistem tanya jawabnya. Langkah ini merupakan otomatisasi dari proses penemuan pengetahuan (KD=*knowledge discovery*) dan penggunaan kembali pengetahuan (KR=*Knowledge Retrieval*) yang berasal dari alur kerja dalam karya-karya mereka. Sebagai kesamaan dapat ditemukan di antara kelas kecil dalam sistem tanya jawab (misalnya 'Perbandingan', 'Pengambilan Keputusan', 'Terstruktur', 'General', 'Alasan') dengan mengeksplorasi fitur tingkat permukaan atas. Set fitur yang spesifik digunakan untuk memperpanjang elemen indikatif dengan menjelajahi sinonim dan antonim dalam WordNet.

Sebuah ringkasan singkat dari ekstraksi jawaban disajikan pada Tabel 5.

Kebanyakan pendekatan tanya jawab medis dalam studi literatur ini, lebih dan kurangnya memanfaatkan informasi semantik untuk seluruh pengolahan pertanyaan, pengolahan dokumen, dan pengolahan jawaban pada tahapan proses tanya jawab pada domain tertentu. Seperti yang diharapkan, mengingat fakta bahwa karakteristik utama sistem tanya jawab domain terbatas menyangkut pemanfaatan pengetahuan semantik pada sumber daya domain tertentu. Dapat dikatakan bahwa penelitian lanjutan tentang efektifitas penggabungan pengetahuan semantik dalam proses tanya jawab sangat diperlukan.

Tabel 5. Metoda yang digunakan Pada Tahap Ekstraksi Jawaban

Metoda Yang Digunakan	Hasil Ekstraksi Jawaban
Lee et al. [36]	Robust information extraction system
Yu et al. [14]	Multiple strategies
Cao et al. [2]	Clustering technique
Patric & Li [5]	Answerable question taxonomy classification

5. Peluang Pengembangan MedQuAn Menggunakan Bahasa Indonesia

Hasil akhir dari survei ini adalah telah ditemukan bahwa potensi tertinggi dalam menggunakan metoda pendekatan MedQuAn adalah *machine learning*, *support vector machine* (SVM), *supervised machine-learning*, *clustering technique* dan *taxonomy classification* untuk pengenalan jawaban dan kesetaraan/ekivalensi pertanyaan belum sepenuhnya dieksplorasi. *Smart model driven* untuk menemukan pencocokan struktural dan cara menggunakan informasi semantik dalam ekstraksi jawaban masih terbuka untuk penelitian masa depan.

Kegiatan progresif akan terus menerus dilakukan dan akumulasi dari penelitian sistem tanya jawab MedQuAn menunjukkan adanya beberapa keterbatasan sehingga hal ini perlu dicarikan cara untuk bagaimana menemukan hal yang baru, dihadapi dan ditangani oleh peneliti lain. Hal inilah yang menjadi tantangan terbesar dalam mengembangkan MedQuAn khususnya untuk bahasa Indonesia.

Penelitian tanya jawab MedQuAn untuk bahasa Indonesia dapat memanfaatkan semantik yang diperoleh dari berbagai situs Web yang berhubungan dengan kesehatan, dengan konsep UMLS (*Unified Medical Language System*) berdasarkan jenis semantik dan hubungan semantik. Sistem MedQuAn ini dapat dilakukan dengan mencari informasi melalui *tag corpus* berdasarkan POS (*Part of Speech*) dengan mencari frase kata benda. Kemudian *tag* frase kata benda dengan konsep metathesaurus dan terkait jaringan semantik berdasarkan jenis semantik. Sistem tanya jawab medis juga dapat menggunakan pola POS untuk mencari bentuk terstruktur [Subjek]-[Kata Kerja]-[Pelengkap], untuk mendeteksi jenis semantik. Jenis semantik yang dihubungkan oleh hubungan semantik akan menghasilkan klausa dengan hubungan semantiknya yang melengkapi proses *tagging*.

Selain itu, fakta bahwa beberapa terminologi dan ontologis sumber daya yang tersedia dalam domain medis terstruktur dapat diakses secara formal berbasis logika. Hal ini menunjukkan relevansi dan kelayakan pendekatan untuk mengeksplorasi sistem tanya jawab medis berbasis logika.

6. Kesimpulan

Dalam makalah ini, telah dilakukan survei terhadap sistem tanya jawab dengan domain medis. Makalah ini memiliki cara baru untuk memahami bahwa penelitian pada domain terbatas (seperti

domain medis, biomedis, dan klinis) ini membutuhkan kerja keras yang sangat signifikan. Survei ini dilakukan terhadap makalah hasil penelitian yang dipublikasikan pada penerbit dan memiliki *top-cited* yang tinggi pula dalam bidang sistem tanya jawab MedQuAn. Peluang dan tantangan dalam mengembangkan MedQuAn khususnya untuk bahasa Indonesia sangat besar. Dengan keragaman pengaruh bahasa daerah sehingga pemahaman semantik dan sintaks pun menjadi tantangan tersendiri.

Penelitian yang akan dilakukan selanjutnya berdasarkan hasil survei makalah ini adalah pembangunan sistem tanya jawab medis berbahasa Indonesia dengan pendekatan pembangkitan pertanyaan. Adapun tujuan penggunaan pendekatan ini adalah menghasilkan pola pasangan pertanyaan dan jawaban yang berasal dari kalimat, teks, dan representasi semantik. Dengan adanya pola pasangan pertanyaan dan jawaban diharapkan proses tanya jawab dengan menggunakan bahasa alami dalam domain medis akan lebih efisien dan efektif.

Daftar Pustaka

- [1] Wren JD., "Question answering systems in biology and medicine—the time is now". *Bioinformatics Journal*, 27(14):2025–2026, 2011
- [2] Y.G. Cao, F. Liu, P. Simpson, L. Antieau, A. Bennett, James, Cimino, J. Ely, Hong Yu., "AskHERMES: An online question answering system for complex clinical questions". *Journal of Biomedical Informatics* 44. 277–288, 2011.
- [3] Yong-gang Cao, James J. Cimino, John Ely, Hong Yu., "Automatically extracting information needs from complex clinical questions". *Journal of Biomedical Informatics*, 43, 962–971, 2010.
- [4] T. Kobayashi, C.-R. Shyu., "Representing clinical questions by semantic type for better classification". In Proceedings of the AMIA 2006 Symposium, AMIA, p. 987,2006.
- [5] Jon Patrick, Min Li., "An ontology for clinical questions about the contents of patient notes". *Journal of Biomedical Informatics* 45, 292–306, 2012.
- [6] Feifan Liu, Lamont, Antieau, Hong Yu., "Toward automated consumer question answering: Automatically separating consumer questions from professional questions in the healthcare domain". *Journal of Biomedical Informatics* 44,1032-1036, 2011.
- [7] X. Huang, J. Lin, D. Demner-Fushman., "Evaluation of PICO as a knowledge representation for clinical questions:.. In Proceedings of the AMIA 2006 Symposium, AMIA, 2006, pp. 359–363, 2006.
- [8] Gobeill J, Tbahriti I, Ehrler F, Ruch P., "Vocabulary-driven passage retrieval for question-

- answering in genomics". In Proceedings of the 16th text retrieval conference. Maryland, USA: TREC, National Institute of Standards and Technology (NIST) 2007.
- [9] Cruchet S, Gaudinat A, Rindfleisch T, Boyer C., "What about trust in the question answering world?", San Francisco, USA: AMIA Annual Symposium 2009.
- [10] Cruchet S, Gaudinat A, Boyer C., "Supervised approach to recognize question type in a QA system for health". *Stud Health Technol Inform* 2008, 136(136):407–412, 2008.
- [11] J.W. Ely, J.A. Osheroff, M.H. Ebell, M.L. Chambliss, D.C. Vinson, J.J. Stevermer, E.A. Pifer., "Obstacles to answering Doctors' questions about patient care with evidence: qualitative study", *Br. Med. Journal.* 324, 710–716. 2002,
- [12] J.W. Ely, J.A. Osheroff, M.L. Chambliss, M.H. Ebell, M.E. Rosenbaum., "Answering physicians' clinical questions: obstacles and potential solutions". *J. Am. Med. Inform. Assoc.* 12, 217–224., 2005,
- [13] W.R. Hersh, M.K. Crabtree, D.H. Hickman, L. Sacherek, C.P. Friedman, P. Tidmarsh, C. Mosbaek, D. Kraemer., "Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions". *J. Am. Med. Inform. Assoc.* 9, 283–293, 2002.
- [14] D.L. Sackett, S. Strauss, W. Richardson, W. Rosenberg, R. Haynes, "Evidence-Based Medicine: How to Practice and Teach EBM", 2nd ed., Churchill Livingstone, Edinburgh, UK; New York, USA, 2000,
- [15] G.R. Bergus, C.S. Randall, S.D. Sinift, D.M. Rosenthal, "Does the structure of clinical questions affect the outcome of curbside consultations with specialty colleagues?" *Arch. Fam. Med.* 9 541–547, 2000.
- [16] J.W. Ely, J.A. Osheroff, M.H. Ebell, G.R. Bergus, B.T. Levy, M.L. Chambliss, E.R. Evans, "Analysis of questions asked by family doctors regarding patient care", *Br. Med. J.* 319, 358–361, 1999.
- [17] R.M. Terol, P. Martinez-Barco, M. Palomar, "A knowledge based method for the medical question answering problem", *Comput. Biol. Med.* 27, 1511–1521, 2007.
- [18] H. Yu, "Towards answering biological questions with experimental evidence: automatically identifying text that summarize image content in full-text articles", In Proceedings of the AMIA 2006 Symposium, AMIA, pp.834–838, 2006
- [19] H. Yu, C. Sable, H.R. Zhu, "Classifying Medical Questions based on an Evidence Taxonomy", in: Proceedings of the AAAI-05 Workshop Question Answering in Restricted Domains, AAAI, Pittsburgh, PA, USA, 2005.
- [20] H. Yu, C. Sable, "Being Erlang Shen: identifying answerable questions", in: Proceedings of the IJCAI'05 Workshop Knowledge and Reasoning for Answering Questions (KRAQ'05), Edinburgh, UK, pp. 6–14, 2005.
- [21] H. Yu, Y. Cao, "Automatically extracting information needs from ad hoc clinical questions", in: Proceedings of the AMIA 2008 Symposium, pp. 96–100, 2008.
- [22] E.T.K. Sang, G. Bouma, M. de Rijke, "Developing offline strategies for answering medical questions", in: Proceedings of the AAAI-05 Workshop Question Answering in Restricted Domains, AAAI, Pittsburgh, PA, USA, 2005.
- [23] P. Jacquemart, P. Zweigenbaum, "Towards a medical question-answering system: a feasibility study", *Stud. Health. Technol. Inform.* 95, pp. 463–468, 2003.
- [24] T. Delbecque, P. Jacquemart, P. Zweigenbaum, "Indexing UMLS semantic types for medical question-answering", in: R. Engelbrecht (Ed.), *Connecting Medical Informatics and Bio-Informatics (ENMI 2005)*, pp. 805–810, 2005.
- [25] Y. Niu, X. Zhu, G. Hirst, "Using outcome polarity in sentence extraction for medical question-answering", in: Proceedings of the AMIA 2006 Symposium, AMIA, pp. 599–603, 2006.
- [26] Y. Niu, G. Hirst, G. McArthur, P. Rodriguez-Gianolli, "Answering clinical questions with role identification", in: Proceedings of the ACL-2003 Workshop Natural Language Processing in Biomedicine, ACL, Sapporo, Japan, pp. 73–80, 2003.
- [27] Y. Niu, X. Zhu, G. Hirst, "Using outcome polarity in sentence extraction for medical question-answering", in: Proceedings of the AMIA 2006 Symposium, AMIA, pp. 599–603, 2006.
- [28] D. Demner-Fushman, B. Few, S.E. Hauser, G. Thoma, "Automatically identifying health outcome information in MEDLINE records", *J. Am. Med. Inform. Assoc.* 13, p : 52–60, 2006a
- [29] D. Demner-Fushman, J. Lin, "Knowledge extraction for clinical question answering: preliminary results", in: Proceedings of the AAAI-05 Workshop Question Answering in Restricted Domains, AAAI Press, Pittsburgh, PA, USA, 2005.
- [30] D. Demner-Fushman, J. Lin, "Answer extraction semantic clustering, and extractive summarization for clinical question answering", in: Proceedings of the 21st Int'l Conf. Comp. Ling. and 44th Ann. Meeting Assoc. Comp. Ling. (COLING-ACL 2006), Sydney, Australia, pp. 841–848, 2006b.
- [31] D. Demner-Fushman, J. Lin, "Answering clinical questions with knowledge-based and statistical techniques", *Comput. Linguist*, pp: 93–103, 2007
- [32] D. Demner-Fushman, S.E. Hauser, S.M. Humphrey, G.M. Ford, J.L. Jacobs, G.R. Thoma, "MEDLINE as a source of just-in-time answers to clinical questions", in: Proceedings of the AMIA 2006 Symposium, AMIA, pp. 190–194, 2006.
- [33] X. Huang, J. Lin, D. Demner-Fushman., 2006, "Evaluation of PICO as a knowledge representation for clinical questions". In Proceedings of the AMIA 2006 Symposium, AMIA, pp. 359–363, 2006.
- [34] J. Lin, D. Karakos, D. Demner-Fushman, S. Khudanpur, "Generative content models for structural analysis of medical abstracts", in:

Proceedings of the HLT-NAACL 2006 Workshop on Linking Natural Language Processing and Biology: Towards Deeper Biological Literature Analysis (BioNLP'06), pp. 65–72, 2006.

- [35] W. Weiming, D. Hu, M. Feng, L. Wenyin, “Automatic clinical question answering based on UMLS relations”, in: 3rd Int'l Conf. Semantics, Knowledge and Grid (SKG 2007), accepted Xi'An, China, 2007.

- [36] M. Lee, J. Cimino, H.R. Zhu, C. Sable, V. Shanker, J. Ely, H. Yu, “Beyond information retrieval—medical question answering”, in: Proceedings of the AMIA 2006 Symposium, AMIA, pp. 469–473, 2006.

Model Pengukuran Proses Bisnis CRM Berbasis eTom dan IT-IL

A Measurement Model of CRM Business Based on eTom and IT-IL

Lihardo Ranjaliba Saragih, Gendut Suprayitno, Dahmir Dahlan

Institut Sains dan Teknologi Nasional

Jl. Moch Kahfi II, Jakarta 12640, Indonesia

Email: ranjaliba@gmail.com; gendut.suprayitno@gmail.com; dddkamang@yahoo.com

Abstract

Nowadays Telecom Company doing business model transformation through Manage services, to reduce cost. Telecommunications companies obviously attached to cutting-edge technology and capital-intensive, therefore needs a different approach in Order to remain competitiveness, not just a simple outsourcing. It made paradigm of outsourcing in global Telecommunications industry changed, in term of Manage service. Company realizes that Manage service is complicated defined, even more complicate executed, so its necessary to have a Maturity level measurement in activities or business Processes scale that will be outsourced. This study is conducted to examine business Processes Maturity level that build Manage services, towards company business model rearrangement. Researchers using the eTOM framework and IT-IL in measuring business unit "Customer Relationship Management - Customer Care" which will apply Manage service. Study results identification aspects that related to potential Manage services outsourced, governance restructuring advice according to reference business Processes Maturity level, and provide appropriate policy recommendations

Keywords: *Business Process Maturity Level, eTOM, IT-IL, CRM – Customer Care*

Abstrak

Saat ini perusahaan Telekomunikasi melakukan transformasi model bisnis melalui Manage service, untuk mengurangi beban biaya perusahaan. Perusahaan telekomunikasi sangat melekat dengan teknologi mutakhir dan padat modal, maka transformasi perlu dilakukan dengan pendekatan khusus sehingga dapat mempertahankan daya saing, bukan sekedar outsourcing karyawan umumnya. Hal ini kemudian mengubah paradigma outsourcing dalam industri telekomunikasi global dalam sebuah istilah Manage service. Perusahaan menyadari bahwa Manage service ternyata rumit didefinisikan apalagi dieksekusi, sehingga perlu dilakukan pengukuran tingkat kematangan dalam skala aktivitas atau proses bisnis yang akan dialihdayakan. Penelitian dilakukan untuk mengukur tingkat kematangan proses bisnis dalam Manage services, menghasilkan penataan ulang model bisnis perusahaan. Peneliti menggunakan framework eTOM dan IT-IL dalam mengukur unit bisnis "Customer Relationship Management - Customer Care" yang akan menerapkan Manage service. Hasil studi adalah berbagai aspek identifikasi terkait dengan layanan potensial outsourcing yang dikelola, saran restrukturisasi tata kelola sesuai dengan referensi tingkat kematangan, dan memberikan rekomendasi kebijakan yang tepat.

Kata kunci: *Tingkat Kematangan Proses Bisnis, eTOM, IT-IL, CRM – Customer Care*

1. Pendahuluan

Perusahaan melakukan transformasi model bisnis mereka dengan beralih menggunakan *Manage service*. Karena penguasaan pasarnya yang tinggi di dalam dan maupun luar negeri, bagi para *vendor Manage service* dan penyedia jasa produk telekomunikasi (operator) percaya bahwa pasar *Manage service* adalah menjanjikan bagi masa depan industri yang sangat kompetitif. Bila dilihat

kembali pada bentuk bisnis perusahaan yang telah berjalan dan dalam posisi yang sudah stabil, maka proses untuk dapat merubah/merombak kembali dengan menempatkan system *Manage service* di dalamnya adalah rumit untuk didefinisikan bahkan lebih jauh lagi untuk dijalankan.

Peneliti melakukan analisis dan sintesis symptom berawal dari adanya fenomena-fenomena tersebut diatas, dengan mendefinisikan proses bisnis dan analisis aktivitas perusahaan yang lengkap, efektif dan efisien. Menentukan pola *Manage service* yang tepat melalui kerangka

Received: 23 November 2014; Revised: 22 Desember 2014;
Accepted: 27 Januari 2015 ; Published online: 30 April 2015
©2014 INKOM 2014/14-NO399

standar proses bisnis perusahaan telekomunikasi melalui TMF (*Telecommunication Modelling Framework*) menggunakan eTOM dan IT-IL, setelah itu perlu mengelola *vendor* sebagai pihak ketiga sebagai perencana dan penyelenggara layanan tersebut, sehingga pengembangan layanan dapat berjalan efektif dan efisien. Pada akhirnya, yang paling utama adalah bagaimana merekomendasikan strategi untuk menjaga atau mencapai keunggulan kompetitif perusahaan.

Berdasarkan konsep symptom, analisis dan akar permasalahan yang terkait, maka penulis merumuskan permasalahan yang dihadapi dalam penelitian yang dilakukan adalah berikut: (1) identifikasi faktor/aspek yang potensial untuk dialihdayakan terkait dengan *Manage service*; (2) restrukturisasi/rancang tata kelola *Manage service*; (3) analisis pengaruh *framework* eTOM dan IT-IL terhadap desain *Manage service*; (4) mengukur tingkat kematangan proses bisnis.

Tujuan penyusunan untuk membantu unsur organisasi perusahaan yang bergerak pada bidang Jasa Telekomunikasi. Dari sisi bisnis, perusahaan tersebut diharapkan dapat bertahan dan makin maju saat tingginya tingkat persaingan antar operator telekomunikasi, dimana sangat penting untuk mengalihdayakan proses bisnis mereka dengan tepat dan menentukan tata kelola proses bisnis *Manage service* yang paling tepat. Dari sisi produk dan teknologi, dalam industri telekomunikasi para perusahaan saling berpacu dalam menggunakan teknologi mutakhir, namun tetap mampu menjaga kualitas dan aliran keuangan yang baik sebagai badan usaha yang profit.

Penelitian diharapkan mampu mengidentifikasi faktor potensial untuk dialihdayakan dengan *Manage service*, mengacu pada kerangka yang standar pada industri telekomunikasi yaitu eTOM dan IT-IL. Perusahaan tersebut melakukan restrukturisasi pada unit bisnis CRM-*Customer Care* merujuk pada proses bisnis eTOM. Proses bisnis perusahaan yang perlu direstrukturisasi akan diteliti dengan melakukan pengambilan data langsung pada staf terkait, untuk dapat dilihat pengaruh *framework* yang menjadi acuan terhadap design *Manage service* yang akan diterapkan. Data yang didapatkan sangat berguna untuk mengukur tingkat kematangan proses bisnis saat ini pada unit bisnis terkait. Berdasarkan acuan kerangka eTOM yang digunakan oleh peneliti maka dapat dipetakan antara kematangan proses saat ini dengan proses bisnis yang diinginkan sesuai standar eTOM. Akhir dari penelitian ini didapatkan nilai kesenjangan, yang kemudian akan dapat dianalisa dan berguna sebagai saran atau rekomendasi bagi proses bisnis

yang telah berjalan, sehingga menjadi lebih baik atau optimal.

2. Landasan Teori

2.1 Perusahaan Telekomunikasi Terhadap Kerangka eTOM dan IT-IL

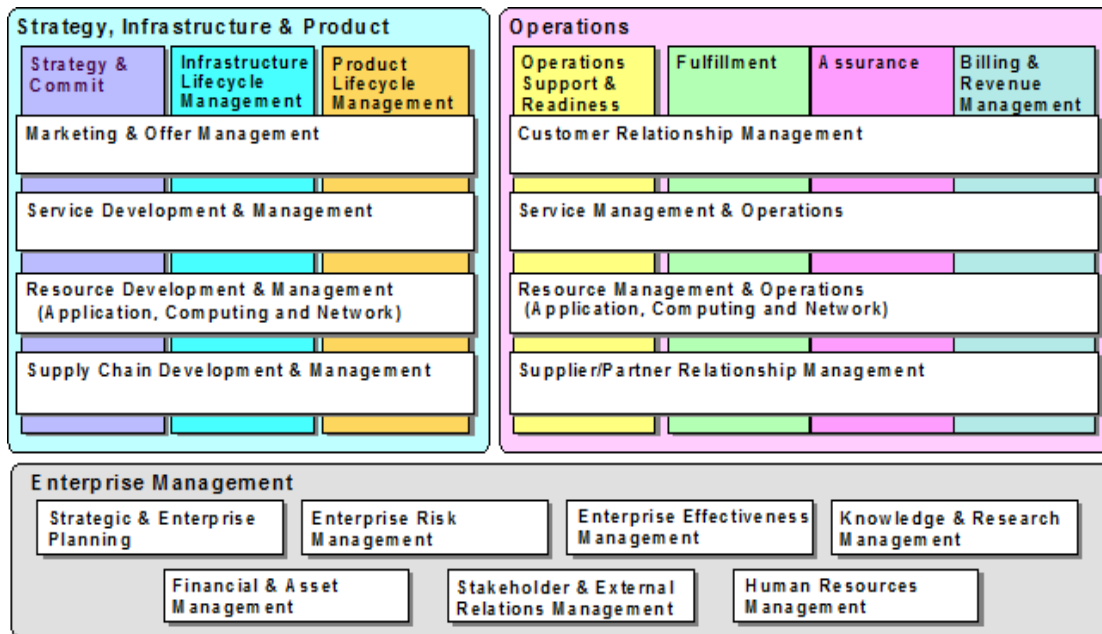
eTOM merupakan *framework* yang diterima secara luas untuk proses bisnis pada industri informasi, komunikasi dan hiburan [12], merupakan *framework* standar yang didefinisikan dalam area proses bisnis telekomunikasi [13], dan saat ini para penyedia layanan ICT, terutama industri Telecom sangat giat menggunakan eTOM untuk mengembangkan, *delivery* dan *Support* servis end-to-end pada perusahaan mereka [4], [14].

Framework eTOM dan IT-IL saling melengkapi, eTOM menangkap Kebutuhan Proses Bisnis SP dan membentuk model proses sesuai parameter kerangka yang telah dimiliki oleh eTOM. Model tsb menghasilkan berbagai aliran proses sesuai model eTOM, dan merepresentasikan solusi yang dapat berjalan untuk skenario bisnis tertentu. Sedangkan IT-IL mendefinisikan IT *good practise* yang diperlukan, kemudian mengelolanya menjadi suatu model pada area yang sebelumnya telah didefinisikan melalui eTOM, misalkan: IT *service delivery*, *Support*, dan lain-lain.

2.2 eTOM v12

eTOM adalah suatu kerangka proses bisnis yang merupakan singkatan dari Enhance Telecom Operations Map yang banyak diterapkan dan digunakan untuk menyediakan layanan end-to-end pada industri telekomunikasi[4]. eTOM sebagai standar untuk *Telecommunications Management Network*, berdasarkan ITU-T *Recommendation M.3050.2*. eTOM dikembangkan mulai awal 90an berdasarkan *input* dari berbagai *Service Provider* untuk menghasilkan model proses bisnis. eTOM adalah standar proses bisnis yang paling banyak digunakan pada industri telekomunikasi[14].

Kerangka/*framework* Enhanced Telecom Operation Map eTOM, dikeluarkan oleh TeleManagement Forum (TM Forum) dan itSMF. Konsep yang perlu selalu diingat bahwa eTOM tidak menunjukkan isu strategis atau jawaban dari siapa seharusnya *Customer* yang layak menjadi target dari *service provider*, market segment seperti apa yang seharusnya dilayani *service provider*, apa saja visi, misi *service provider* tersebut dll, akan tetapi berfungsi sebagai kerangka proses bisnis dan referensi, untuk memandu pengembangan dan manajemen perusahaan telekomunikasi. Terdapat tiga proses utama pada eTOM: "Strategy,



Gambar 1. eTOM level-0 [11].

Infrastructure & Product", "Operation", dan "Enterprise Management"

Secara pendefinisian proses bisnis, eTOM terdiri dari empat level Level-0, Level-1, Level-2 dan Level-3 pada kerangka eTOM, dimana tiap tingkat hirarki membungkus kelompok proses lebih detail di tingkat berikutnya. Pada baris/tingkat atas adalah kegiatan pelanggan (seperti pemasaran yang dihadapi), dan kemudian baris/tingkat bawah adalah *supplier* dan aktivitas pendukung. Gambar 1 adalah eTOM level 0, yaitu keseluruhan *framework* eTOM pada hierarki proses bisnis teratas atau paling luas.

2.3 IT-IL (Information Technology Infrastructure Library) v3

Information Technology Infrastructure Library IT-IL, dikeluarkan oleh UK Office of Government Commerce(OGC), yang terdiri lima volume layanan: Strategy, Design, Transition, Operation, Continual Improvement, dimana IT-IL mampu mengelola semua kegiatan dan antarmuka Manage service, dengan menyediakan tata kelola layanan.

Dalam organisasi IT, IT-IL berfungsi Referensi manajemen bisnis layanan IT, yang mampu memfokuskan pada penyelarasan layanan(service) dan pendukung(Support) layanan tersebut. IT-IL dapat Menjadikan layanan IT lebih efisien untuk berkolaborasi di seluruh lifecycle organisasi sesuai Gambar 2. Dengan IT-IL Kebutuhan pelanggan dapat lebih efektif untuk diperoleh/dikenali dan

dikelola, IT-IL dapat Menjadikan bisnis layanan pendukung IT berjalan selaras. Melalui IT-IL para pengambil kebijakan dapat proaktif mengambil kendali perubahan organisasi.

2.4 Kerjasama Framework eTOM dan IT-IL

Proses eTOM dan ITIL dapat saling terkait, dan mampu memberikan informasi mengenai pemetaan dari satu tampilan ke tampilan proses yang lain, melalui TMF yaitu sebuah Panduan Interpreter untuk praktisi eTOM dan ITIL [4].

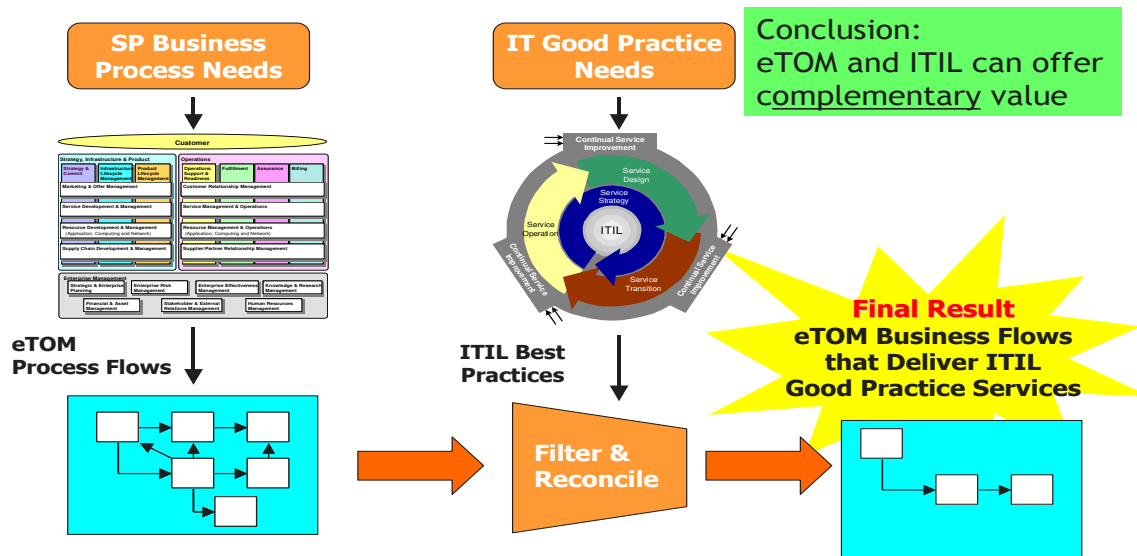
Model pada Gambar 3 menghasilkan berbagai aliran proses sesuai model eTOM, merepresentasikan solusi yang dapat berjalan pada skenario bisnis. Di sebelah kiri: eTOM menangkap Kebutuhan Proses Bisnis SP dan membentuk model proses. Di tengah: IT-IL mendefinisikan IT *good practise* tingkat kematangan yang diperlukan, mengelolanya menjadi suatu model pada area: IT Service Delivery, Support, dalam tesis ini yaitu CRM-CC. Di sebelah kanan: eTOM dan IT-IL yang saling melengkapi; menjadi aliran bisnis yang menampilkan layanan *good practice*.

2.5 IT-IL Maturity Level

Untuk mendesain proses bisnis Manage service, perlu mereferensi pada IT-IL yang spesifik pada volume design, yaitu bagaimana mengetahui tingkat kematangan proses bisnis pada saat ini untuk digantikan oleh Manage service, melalui konsep IT-IL Maturity Level.



Gambar 2. IT-IL versi 3 [7].



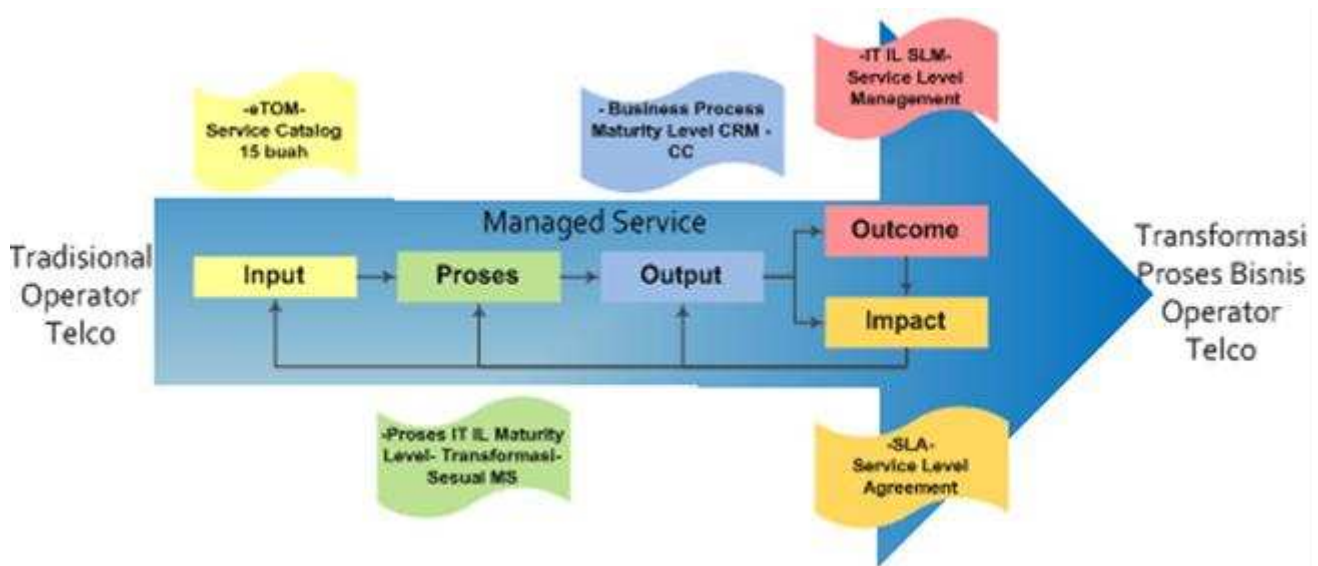
Gambar 3. ITIL and eTOM – working together (TR143_Building_Bridges-ITIL_and_eTOM – [11].

3. Metodologi Penelitian

3.1 Kerangka Berfikir Penelitian

eTOM menyediakan kerangka proses bisnis yang mendukung penelitian ini, yang sudah mendukung secara lengkap proses bisnis pada unit bisnis *Customer Relationship Management (CRM) Customer Care*, sehingga eTOM menjadi nilai

input/masukan. Pada kerangka eTOM dipilih proses bisnis hingga *level 4* secara detail, hal ini sebagai *input* awal penelitian, adapun eTOM sudah member referensi pada kerangka IT-IL *service level Management level 2* sebagai dasar dari manajemen efektif perusahaan. Proses digunakan IT-IL *good practice*, dengan spesifikasi IT-IL *service design*. Mengukur tingkat kematangan proses bisnis yang diperoleh melalui



Gambar 4. Kerangka Berfikir Umum Penelitian.

proses sebelumnya (eTOM), kemudian menggunakan teknik IT-IL *Maturity level* menjadikan proses-proses tersebut menjadi terukur tingkat kematangannya pada unit prosedural dan teknis operasional. Output yang diharapkan ialah nilai kematangan tiap proses bisnis eTOM yang relevan pada unit CRM- *Customer Care*. Penelitian dilakukan dengan menyebarkan kuesioner dan borang untuk diperoleh kondisi teknis pada *level* operasional staff sehari-hari, berdasarkan hasil dari masing-masing populasi maka akan didapatkan nilai tingkat kematangan yang objektif dari proses bisnis yang telah berlangsung pada saat ini. Nilai kematangan yang didapatkan pada saat ini dianggap sudah merepresentasikan kondisi kematangan yang sesungguhnya, untuk dianalisa pada tahap selanjutnya baik-buruk, maju-mundur, dan sudah maju-mundur nya tingkat pemahaman, praktik, dan kepuasan pelanggan secara langsung pada *level* teknis. Outcome penelitian adalah bagaimana mendefinisikan *Manage service* melalui *service level Management*, melalui pemanfaatan hasil perhitungan *level* kematangan dan kerangka eTOM yang menjadi acuan terlengkap pada industry telekomunikasi saat ini. *Service level Management* menghendaki adanya kondisi kesepakatan yang tepat, berdaya guna dan menguntungkan dari sisi operasional saat dijalankan pada unit bisnis CRM- *Customer Care*. Setelah mendefinisikan *service level Management*, kemudian eTOM dan IT-IL yang telah menghasilkan tingkat kematangan proses bisnis saat ini, melalui pemetaan pada

proses bisnis *Customer Relationship Management* sesuai dan diproses melalui *service design bestpractice* IT-IL, diperoleh kesepakatan SLA (*Service License Agreement*) dan KPI (*Key Performance Indicator*) pada tiap proses bisnis yang dijalankan. SLA dan KPI dan berbagai data sumber referensi di antaranya mengetahui kondisi proses bisnis saat ini melalui perhitungan *level* kematangan eTOM saat ini, dan juga *service level Management* sebagai definisi dari proses baru yang berisi manajemen proses bisnis baru *Manage service* yang bisa diterapkan pada perusahaan operator. Dilakukan monitoring dan evaluasi regular terhadap pencapaian KPI yang dikerjakan oleh *vendor*, hal ini berpengaruh atau impact pada kinerja dan secara langsung keuntungan perusahaan operator karena merupakan layanan yang dirasakan langsung oleh pelanggan pada sisi akhir.

3.2 Tahapan Penelitian

Fase 0: Studi Literatur. Aktivitas: Melakukan kajian pustaka pada teori eTOM dan ITIL sebagai dasar *framework* pendukung pada penelitian ini.

Fase 1: Identifikasi komponen *Framework*. Aktivitas: Melakukan dekomposisi Bisnis Proses (BP) *Customer Relationship Management* pada *Framework* eTOM, dan menyesuaikan pada *service catalog* MS.

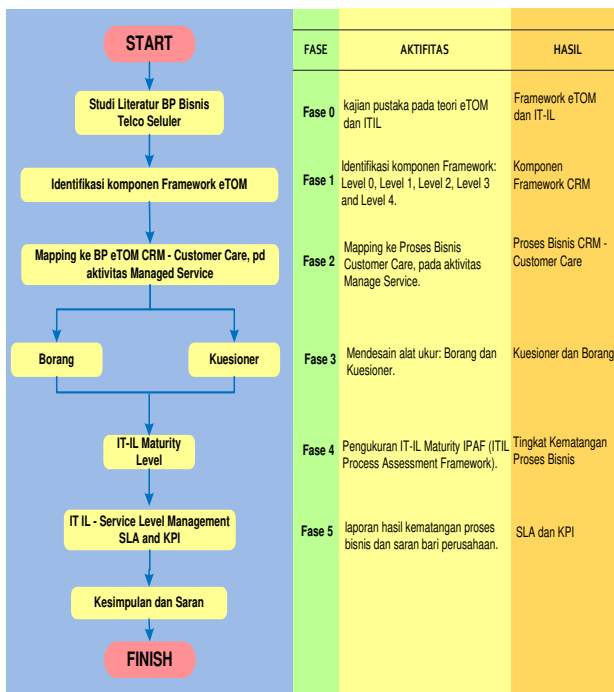
Fase 2: Mapping ke BP *Customer Care*, pd aktivitas *Manage service*. Aktivitas: Penurunan Bisnis Proses eTOM *Customer Relationship*

Management (CRM), pada bagian *Customer Care* untuk *Manage Service*.

Fase 3: Borang dan Kuesioner. Aktivitas: Perancangan Borang dan Kuesioner sesuai konsep akademis sebagai alat ukur, dan melakukan distribusi untuk mendapatkan data penelitian.

Fase 4: Pengukuran IT-IL *Maturity*. Aktivitas: Menggunakan teknik *Self Assessment*. Ini memberikan indikasi yang berguna dari pandangan praktisi IT pada seberapa baik proses telah dikerahkan dan dieksekusi, melalui kuesioner.

Fase 5: IT IL - *Service Level Management* SLA and KPI. Aktivitas: Menyediakan laporan hasil kematangan proses bisnis dan saran yang diberikan terhadap perusahaan. Saran yang diberikan berupa proses bisnis yang kurang optimal atau saran akan proses bisnis baru.



Gambar 5. Tahapan Penelitian

4. Pelaksanaan Penelitian dan Pengolahan Data

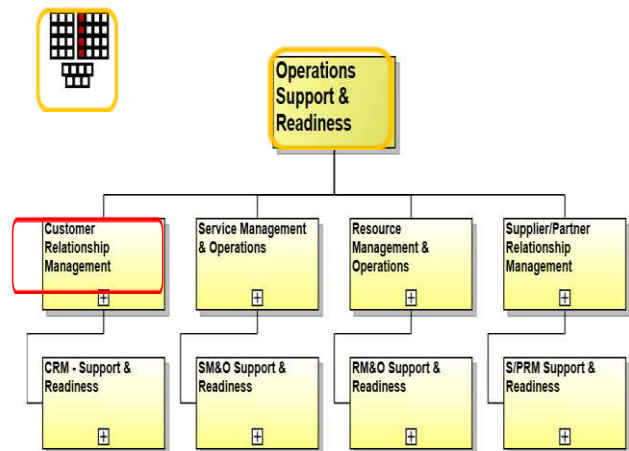
4.1 Fase 1 Penelitian

Kotak kuning adalah *framework* eTOM menyeluruh, penelitian ini mendekomposisikan CRM berdasarkan pada kotak merah, dimana Penurunan *Level 0* - Komponen *Framework* CRM. Urutan tingkat kerangka eTOM bernama: 1. eTOM Operations Horizontal Functional *Process* Grouping.



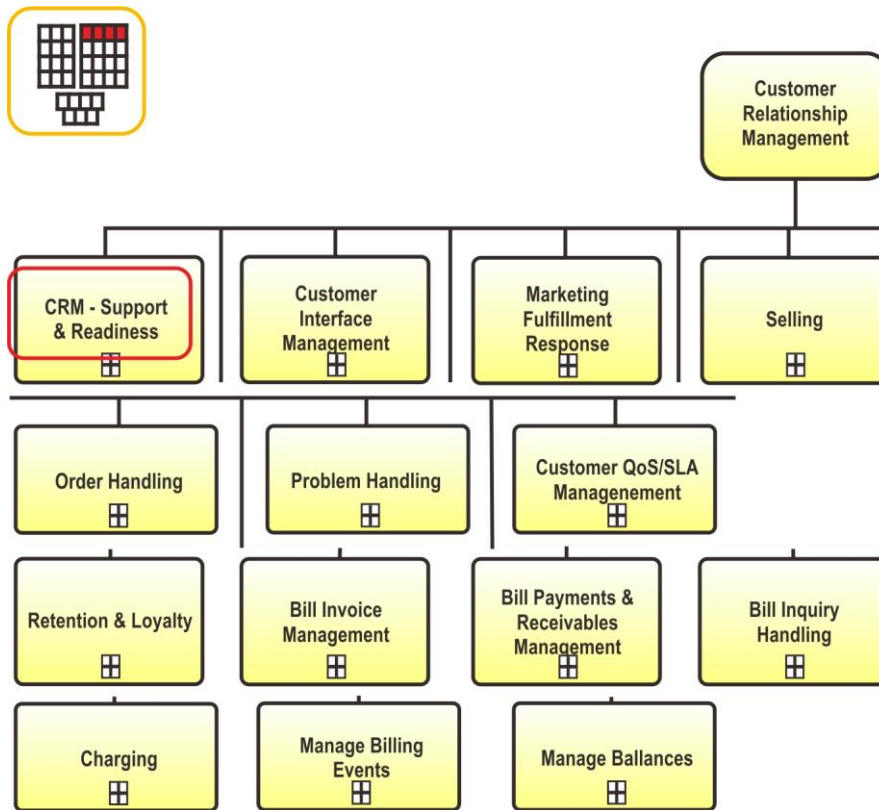
Gambar 6. Penurunan *Level 0*, identifikasi komponen *level framework* [11]

Menuju proses bisnis yang lebih detail dari operation CRM, maka dilakukan dekomposisi pada kotak merah pada gambar di bawah, dimana Penurunan bernama: *Level 1* - Komponen *Framework* CRM.

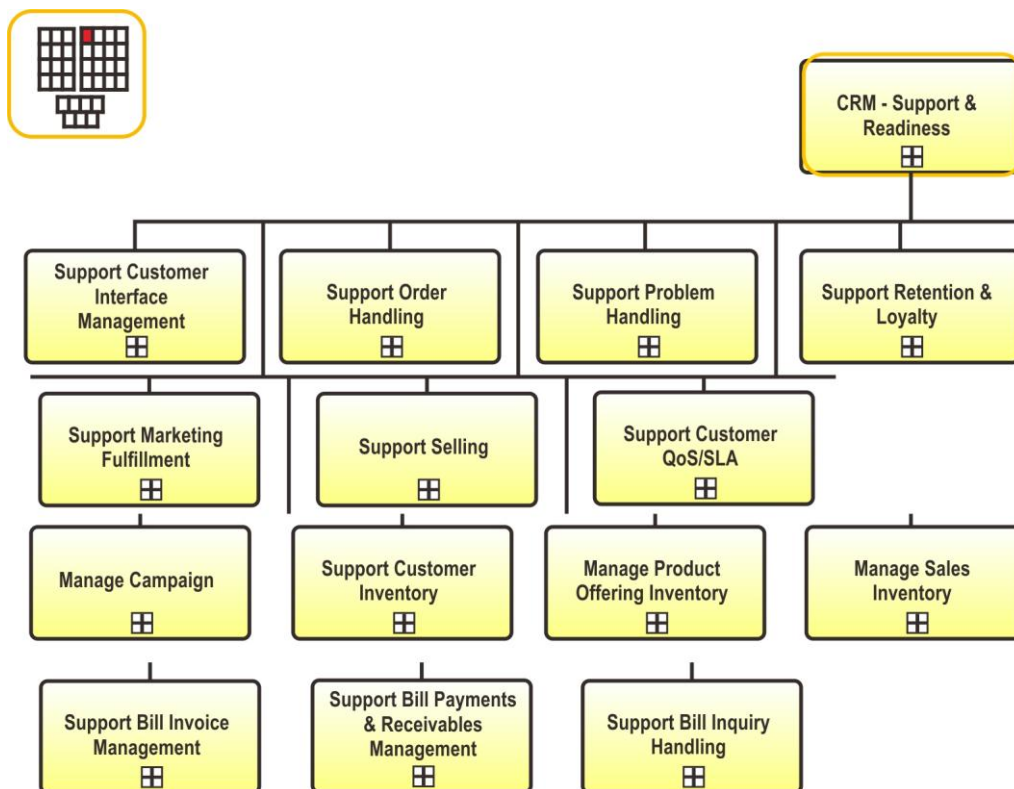


Gambar 7. Penurunan *Level 1*, identifikasi komponen *level framework* [11]

Penurunan *Level 2* - Komponen *Framework* operations Support & Readiness, memiliki 8 subproses, maka dekomposisi pada urutan tingkat kerangka eTOM bernama: 1.1.1 Customer Relationship Management.



Gambar 8. Penurunan *Level 2*, identifikasi komponen *level framework* [11]



Gambar 9. Penurunan *Level 3*, identifikasi komponen *level framework* [11].

Penurunan Level 3 - Komponen Framework Customer Relationship Management, memiliki 14 subproses, maka dekomposisi pada urutan tingkat kerangka eTOM ini bernama: 1.1.1.1 CRM - Support & Readiness.

Penurunan Level 4 - Komponen Framework pada proses bisnis CRM-Support & Readiness level 3 memiliki 15 subproses: 1.1.1.1.1 Support Customer Interface Mgt, 1.1.1.1.2 Support Order Handling, 1.1.1.1.3 Support Problem Handling, 1.1.1.1.4 Support Billing & Collections, 1.1.1.1.5 Support Retention & Loyalty, 1.1.1.1.6 Support Marketing Fulfillment, 1.1.1.1.7 Support Selling, 1.1.1.1.8 Support Customer QoS/SLA, 1.1.1.1.9 Manage Campaign, 1.1.1.1.10 Manage Customer Inventory, 1.1.1.1.11 Manage Product Offering Inventory, 1.1.1.1.12 Manage Sales Inventory, 1.1.1.1.13 Support Bill Invoice Mgt, 1.1.1.1.14 Support Bill Payments&Receivables Mgt, 1.1.1.1.15 Support Bill Inquiry Handling

4.2 Fase 2 Penelitian

Mapping ke Proses Bisnis Customer Care, pada aktivitas Manage service. Penurunan Proses Bisnis eTOM Customer Relationship Management (CRM), sesuai pada criteria unit bisnis Customer Care Manage service.

Tabel 1. Penentuan Proses Bisnis, kriteria Customer Care

Level 0	Level 1	Level 2	Level 3	Level 4	Customer Care
1. eTOM Operations Horizontal Functional Process Grouping	1.1 Operations Support & Readiness	1.1.1 Customer Relationship Management	1.1.1.1 CRM - Support & Readiness	1.1.1.1.1 Support Customer Interface Mgt	√
				1.1.1.1.2 Support Order Handling	√
				1.1.1.1.3 Support Problem Handling	√
				1.1.1.1.5 Support Retention & Loyalty	
				1.1.1.1.6 Support Marketing Fulfillment	
				1.1.1.1.7 Support Selling	
				1.1.1.1.8 Support Customer QoS/SLA	√
				1.1.1.1.9 Manage Campaign	
				1.1.1.1.10 Manage Customer Inventory	√
				1.1.1.1.11 Manage Product Offering Inventory	
				1.1.1.1.12 Manage Sales Inventory	
				1.1.1.1.13 Support Bill Invoice Management	
				1.1.1.1.14 Support Bill Payments & Receivables Management	
				1.1.1.1.15 Support Bill Inquiry Handling	

Dekomposisi pada tahap ini menghasilkan 5 proses bisnis sebagai berikut: L4 : 1.1.1.1.1 Support Customer Interface Mgt, L4 : 1.1.1.1.2 Support Order Handling, L4 : 1.1.1.1.3 Support Problem Handling, L4 : 1.1.1.1.8 Support Customer QoS/SLA, L4 : 1.1.1.1.10 Manage Customer Inventory

4.3 Fase 3 Penelitian

Mendesain alat ukur: Borang dan Kuesioner. Teknik purposive random sampling yaitu

pengambilan sampel untuk tujuan dan dengan syarat-syarat tertentu [15].

Alat ukur Kuesioner: *Maturity Level* pada kuesioner berdasarkan kriteria *level IT SM Benchmark*, agar sesuai tujuan yang ingin dicapai oleh konteks proses bisnis eTOM, yaitu *Framework Process Report 12.5.0*. Alat Ukur borang: Mengkonstruksikan mengenai organisasi, kejadian dan tuntutan pada proses bisnis.

4.4 Fase 4 Penelitian

Pengukuran IT-IL bersumber dari *Maturity IPAF (ITIL Process Assessment Framework)*.

Perspektif Kematangan Proses: Teknik *Self Assessment*, Memberikan indikasi dari pandangan praktisi IT, seberapa baik proses telah dikerahkan dan dieksekusi

Deskripsi *Maturity Level* pada IT-IL [8] adalah sebagai berikut:

0: *Non-existent, Nothing present*

1: *Initial, No standard Processes, Ad hoc approaches.*

2: *Repeatable, No formal training or communication.*

3: *Defined, Standarized and documented.*

4: *Manage, Monitor and measure compliance to procedures.*

5: *Optimized, Proactive and continues self improvement.*

4.5 Fase 5 Penelitian

IT IL - *Service Level Management SLA and KPI*. Laporan kematangan proses bisnis dan saran yang diberikan kepada perusahaan, berupa proses bisnis yang kurang optimal atau proses bisnis baru yang dianggap perlu muncul sesuai standar eTOM.

4.6 Pengujian Sistem

Uji tabulasi hasil pengambilan data kuesioner terdiri dari uji validitas dan uji realibilitas. Uji validitas untuk mengetahui kelayakan butir-butir pertanyaan yang diajukan dalam kuesioner, sedangkan uji realibilitas melihat konsistensi/kehandalan jawaban kuesioner.

4.6.1 Uji Validitas

Menggunakan korelasi *Bevariate Pearson* untuk uji validitas data, dengan mengkorelasikan masing-masing skor item pertanyaan dengan skor total. Berikut adalah hipotesis dalam pengujian validitas:

Ho: Jika nilai $r_{hitung} > r_{tabel}$, maka item pertanyaan atau pernyataan dalam angket berkorelasi signifikan terhadap skor total (artinya item kuesioner dinyatakan valid).

Ha: Jika nilai $r_{hitung} < r_{tabel}$, maka item pertanyaan atau pernyataan dalam angket tidak berkorelasi signifikan terhadap skor total (artinya item kuesioner dinyatakan tidak valid).

4.6.2 Uji Realibilitas

Menggunakan rumus cronbach alpha untuk uji reliabilitas. Berikut hipotesis pengujian realibilitas:

Ho: Jika nilai Cronbach Alpha $> 0,60$, maka konstruk atau variabel dikatakan reliabel (artinya kuesioner dinyatakan realibel).

Ha: Jika nilai Cronbach Alpha $< 0,60$, maka konstruk atau variabel dikatakan tidak reliabel (artinya kuesioner dinyatakan tidak realibel).

4.7 Hasil Uji Validitas dan Realibilitas

Pada pengolahan data pertama, didapatkan proses bisnis SCIM dan SPH memenuhi syarat validitas dan realibilitas, sedangkan tiga proses lainnya SOH, SCQ, MCI memerlukan percobaan lanjutan.

Tabel 2. Hasil Uji Validitas dan Realibilitas

Proses Bisnis	Pengolahan Data Pertama		
	Validitas pearson	Reliabilitas Cronbach	Butir Valid
<i>Support Customer Interface Mgt (SCIM)</i>	Ya	Ya	15
<i>Support Order Handling (SOH)</i>	Ya	Ya	11
<i>Support Problem Handling (SPH)</i>	Ya	Ya	14
<i>Support Customer QoS/SLA (SCQ)</i>	Ya	Ya	14
<i>Manage Customer Inventory (MCI)</i>	Ya	Ya	9

4.8 Tingkat Kematangan Proses Bisnis

Tingkat Kematangan atribut dihitung dengan menggunakan rumus rata-rata, yaitu total bobot jawaban kuesioner dibagi dengan jumlah pertanyaan.

Distribusi skor proses bisnis 1: *Support Customer Interface Management – SCIM*

Pertanyaan	Q1	Q3	Q5	Q6	Q7	Q8	Q9	Q10	Q12	Q13	Q14	Q15	Q16	Q18	Q19
Rata-rata	3.6	3.44	3.56	4.08	4.04	3.76	4.16	3.76	3.92	3.6	2.84	4.12	3.36	3.16	3

Didapatkan rata-rata *level* kematangan di tingkat 3,63

Distribusi skor proses bisnis 2: *Support Order Handling – SOH*

Pertanyaan	Q20	Q21	Q22	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q32
Rata-rata	3.28	3.48	4.08	4.32	3.32	3.96	3.96	4.08	3	4.28	2.96

Didapatkan rata-rata *level* kematangan di tingkat 3.7

Distribusi skor proses bisnis 3: *Support Problem Handling – SPH*

Pertanyaan	Q34	Q35	Q36	Q38	Q39	Q40	Q42	Q43	Q44	Q47	Q49	Q50	Q51	Q53
Rata-rata	3.6	4.56	3.6	4.28	3.88	3.88	2.84	3.52	2.8	2.4	4.52	3.84	3.92	3.4

Didapatkan rata-rata *level* kematangan di tingkat 3.65

Distribusi skor proses bisnis 4: *Support Customer QoS/SLA – SCQ*

Pertanyaan	Q54	Q56	Q57	Q58	Q59	Q60	Q63	Q64	Q65	Q70	Q71	Q72	Q75	Q79
Rata-rata	3.48	2.84	3.88	3.08	4.16	3.88	3	2.64	2.84	3.8	3.04	2.6	2.72	2.56

Didapatkan rata-rata *level* kematangan di tingkat 3,18

Distribusi skor proses bisnis 5: *Manage Customer Inventory – MCI*

Pertanyaan	Q82	Q83	Q84	Q86	Q87	Q89	Q91	Q94	Q95
Rata-rata	2.88	2.32	3.4	2.92	3.04	3.92	4.04	3.32	2.72

Didapatkan rata-rata *level* kematangan di tingkat 3.17

5. Analisis dan Hasil Penelitian

Dari hasil pengolahan data didapatkan rata-rata skor pertanyaan valid, sebagai tingkat kematangan proses bisnis saat ini (as-is). Sedangkan tingkat kematangan yang akan dicapai adalah nilai standar “*Framework eTOM Assurance CRM*” proses bisnis terkait.

Tabel 3 : Tingkat kematangan Proses Bisnis

Proses Bisnis	Tingkat Kematangan	
	as-is	to-be
<i>Support Customer Interface Management - SCIM</i>	3.63	4
<i>Support Order Handling - SOH</i>	3.70	4
<i>Support Problem Handling - SPH</i>	3.65	4
<i>Support Customer QoS/SLA – SCQ</i>	3.18	4
<i>Manage Customer Inventory – MCI</i>	3.17	4

5.1 Analisa Tingkat Kematangan Saat Ini

Indikasi 80% tingkat kematangan saat ini > 3 (terdefinisi), maka kebutuhan akan *Management CRM* telah dipahami dalam organisasi. Walaupun organisasi belum mencapai kematangan standar eTOM.

Indikasi Proses bisnis *Support Problem Handling* memiliki tingkat kematangan tertinggi 3,9, maka Organisasi CRM telah memahami tentang aturan, prinsip dan konsep dasar dari *Order/Pemesanan layanan* oleh pelanggan. Pelatihan IT informal telah dilakukan secara baik, melalui praktik langsung.

Indikasi Proses bisnis *Support Customer QoS/SLA* memiliki tingkat kematangan terendah dalam organisasi, maka Organisasi CRM belum memanfaatkan Aplikasi CRM, untuk mendukung tercapainya QoS / SLA. Belum adanya proses yang proaktif memantau dan menilai kinerja terhadap parameter QoS / SLA. Adanya kekurangan dalam manajemen dan pelaporan, terhadap kemampuan proses mencapai QoS / SLA. Menurut indikasi hasil borang pada proses bisnis *Support Customer QoS/SLA* yang memiliki *level* kematangan terendah

20	Bagaimana model loyalitas, dan model pengukuran kelas / kategori pelanggan yang digunakan? Jawaban: pengukuran loyalitas di ukur dari penggunaan/usage ber bulan yang di lakukan oleh customer
21	Seberapa sering perusahaan menjalankan analisis derajat loyalitas pelanggan? Apakah mereka secara teratur melakukannya atau berdasarkan pada basis per-permintaan? Jawaban: 3 bulan sekali di lakukan review loyalitas pelanggan
26	Apakah ada model awal yang perlu dibangun dari perilaku pelanggan dan analisis nilai wawasannya, di dalam laporan layanan pelanggan? Jawaban: tidak ada

Gambar 11. Hasil Borang Penelitian

Berdasarkan borang ditemukan kekurangan sebagai berikut: Review loyalitas dilakukan dalam waktu yang terlalu lama, Parameter loyalitas yang diukur sangat terbatas hanya pada usage/penggunaan, Prosedur *Management* data belum didokumentasikan secara menyeluruh karena tidak ada model.

5.2 Analisa Tingkat Kematangan yang Diharapkan

Tingkat kematangan proses bisnis yang ingin dicapai adalah mengacu pada proses bisnis CRM kerangka eTOM pada subproses terkait. Keempat proses bisnis tersebut, pada kerangka eTOM memiliki tingkat kematangan bernilai 4 (*Managed*).Kematangan tingkat 4 IT-IL: Terukur dan terkontrol, yaitu terdapat kumpulan data dan proses rinci dengan kualitas baik.

Analisa TO-BE (*Operations - Assurance*) menurut eTOM sesuai kematangan tingkat 4 eTOM CRM adalah proses bisnis telah memiliki kriteria berikut: *Create Customer Problem Report, Correct & Recover Customer Problem, Track & Manage Customer Problem, Report Customer Problem, Close Customer Problem Report, Isolate Customer Problem*.



Gambar 12. Analisis TO-BE eTOM (Operations-Assurance)

5.3 Analisa Kesenjangan (Gap)

Menganalisa kesenjangan antara aktual proses bisnis saat ini terhadap eTOM, maka melalui kesenjangan atribut *Support Customer Interface Management, Support Order Handling* dan *Manage Customer Inventory* adalah relatif sama sebesar 0.5 s/d 0.6, dan *Support Customer QoS/SLA* memiliki kesenjangan terbesar yaitu 1 tingkat. Adapun *Support Problem Handling* dianggap tidak memiliki kesenjangan karena hanya memiliki selisih kecil 0.1.

Berdasarkan standar proses bisnis eTOM yang menjadi acuan, maka semua atribut diharapkan mencapai kematangan *level* 4. Perlu adanya perbaikan berupa proses pencapaian tingkat kematangan *level* 4 pada *Support Customer QoS/SLA* berdasarkan *Framework eTOM Assurance CRM*.

5.4 Pencapaian Tingkat Kematangan 4

Terdapat enam kriteria mencapai *level* kematangan eTOM "*Support Customer QoS/SLA - SCQ*", yang perlu diterapkan pada organisasi CRM operator, sesuai *Framework eTOM Assurance CRM*.

1. *Access Customer QoS/SLA Performance.* Mengelola keseluruhan penilaian kinerja QoS/SLA pelanggan dan Melakukan pemantauan proaktif
2. *Create Customer QOS Performance Degradation Report.* Membuat laporan yang menilai kinerja QOS yang menurun
3. *Manage QOS/SLA violation.* Memastikan pelanggan dan proses internal terkait mendapatkan informasi yang cukup
4. *Track & Manage Customer QOS performance resolution.* Secara efisien Menunjuk, mengkordinasikan dan menelusuri pelanggan yang membeli produk.
5. *Report Customer QoS Performance.* Analisis agregat kinerja QoS seluruh kelompok pelanggan atau analisis jenis SLA, melacak dan memantau proses QoS/SLA Pelanggan dan biaya yang terkait, dan melaporkannya sebagai kemampuan pelanggan dalam memproses QoS/SLA

6. *Close Customer QoS performance Degradation Report*. Menutup laporan penurunan kinerja QOS, tepat saat masalah telah diselesaikan

6. Kesimpulan

Penelitian mengidentifikasi faktor atau aspek potensial pada perusahaan operator yang akan dialihdayakan melalui *Manage service*, yaitu dalam unit bisnis *Customer Relationship Management*. eTOM digunakan untuk mendefinisikan proses bisnis beserta kriteria pendukungnya, dalam organisasi CRM - *Customer Care*. Aspek faktor atau potensial perusahaan diidentifikasi melalui komponen proses bisnis *framework* eTOM, untuk kemudian dipetakan pada proses bisnis *Customer Care*, dimana didapatkan 5 dari 16 proses bisnis hasil identifikasi: *Support Customer Interface Management*, *Support Order Handling*, *Support Problem Handling*, *Support Customer QOS/SLA*, *Manage Customer Inventory*.

Restrukturisasi atau perancangan tata kelola *Manage service* pada organisasi perusahaan mampu dilakukan melalui evaluasi *Level* kematangan IT-IL, dengan mengetahui proses bisnis saat ini (as-is) dan kondisi yang diharapkan (to-be). Mengetahui kondisi yang dapat diharapkan dalam rangka penglihdayaan unit bisnis melalui *Manage service*, dengan melakukan evaluasi maka proses bisnis yang sudah baik dapat dipertahakan, sedangkan proses bisnis yang kurang atau tidak baik dapat dilakukan perbaikan berdasarkan kondisi yang diharapkan. Menerapkan *Service Level Management (SLM)* pada unit kerja *Customer Relationship Management – Customer Care*, dengan output berupa matriks tanggungjawab RACI dan dokumen KPI (Key Performance Indicator)

Pengaruh *framework* eTOM dan IT-IL terhadap desain *Manage service*, sesuai evaluasi yang telah dilakukan. Organisasi CRM pada operator yang diteliti saat ini (*as is*) berdasarkan *framework* eTOM dan IT-IL secara umum sudah baik dikarenakan memiliki tingkat kematangan yang hampir mencapai standar nilai kematangan eTOM. Yaitu pada 3 proses bisnis yaitu: *Support Customer Interface Management*, *Support Order Handling*, dan *Manage Customer Inventory*. Karena sudah cukup baik dengan *level* kematangan (3.6 - 3.7), maka *Manage service* tetap dapat menggunakan mekanisme proses bisnis yang sedang berlaku saat ini tanpa melakukan perubahan signifikan.

Tingkat kematangan proses bisnis *Support Order Handling*, memiliki *level* tertinggi yaitu 3.7, nilai tersebut hamper mendekati *level* kematangan tertinggi eTOM pada proses bisnis yang sama. Pengalihdayaan melalui *Manage service* tanpa perlu melakukan perubahan mekanisme proses bisnis yang telah berlaku saat ini.

Daftar Pustaka

- [1] Arrianto MW, *IT Governance Maturity At Indonesian State Owned Enterprises:Contingent Factors & Impacts*, Universitas Indonesia, Jakarta2011.
- [2] Francesco L, *Outsourcing Of Manage Tellecommunication Services: Reasons and Implications on Service Development and Business Model*, Politecnico di Milano, Milan, 2011.
- [3] Hou J, *A Practical Approach to the Operation of Telecommunication Services driven by the TMF eTOM Framework*, Universitat Politècnica de Catalunya, Barcelona Spain, 2009.
- [4] Jenny H, "eTOM and ITIL - An IT Outsourcing Service Provider", BPTrends Journal eTOM and ITIL,2005.
- [5] Kannamani, *Effective Implementation of Problem Management in ITIL Service Management*, *International Journal of Scientific & Engineering Research*, Houston USA, 2013.
- [6] Nontobeko, *Investigating The Use of The ITIL Framework Towards IT Service Delivery*, NMMU, South Africa, 2009.
- [7] OGC, *ITIL V3 - Service Design*, TSO for OGC, United Kingdom, 2011.
- [8] Peter AJ, *Information Technology Infrastructure Library Problem-A Maturity Model Approach*, Boise State University, Idaho United States, 2009.
- [9] Peter L, *Manage service in Telecommunication Sector Getting messages across*, Gavle University, Swedia, 2008.
- [10] Raninditha, *Aplikasi Manage services Pada Operator Telekomunikasi, (XL Axiata) dan operator jaringan (PT Huawei)*, Universitas Indonesia, Jakarta, 2012.
- [11] TM Forum, *Business Process Framework eTOM - v12*, TM Forum, USA, 2012.
- [12] Nimmy J, *Integrating the eTOM Business Process Framework into the Ericsson Seamless Modeling*, Technische Universiteit Eindhoven, Belanda, 2010
- [13] Faride L, *Enriched eTOM Framework in Service Deliver Operation through Alignment with COBIT5 Strategic Objectives*, Islamic Azad University, Iran, 2014.
- [14] Mursyid H, *Rute Menuju Best Practice*, SBM ITB, Bandung, 2008.
- [15] Sugiyono, *Metode Penelitian Administrasi*, Alfabeta, Bandung, 2001.

Perbandingan Performa Pelacakan antara Repetitive Controller dan PI Controller

On Comparison Between Repetitive Controller and PI Controller Tracking Performance

Edi Kurniawan

Research Center for Informatics, Indonesian Institute of Sciences

Komplek LIPI, Jl Cisitua No.21/154D Bandung, Indonesia

Email: edi.kurniawan@lipi.go.id

Abstract

Tracking periodic signals are common task in many control problems. One of the examples is movement control of pick and place robot in industry. The requirement of high tracking accuracy becomes very important in many applications. Therefore, a sophisticated control algorithm that manages to achieve high accuracy tracking of periodic command is required. Repetitive Control (RC) based on internal model principle is one of control schemes that can be employed to achieve perfect tracking of periodic signal. On the other hand, Proportional Integral (PI) controller can also be used for tracking. This paper compares the tracking performance of PI controller, RC, and PI with RC, where PI with RC here is integration between PI controller and RC. Step by step design to obtain the parameters of PI, RC and PI with RC are given. A simulation on servo motor system is carried out to assess the performance of RC, PI, and PI with RC respectively. From the simulation results, the transient response and tracking accuracy are thoroughly discussed.

Keywords: Repetitive Control, PI controller, tracking, periodic signal

Abstrak

Pelacakan sinyal periodik adalah kegiatan umum dalam banyak permasalahan kontrol. Salah satu contoh adalah kontrol gerakan *pick and place* robot di industri. Kebutuhan akurasi yang tinggi menjadi sangat penting dalam banyak aplikasi kontrol. Oleh karena itu, algoritma kontrol untuk mencapai pelacakan akurasi yang tinggi khususnya untuk sinyal periodik sangat diperlukan. *Repetitive Controller* (RC) berdasarkan prinsip *internal model* adalah salah satu dari banyak algoritma kontrol yang dapat digunakan untuk pelacakan sempurna sinyal periodik. Di sisi lain, *Proportional Integral* (PI) controller juga dapat digunakan untuk pelacakan. Makalah ini membandingkan kinerja pelacakan PI controller, RC, dan PI dengan RC, di mana PI dengan RC di sini merupakan integrasi antara PI controller dan RC. Langkah-langkah untuk mendapatkan parameter PI, RC, dan PI dengan RC di berikan dalam makalah ini. Simulasi untuk sistem servo motor dilakukan untuk menguji kinerja RC, PI, dan PI dengan RC. Dari hasil simulasi, *Transient response* dan akurasi pelacakan dibahas secara mendalam.

Kata kunci: Repetitive Control, PI controller, pelacakan, sinyal periodik

1. Introduction

Tracking control can be found in many industrial applications such as contour tracking in machining processes [1], trajectory tracking of robot manipulator [2], and etc. Proportional Integral (PI) controller becomes most common feedback control, and considered as first basic solution in the control of industrial system [3]–[5]. Proportional Integral is part of Proportional

Integral Derivative (PID) controller, which is used more than 90% of control loop in today use [4]. Most loops are PI because derivative action is not used very often.

In some repetitive processes, using PI controller may not be satisfactory, and a more sophisticated controller is needed to achieve control objectives. There is another tracking controller named as Repetitive Controller (RC), that can be used for tracking, especially for periodic signal. Tracking periodic signal is also common problem found in many industrial applications. As listed in [6]–[7], RC has been successfully used for robot control, accurate

position control of piezoelectric actuators, and etc. Recently, RC has been used for tracking control in underwater applications [8], tracking control of engine valve system [9][10], and tracking of contouring tasks in an industrial biaxial precision gantry [11].

RC is based on the idea of internal model principle by [12], that uses the model of reference signal in the controller. The internal model has a capability to learn from the previous cycle error, then generate control signal that can refine the tracking output to be as close as possible to the reference signal. Besides tracking periodic reference, RC can also be employed for rejecting period disturbance. However, the control objective discussed in this paper is only for tracking reference signal.

This paper investigates the tracking performance between PI controller, RC, and PI with RC. The pros and cons of each control algorithm will be reviewed.

This paper is structured as follows; Section 2 and 3 presents the overview of PI and RC respectively. Numerical example that covers the simulations results and discussion is given in Section 4. Section 5 concludes the paper.

2. Overview: PI Controller

The PID controller has the following transfer function:

$$C(s) = K_p + K_i \frac{1}{s} + K_d s \quad (1)$$

where K_p is proportional gain, K_i is integral gain, K_d is derivative gain, $\frac{1}{s}$ is integrator, and s is differentiator, and it is also a Laplace operator.

If we omit derivative part as derivative action is not used very often [4], then we have the following PI transfer function.

$$C(s) = K_p + K_i \frac{1}{s} \quad (2)$$

Proportional, Integral control is based on the present (P) and the past (I) control error, where error is obtained from the difference between the actual and desired output. This can be seen from the control law as follows:

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau \quad (3)$$

We can see that Proportional part uses the current control error, while the Integral part accumulates the previous control error values.

Tuning PI gains becomes a very crucial part in the design of PI controller, and PI tuning method has been a large research area. There are many aspects that should be taken into account when designing the PI controller. Desirable features of a design procedure are [4]:

- (a) It should give a controller that meets the design specifications.
- (b) It should be based on the available process knowledge
- (c) It should meet limitations on computation power and resources available.

There are several methods for tuning PI gains such as *Manual tuning*, *Ziegler–Nichols*, *Software tools*, *Cohen–Coon*, and etc. The most effective ways usually involve the knowledge of process/plant model, then choosing P, I, gains based on the plant model parameters. Effects of increasing PI gains independently is shown in Table 1 [13].

Table 1. Effects of increasing K_p and K_i independently

Gain	Rise time	Over-shoot	Settling time	Steady-state error
K_p	Decrease	Increase	Small Change	Decrease
K_i	Decrease	Increase	Increase	Eliminate

From the Table 1, it shows that increasing K_i eliminates the steady state error which is required for tracking accuracy. However, it comes with the trade-off such as increasing overshoot and rise time, which represents poor transient response. Therefore, tuning PI gains is an effort to obtain optimal parameters that meets the design specifications, or not to obtain best parameters that satisfies all aspects.

PI controller shown in (2) is in continuous-time form, in which it needs to be digitized in order to be implemented in digital computer. The digital/discrete-time PI is basically an approximation of the continuous-time form. Using forward difference approximation, the discrete-time PI is formulated as follows:

$$C(z) = K_p + K_i T \frac{1}{z-1} \quad (4)$$

where T is sampling time, and z is a discrete frequency domain operator.

3. Overview: Repetitive Control

Since the introduction of the digital computer, the use of digital control has greatly expanded for several reasons, such as being cheaper, smaller, and more flexible than analogue hardware. The

first digital RC was introduced in [14], where the digital RC has the following transfer function:

$$C_{rc}(z) = F(z) \frac{z^{-N}}{1 - z^{-N}} \quad (5)$$

where $N = T_r/T$ is the number of samples per reference period and it has to be integer, T_r is reference period, T is sampling time, and $F(z)$ is RC compensator.

RC involves two main designs; Internal model and compensator design. The internal model is a generator of periodic signal which has a capability to generate periodic signal so the perfect tracking of periodic reference can be achieved. The internal model is shown as the term $(z^{-N}/1 - z^{-N})$ in (5). RC compensator $F(z)$ is required to stabilize the closed-loop system. RC can be designed either in standard or plug-in manner as shown in Figure 1(a) and 1(b) respectively.

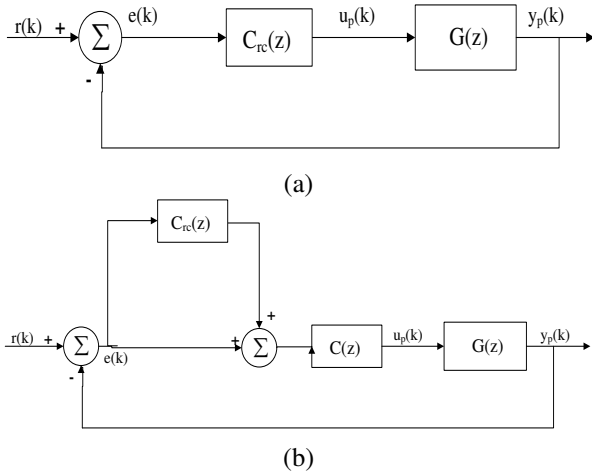


Figure 1. Block diagram of Repetitive Control System (a) Standard (b) Plug-in.

The plug-in RC has been introduced by Cosner et al [15], where RC is added in plug-in manner as shown in Figure 1(b). $C(z)$ is a nominal controller used (e.g P, PI, PID, Lead, Lag). A compensator design based on Zero Phase Tracking Error Controller (ZPTEC) proposed in [16], [17] can be used to obtain a compensator $F(z)$. For standard RC system shown in Figure 1(a), $F(z)$ is designed to compensate the dynamics of plant $G(z)$. If $G(z)$ is stable minimum phase plant, $F(z)$ can be designed as the inverse of $G(z)$

$$F(z) = \frac{1}{G(z)} \quad (6)$$

While for plug-in RC shown in Figure 2(b), $F(z)$ is designed to compensate the dynamics of closed-loop model as follows:

$$G_c(z) = \frac{C(z)G(z)}{1 + C(z)G(z)} \quad (7)$$

$$F_p(z) = \frac{1}{G_c(z)} \quad (8)$$

where $F_p(z)$ is a compensator of the plug-in RC system.

The standard RC system is stable if the following two conditions are satisfied [15][17]:

1. $G(z)$ is a stable transfer function
2. $\|1 - G(z)F(z)\|_\infty < 1$

where $\|\cdot\|_\infty$ denotes the norm infinity of the transfer function.

For plug-in RC system to be stable, the term $G(z)$ in stability conditions (1) and (2) is replaced with $G_c(z)$.

4. Numerical Example

Numerical examples are given to investigate the performance of PI, RC, and PI with RC (Plug-in RC). The following continuous plant model is used in the simulation

$$G(s) = \frac{649.3}{s^2 + 37.31s + 649.3} \quad (9)$$

which is a transfer function of stabilized servo motor used in [18].

4.1 Standard RC Design

Let the sampling time be $T = 0.005s$, and the reference signal $r(k)$ to be tracked has a period 1.25 s. This gives the number of samples N as 250. The discrete model of (9) with the chosen sampling period is given as follows:

$$G(z) = \frac{0.007623z + 0.007164}{z^2 - 1.815z + 0.8298} \quad (10)$$

which is a minimum phase stable plant as its zero and poles located inside the unit circle.

The first stability condition of RC system is satisfied due to the plant model is stable. The choice of compensator $F(z)$ as the inverse of $G(z)$ makes the second stability condition is fulfilled. Thus, the digital RC can be formulated as follows:

$$C_{rc}(z) = 131.23 \frac{z^2 - 1.815z + 0.8298}{z^{251} + 0.9402z^{250} - z - 0.9402} \quad (11)$$

4.2 PI Design

Let the design in continuous time, then PI controller will be digitized with the sampling period $T = 0.005s$. The function `pdtune` in MATLAB can

used to tuned gain of PI. The obtained gain gives the following PI Controller in continuous time.

$$C(s) = 1.100 + 18.900 \frac{1}{s} \quad (12)$$

The digital PI is given as follows:

$$C(z) = 1.100 + 18.900T \frac{1}{z-1} \quad (13)$$

, where T is sampling period.

4.3 Plug-in RC (PI with RC) Design

Let the PI controller given in (13) is the nominal controller used in the plug-in RC system as shown Figure 1(b). The closed-loop model as formulated in (7) is :

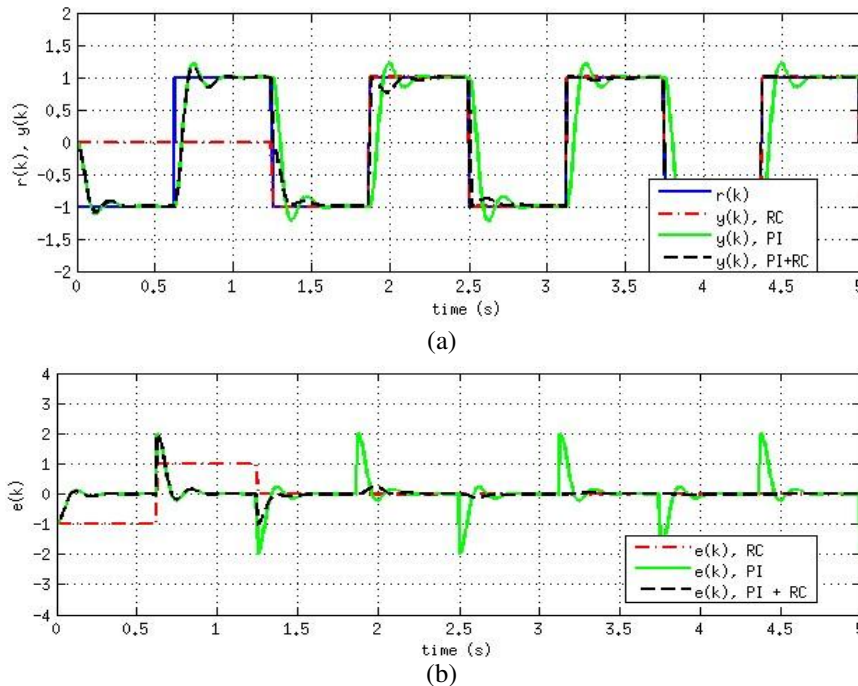


Figure 2. (a) Tracking Output of Square Signal, (b) Tracking Error.

$$G_c(z) = \frac{8.38 \times 10^{-3}(z^2 + 0.025z - 0.859)}{z^3 - 2.807z^2 + 2.645z - 0.837} \quad (14)$$

Then, the digital plug-in RC is :

$$C_{rc}(z) = \frac{119.24[z^3 - 2.807z^2]}{z^{252} + 0.026z^{251} + 2.645z - 0.837} \quad (15)$$

$$\frac{+0.859z^{250} - z^2 - 0.025z + 0.858}{z^{252} + 0.026z^{251} + 2.645z - 0.837}$$

4.4 Results and Discussion

Two tracking scenarios are carried out in the simulation. SIMULINK is used to simulate the tracking controls of RC, PI, and PI with RC. The block diagram of the SIMULINK model is given in Appendix. The first scenario is that the plant output is required to track square reference signal. The tracking outputs and errors of RC, PI, and PI with RC are shown in Figure 2(a) and (b) respectively. From Figure 2(a), tracking output of RC starts to perfectly follow the reference signal at second cycle (1.25 – 2.5 s), but the output is zero in the first cycle (0 – 1.25 s). We can also observe from

Figure 2(b) that the tracking error of RC in the first cycle is equal to the reference signal, while the tracking error in the next cycle is equal to zero. This is due to RC uses the first cycle as learning period, and no control signal is generated in this interval. This also means that standard RC gives zero response in the first cycle, but a good response in the second cycle and ahead.

The tracking output of PI is shown in green line. It gives fast response in the first cycle, but overshoot and oscillation still remain in every pulse change. After overshoot and oscillation occur, the tracking output moves and settles to the reference signal value. However, this steady state condition does not last longer due to pulse change. This implies that no matter how many cycles of reference signal have been tracked, zero tracking error can not be achieved by using PI controller.

The tracking output of PI with RC (Plug-in RC) is indicated by black dash line. In the first cycle, the tracking output shows better response compared to the RC's, and also shows similar response to the PI's. This is due to the RC control signal is still inactive, while PI control signal is

already active and do the tracking. On the second cycle, small overshoot still appear, but it is smaller than PI's. Starting at the third cycle, zero tracking error is achieved as shown in Figure 2(b).

The first scenario uses periodic signal with step type as a reference. The second scenario is that the plant output is required to track triangle reference signal. Triangle signal is ramp type signal that may effect to the tracking performance of PI.

For RC, tracking performance is similar to the performance as shown in the first scenario. An error equal to the reference signal appears in the first cycle, and zero tracking error is started at the second cycle.

For PI, the tracking output shows small overshoot for every peak change. After the overshoot occurs, the tracking output never settle to the reference value. This performance differs with the performance shown in the first scenario.

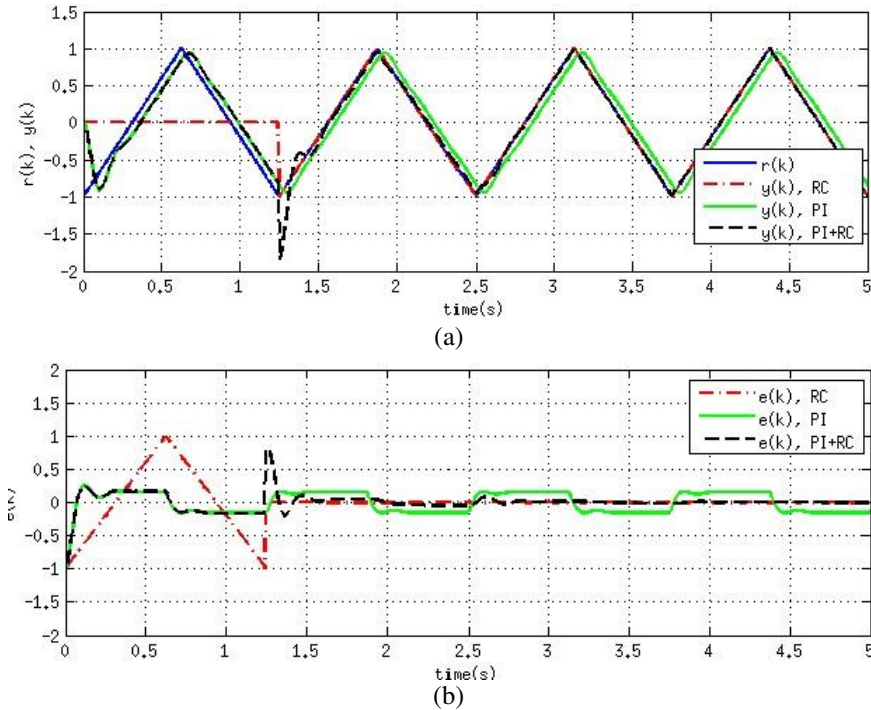


Figure 3. (a) Tracking Output of Triangle Signal, (b) Tracking Error.

This error is typical response when Type-1 system is fed by ramp input. When the plant model (9) is cascaded with PI controller, results in an open-loop model with a single integrator, which is a Type-1 system. Therefore, steady-state error of Type-1 system fed by ramp input is always non-zero, while steady-state error of Type-1 system fed by step input is zero. This can also be proven from final value theorem.

The tracking output of PI with RC shows similar performance compared to the performance shown in the first scenario.

In summary, PI with RC provides better tracking performance compared to both PI and RC. PI with RC combines the fast response feature of PI and learning feature of RC. Therefore, both fast response and zero-tracking error can be achieved by employing this controller. However, PI with RC gives higher order controller compared to both PI and RC. We can see from (11) (13), and (15) that the order of PI with RC is 252, while the order of RC and PI is 250 and 2 respectively.

5. Conclusion

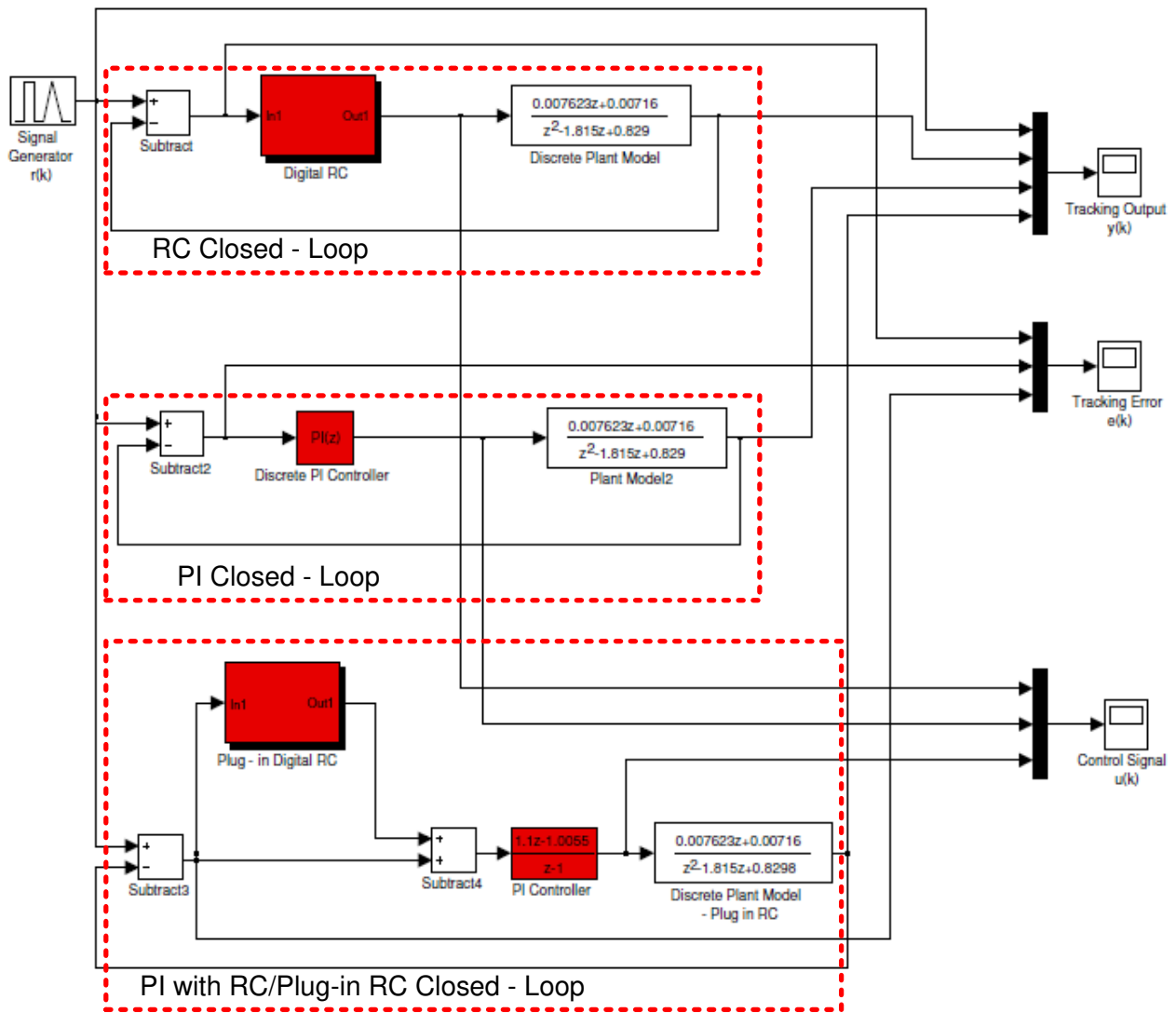
This paper presents the tracking performance comparison of PI controller, RC, and PI with RC. The transient response and tracking accuracy are also discussed. Overall, PI with RC provides better tracking performance compared to both PI and RC. This is due to PI with RC integrates the fast response feature of PI and learning feature of RC at once. In exchange, PI with RC gives higher order controller compared to both PI and RC.

References

- [1] P. R. Ouyang, V. Pano, and T. Dam, "PID contour tracking control in position domain," Industrial Electronics (ISIE), 2012 IEEE International Symposium on. pp. 1297–1302, 2012.
- [2] A. Preyadarchane and S. Alavandar, "Fractional order controller for trajectory tracking of a two degree of freedom robot manipulator," Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. pp. 1–4, 2013.

- [3] C. Knospe, "PID control," *Control Systems*, IEEE, vol. 26, no. 1, pp. 30–31, 2006.
- [4] K. J. Åström and T. Hägglund, "The future of PID control," *Control Eng. Pract.*, vol. 9, no. 11, pp. 1163–1175, Nov. 2001.
- [5] M. S. Tavazoei, "From Traditional to Fractional PI Control: A Key for Generalization," *Industrial Electronics Magazine*, IEEE, vol. 6, no. 3, pp. 41–51, 2012.
- [6] J. Sternby, "Repetitive Control Theory and Applications - A Survey," *proc 13th IFAC World Congr.*, vol. D, pp. 1–6, 1996.
- [7] C. Li, D. Zhang, and X. Zhuang, "Theory and applications of the repetitive control," *SICE Annu. Conf. Sapporo, Hokkaido Inst. Technol. Japan*, 2004, pp. 27–34, 2004.
- [8] K. K. Leang, S. Yingfeng, S. Song, and K. J. Kim, "Integrated Sensing for IPMC Actuators Using Strain Gages for Underwater Applications," *Mechatronics*, IEEE/ASME Trans., vol. 17, no. 2, pp. 345–355, 2012.
- [9] H. Liao, M. J. Roelle, C. Jyh-Shin, P. Sungbae, and J. C. Gerdes, "Implementation and Analysis of a Repetitive Controller for an Electro-Hydraulic Engine Valve System," *Control Syst. Technol. IEEE Trans.*, vol. 19, no. 5, pp. 1102–1113, 2011.
- [10] S. Zongxuan and K. Tang-Wei, "Transient Control of Electro-Hydraulic Fully Flexible Engine Valve Actuation System," *Control Syst. Technol. IEEE Trans.*, vol. 18, no. 3, pp. 613–621, 2010.
- [11] H. Chuxiong, Y. Bin, C. Zheng, and W. Qingfeng, "Adaptive Robust Repetitive Control of an Industrial Biaxial Precision Gantry for Contouring Tasks," *Control Syst. Technol. IEEE Trans.*, vol. 19, no. 6, pp. 1559–1568, 2011.
- [12] B. A. Francis and W. M. Wonham, "The internal model principle for linear multivariable regulators," *Appl. Math. Optim.*, vol. 2, no. 4, p. 380, 1975.
- [13] K. H. Ang, G. Chong, and Y. Li, "PID control system analysis, design, and technology," *Control Systems Technology*, IEEE Transactions on, vol. 13, no. 4, pp. 559–576, 2005.
- [14] M. Nakano and S. Hara, "Microprocessor-Based Repetitive Control," in *Microprocessor-Based Control Systems*, vol. 4, N. Sinha, Ed. Springer Netherlands, 1986, pp. 279–296.
- [15] C. Cosner, G. Anwar, and M. Tomizuka, "Plug in repetitive control for industrial robotic manipulators," in *Robotics and Automation, 1990. Proceedings., 1990 IEEE International Conference on*, 1990, pp. 1970–1975 vol.3.
- [16] M. Tomizuka, "ZERO PHASE ERROR TRACKING ALGORITHM FOR DIGITAL CONTROL," *J. Dyn. Syst. Meas. Control. Trans. ASME*, vol. 190, no. 1, pp. 65–68, 1987.
- [17] M. Tomizuka, T.-C. Tsao, and K.-K. Chew, "Analysis and Synthesis of Discrete-Time Repetitive Controllers," *J. Dyn. Syst. Meas. Control*, vol. 111, no. 3, pp. 353–358, 1989.
- [18] E. Kurniawan, "Robust Repetitive Control and Applications," *PhD Thesis*, Swinburne University of Technology, 2013.

Appendix



Block Diagram of SIMULINK model

Pengenalan Entitas User Profile Pada Twitter

Entity Recognition of User Profile on Twitter

Titin Pramiyati, Iping Supriana, Ayu Purwarianti

STEI-Institut Teknologi Bandung

Jl. Ganesha 10, Bandung, Indonesia

Email: titin.harsono@gmail.com; iping@informatika.org; ayu@informatika.org

Abstract

Trust scope attribute as an attribute to determine the level of trust resources, will be filled with the data contained in the user profile of Twitter –one of social media-known as Bio Twitter. However, these data should be in accordance with the characteristics and functions of each attribute, such as education attribute must be filled in with the information relating to the educational background of the owner of the profile. To obtain the data corresponding to the trust scope attributes, we perform named entity recognition, which is one of the activities in the process of information extraction. Therefore, this paper describes the results of the entity recognition process performed on data contained in the user profile. Software used to recognize the data as an entity is Indonesia Netagger, which is to perform entity recognition that written in Indonesian language. The software recognizes only five entities namely Person, Organization, Location, Position and Other. We carried out the research by conducting four stages namely entity recognition-with original data-Bio Twitter, error identification, formalizing data, and final test. The results show the success of entity recogniton as follow; Person entity is recognized correctly by 71% of the total data available, the entity Organization recognized correctly by 50%, 20% Position entity recognized correctly, and 50% recognized correctly as Location entity.

Keywords: *trust scope atributes, level trust, user profile, named entity recognition.*

Abstrak

Atribut *trust scope* sebagai atribut untuk menentukan tingkat kepercayaan sumber informasi, akan diisi dengan data yang terdapat pada *user profile Twitter* yang dikenal sebagai *Bio Twitter*. Hanya saja, data tersebut harus sesuai dengan karakteristik dan fungsi dari masing-masing atribut *trust scope*, seperti atribut pendidikan harus diisi dengan informasi yang berkaitan dengan latar belakang pendidikan dari pemilik profil tersebut. Untuk mendapatkan data yang sesuai dengan atribut, kami melakukan *named entity recognition*, yang merupakan salah satu kegiatan pada proses ekstraksi informasi. Oleh karena itu, paper ini menjelaskan hasil proses pengenalan entitas yang dilakukan terhadap data yang terdapat pada *user profile*. Perangkat lunak yang digunakan untuk mengenali data sebagai entitas adalah *IndonesiaNetagger*. *IndonesiaNetagger*, merupakan perangkat lunak untuk mengenali entitas yang ditulis dalam bahasa Indonesia. Kami melakukan penelitian dalam empat tahap, yaitu pengenalan *entity* dengan data *Bio twitter* yang asli, identifikasi kesalahan proses pengenalan, formalisasi data dan pengujian pengenalan entitas akhir. Hasil penelitian menunjukkan keberhasilan sebagai berikut; entitas *Person* dikenali dengan benar adalah sebesar 71% dari total data entitas yang tersedia, entitas *Organization* dikenali dengan benar sebesar 50%, entitas *Position* 20% dikenali dengan benar, dan 50% entitas *Location* dikenali dengan benar.

Kata kunci: atribut *trust scope*, tingkat kepercayaan, pemilik profil, pengenalan entitas

1. Pendahuluan

User profile adalah tampilan visual dari data pribadi yang dikaitkan dengan pengguna tertentu, dan dapat dianggap sebagai representasi dari sebuah model pengguna. Sebuah *user profile* akan mengacu pada representasi digital identitas

seorang secara eksplisit. Oleh karenanya *user profile* menjadi salah satu layanan yang disediakan pada beberapa layanan internet seperti media sosial.

User profile menyediakan berbagai informasi yang berkaitan dengan jati diri penggunanya, termasuk diantaranya pernyataan tentang tempat bekerja, siapa yang mereka kenal, tempat tinggal, riwayat pendidikan dan sebagainya, umumnya fasilitas ini digunakan untuk berbagi informasi

Received: 26 Februari 2015; Revised: 16 Maret 2015;
Accepted: 20 Maret 2015 ; Published online: 30 April 2015
©2014 INKOM 2014/14-NO411

pribadi dengan teman, kerabat, pegawai, atau kepada dunia.

Pengguna internet merupakan sumber informasi yang potensial untuk dimanfaatkan, sehingga ketersediaan informasi tidak hanya berasal dari organisasi resmi saja, akan tetapi masing-masing pengguna dapat mewakili masyarakat untuk berpartisipasi menyediakan informasi yang berkualitas dan dapat dipercaya.

Informasi yang dipercaya dapat diperoleh berdasarkan pada kepercayaan yang dimiliki oleh sumber informasi [1], reputasi sumber informasi [2], dan kepercayaan yang diberikan oleh entiti dengan memperhatikan tingkat kepercayaan (*trust level*) yang dimiliki oleh entiti tersebut [3].

Berbagai model kepercayaan telah banyak dibangun untuk menentukan tingkat kepercayaan, diantaranya model kepercayaan untuk menilai kepercayaan pengguna terhadap aplikasi [4], penentuan kepercayaan pengguna internet [5], dan mekanisme penentuan kepercayaan dan reputasi untuk pengambilan keputusan mendapatkan rekan kerja tanpa harus mengenalnya terlebih dahulu [6].

Terdapat dua jenis penilaian kepercayaan yaitu *direct trust* dan *recommended trust* [7], *direct trust* adalah penilaian kepercayaan berdasarkan pada interaksi langsung yang terjadi, sementara *recommended trust* diperoleh berdasarkan reputasi.

Penilaian kepercayaan langsung, dapat ditentukan berdasarkan *feedback* yang diberikan oleh entiti atau pengguna lain saat terjadi interaksi [1], berdasarkan padakekerapan interaksi yang terjadi [8], dapat ditentukan berdasarkan pada konteks [7] dan *trust scope* [9].

S.Ibotombi Singh dan Smriti K. Sinha [7] menyatakan bahwa kepercayaan dapat dibangun berdasarkan pada *context-sensitive*, *transferable*, *dynamic* dan *history-based*. Penggunaan *context-sensitive* dalam membangun kepercayaan dapat menjadikan suatu *agent* dipercaya pada satu konteks, dan tidak dipercaya pada konteks lain. Sebagai contoh adalah konteks pada layanan pemesanan tiket perjalanan dapat didefinisikan dengan atribut *keberangkatan*, *tujuan*, *tanggal keberangkatan*, dan *kelas*. Sedangkan penggunaan *transferable* dalam menentukan kepercayaan dimaksudkan untuk memberi kepercayaan berdasarkan pada satu konteks untuk memberi kepercayaan pada konteks lainnya, misal seorang yang dipercaya sebagai seorang politisi unggul, akan dipercaya juga sebagai pembicara unggul pula, hal ini karena keterkaitan konteks profesi politisi dengan kemampuannya sebagai pembicara.

Thirunarayan dkk [9] menentukan lingkup kepercayaan atau *trust scope* pada jaringan sosial dan interaksi yang terjadi. Pada kajiannya,

Thirunarayan menyatakan bahwa lingkup kepercayaan atau *trust scope* dapat menangkap konteks, kegiatan, fungsi atau domain dari hubungan kepercayaan yang terjadi, misal seorang A mempercayai seorang B karena kemampuan yang dimiliki B, kemampuan B akan menjadi rekomendasi dalam lingkup kepercayaan karena pengetahuan yang dimiliki oleh B. Lingkup kepercayaan yang didasarkan pada pengetahuan yang dimiliki oleh seseorang disebut *referral trust*.

Demikian halnya jika seorang A mempercayai seorang B karena dapat melakukan pekerjaan sesuai dengan lingkup kepercayaannya, maka kepercayaan yang didasarkan pada kemampuan dalam melakukan pekerjaan disebut *functional trust*.

Pada masalah tertentu, seperti penentuan kepercayaan untuk seorang teknisi atau mekanik, *referral trust* dan *functional trust* dapat digunakan secara bersama, karena akan memberikan lingkup kepercayaan yang lebih baik, dibandingkan jika hanya menggunakan salah satu *referral trust* atau *functional trust*. Penggunaan salah satu lingkup kepercayaan akan memberi kemungkinan pemberian kepercayaan yang kurang tepat.

Salah satu properti yang digunakan pada model kepercayaan yang diusulkan oleh Kyounghee Jung adalah properti *interaction significance based on knowledge*. Properti ini digunakan untuk mendapatkan nilai kepercayaan perorangan (*personal trust*), yaitu penilaian yang diberikan oleh pengguna sebuah layanan berdasarkan pada pengetahuan yang dimiliki oleh penggunanya, karena perbedaan pengetahuan akan memberikan nilai yang berbeda untuk tiap pengguna [10].

Selanjutnya, memperhatikan pembahasan penentuan lingkup kepercayaan berdasarkan pada aspek konteks, *transferable*, pengetahuan dan kemampuan dalam menyelesaikan tugas, maka *referral trust* dan *functional trust* dapat mewakili ke-empat aspek tersebut. Penentuan lingkup kepercayaan yang didasarkan pada aspek konteks dan pengetahuan dimasukkan ke dalam *referral trust*, sedangkan *transferable*, kemampuan dan pengetahuan dapat dimasukkan ke dalam *functional trust*.

Untuk menentukan atribut apa saja yang dapat dijadikan sebagai atribut *trust scope*, telah dilakukan survey yang melibatkan 257 responden dan atribut *user profile* dari 4 media sosial yaitu *Facebook*, *Google+*, *Twitter* dan *LinkedIn*. Hasil yang diperoleh dari survey adalah terdapat 8 atribut *trust scope*, yaitu atribut pendidikan, tempat pendidikan, pekerjaan, tempat bekerja, profesi, jabatan, minat dan komunitas.

Twitter adalah layanan jejaring sosial yang lebih dikenal sebagai media sosial yang memungkinkan

penggunanya membuat akun tanpa harus membayar dan dapat mengirim dan membaca pesan teks hingga 140 karakter, yang disebut sebagai kicauan (*tweet*).

Kicauan pengguna *Twitter* dapat terlihat oleh pengguna lain walaupun tidak terdapat ikatan berlangganan dengan cara mengikuti (*follow*) pengguna yang bersangkutan. *Twitter* membedakan pengguna terhadap pengguna yang diikuti sebagai kelompok *following* dan pengguna yang mengikuti adalah kelompok *follower*.

Berdasarkan struktur koneksi pada *Twitter*, berikut kategori utama dari pengguna *Twitter*:

- a. Sumber informasi (*Information Source*), pengguna *Twitter* pada kategori ini memiliki pengikut yang besar. Pengguna ini dapat melakukan posting dengan interval umum atau jarang. Meskipun jarang melakukan perubahan informasi, pengguna dengan pengikut yang besar menjadikan perubahan informasi menjadi bernilai. Beberapa sumber informasi juga ditemukan menjadi alat posting berita otomatis dan informasi berguna lain di *Twitter*
- b. Teman (*Friends*), pada umumnya hubungan yang terdapat di *Twitter* berada pada kategori ini, dan terdapat berbagai sub-kategori pertemanan yang tersedia, sebagai contoh seorang pengguna dapat memiliki teman, family dan rekan kerja pada daftar pertemanan atau pengikut. Kadangkala pengguna yang tidak kenal dapat menambahkan seseorang sebagai teman
- c. Pencari informasi (*Information seeker*), adalah seseorang yang jarang melakukan posting, tetapi mengikuti pengguna lain secara reguler.

Sebagai media sosial, *Twitter* juga menyediakan layanan *user profile* yang dikenal sebagai *Bio Twitter*. Berbeda dengan layanan *user profile* pada media sosial lainnya, *Bio Twitter* tidak memisahkan setiap data ke dalam atribut tertentu, sehingga isi dari *Bio Twitter* terlihat seperti sebuah dokumen teks.

Untuk mendapatkan data yang sesuai dengan kriteria atribut *trust scope*, perlu dilakukan pemisahan data yang terdapat pada *Bio Twitter*. Proses pemisahan data ini dapat dilakukan dengan menggunakan proses pengenalan entitas (*Named Entity Recognition*).

Named Entity Recognition (NER) memiliki peran penting pada area aplikasi *Natural Language Processing* (NLP) yang banyak tersedia, seperti ekstraksi informasi, *retrieval information*, tanya-jawab dan peringkasan otomatis. Ciri utama dari tugas NER adalah melakukan identifikasi dan membuat *tag context* pada kata-kata yang tersedia berdasarkan pada kemungkinan kombinasi atas

kata-kata tersebut, seperti penentuan panjang minimal kata *word* yang akan diidentifikasi sebagai nama entiti, kata awal, dan sebagainya [11].

Metoda NER diklasifikasikan ke dalam tiga kategori, yaitu *rule-based method*, *statistical-based method*, dan *rule-statistical combined method*. *Rule-based method* menggunakan aturan yang dibuat, dan mengidentifikasi nama entiti yang berbeda dengan cara mencocokkan kata dengan aturan yang telah ditentukan. *Statistical-based method* menggunakan korpus yang dianotasi untuk menentukan peluang sebuah kata sebagai nama entiti, jika nilai peluang sebuah kata lebih besar dari nilai *threshold* yang ditentukan, maka kata tersebut akan diidentifikasi sebagai nama entiti. *Rule-statistical combined method*, adalah metoda yang mengkombinasikan antara kedua metoda *rule-based* dan *statistical-based*, seperti penggunaan *rule-conditional random field* (CRF) *combined method* [12].

Named entity seperti *Person*, *Organization*, *Position*, dan *Location* dalam proses identifikasi akan membutuhkan ciri (*feature*) yang merefleksikan properti dari sebuah nama entiti, seperti tipe, kemunculan dan berbagai ukuran umum, baik untuk skala dokumen maupun korpus. Salah satu contoh penggunaan *feature* dalam penentuan nama entiti adalah *feature* kemunculan sebuah kata pada urutan pertama (*first sentence occurrence*), karena urutan kemunculan kata dapat menentukan tingkat kepentingan dari kata tersebut [13].

Penelitian yang dilakukan Khodra dan Purwarianti (2013) menggunakan vektor fitur untuk model klasifikasi dan untuk tiap token pada proses ekstraksi informasi dari transaksi *online* di *Twitter*. Vektor fitur yang digunakan pada model klasifikasi menggunakan kategori yang dibangun dengan berdasarkan notasi BIO (*Begin In Other*)-<jenis informasi>, sedangkan vektor fitur untuk tiap token didefinisikan berdasarkan atribut leksikal token tersebut dan tetangganya [14].

Notasi BIO juga digunakan pada sistem NER dengan kerangka kerja CRF (*Conditional Random Field*), untuk memberikan *tag* yang sudah dikenali pada setiap karakter atau kata bahasa China yang terdapat pada dokumen input untuk mengidentifikasi nama entiti [15].

Berdasarkan fungsi *trust scope* dalam menentukan tingkat kepercayaan sumber informasi, dan kebutuhan akan kebenaran data yang sesuai dengan karakteristik dari atribut *trust scope*, maka paper ini akan membahas hasil pengujian yang dilakukan pada proses pengenalan entitas untuk merepresentasikan data *user profile* ke dalam atribut *trust scope*, menggunakan *Bio Twitter*

sebagai sumber data dan menggunakan perangkat lunak *IndonesiaNetagger* dalam proses pengenalan entitas untuk sumber data dalam Bahasa Indonesia.

2. Metodologi dan Pengumpulan Data

Pengujian yang dilakukan merupakan proses pengenalan entitas terdiri dari beberapa tahap yaitu pengenalan entiti dengan data *Bio twitter* yang asli, identifikasi kesalahan proses dan pengenalan, formalisasi data dan pengujian pengenalan entitas akhir.

Proses pengenalan entiti terhadap data *Bio Twitter* yang asli dilakukan untuk mengetahui proses pengenalan entitas dapat berjalan dengan baik atau tidak. Hal ini dikarenakan data yang terdapat pada *Bio Twitter* dapat ditulis dengan bentuk yang disukai oleh penggunanya.

Tahap identifikasi kesalahan proses pengenalan dilakukan untuk mengidentifikasi hal-hal yang menimbulkan kegagalan proses pengenalan, dan sekaligus dipakai untuk melakukan perbaikan agar proses pengenalan tidak mengalami kegagalan.

Tahap formalisasi data adalah tahap perbaikan atas sumber data (korpus) yang digunakan berdasarkan hasil identifikasi kesalahan selama proses pengenalan. Tahap terakhir adalah pengujian akhir jika formalisasi data telah selesai dikerjakan, untuk mengetahui hasil pengenalan entitas yang dilakukan.

Proses pengenalan entitas ini menggunakan data *user profile Bio Twitter* yang dikumpulkan melalui pengambilan data *Bio Twitter* secara manual, dan pengambilan dengan memanfaatkan *Application Program Interface (API)* yang disediakan oleh *Twitter*.

Perangkat yang digunakan pada pengujian proses pengenalan entitas ini adalah perangkat lunak *IndonesiaNETagger* [16] yaitu perangkat untuk melakukan pengenalan entitas untuk korpus Bahasa Indonesia.



Gambar 1. Korpus (BioTwitter.txt).

Pengumpulan data *Bio Twitter* yang dilakukan secara manual berasal dari 20 akun *Twitter* yang dipilih secara acak. Hasil dari pengumpulan data dijadikan sebagai sebuah korpus dengan format file teks (.txt) seperti terlihat pada Gambar 1.

Sedangkan pengumpulan data *Bio Twitter* yang diambil dengan menggunakan API, merupakan *Bio Twitter* dari kelompok *following* dan *follower* dari akun *Twitter* tertentu, data yang berhasil diambil disimpan pada sebuah relasi (*file*) dan diberi nama *UserTweetBio(DBMS:Microsoft Access)*, seperti terlihat pada Gambar 2.

username	name	followers_count	description	status
aceanonymus	Rian A.	145	180 SELECT * FROM Bio WHERE Username = 'aceanonymus'	11195
SBYudhoyono	S. B.	152	6099887 Akun Resmi Presiden Ke-6 RI (2004-2014) Susilo Bambang Yudhoyono. Dikelola oleh Staf Pribadi. Tweet dari	3413
johnriady	John R.	471	4351 FH UPH. BeritaSatu. Columbia University Law School (JD 2011). Wharton School (MBA 2008). Georgetown U	477
dha_jy	dha-d	153	104 imple	347
moeldika	Leljen	20	772 The Official Twitter Account Of Moeldika, Panglima TNI (Tentara Nasional Indonesia) REPUBLIK INDONESIA	2
raindra18	Israni	357	105 Fans berat ayah @dharmaawan234 & kaka @ibairasyidharmawan	173
felixsiauw	Felix S.	318	1086075 penulis, pembelan dakwah, bersama yang menginginkan tegaknya syariah-khilaafah, hamba yang sangat t	46990
Fachdian	Fachd	614	263 Business Development Manager at Rebel Creative Syndicate, President at IT Gema Solusindo, CEO at OLS, M	996
hamsanzoeila	Hams	3549	22179 Ketua Mahkamah Konstitusi Republik Indonesia, Dosen Magister Hukum UIA Jakarta, Guest Professor Chin	189
hatterajasa	Hatta	67	924721 Ketua Umum Partai Amanat Nasional	1322
ni_magasa	Reni N.	119	321 @ajukibiniu	4785
snaptu	Snapti	628	3348746 We are now part of Facebook and our great technology has become the official Facebook app for Every Ph	3549
ugang_arung	agung	59	50	27
NajwaShihab	Najwa	474	1497099 TV Anchor, Host of a weekly program 'Mata Najwa', every Wednesday 8 PM on Metro TV, Indonesia	2021
Witonsu	Witons	185	245 The Official Twitter Page of wisnu:statorji line wisnu:mas instagram wisnu:mas	8328
dimas699	Dimas	127	204 "CEMAN"	12827
DidiJembroz	Didi	83	42 Nothing impressed me much.	282
awanadiprakoso	awan	1508	437 sampai kapan pun akan terus bermusik dan selalu main bass aminn	4455
rindywidia	rindy	129	125	815
eni_wijanaric	Ania w	103	61	19
vhviva	Novia	552	363 khiaa, tauakkal, bersyukur & "YA ALLAH SWT & "MAs ipaa & "Family belajar mengerti, belajar bjaksana	3803
UberDoc	Uber5	2307	11870499 A full-featured, customizable Twitter app just for you. For support, please submit tickets here: http://t.co/	34523

Gambar 2. File UserTweetBio.

Data *Bio Twitter* yang diambil seperti terlihat pada Gambar 3, yaitu akun @anismatta. Data yang tercantum pada *Bio Twitter* terdiri dari teks dan beberapa karakter khusus yang digunakan seperti simbol))((yang dibuat dari beberapa karakter khusus..

Data *Bio Twitter* yang diambil umumnya menggunakan bahasa Indonesia, akan tetapi ada beberapa data *Bio Twitter* milik akun tertentu yang menuliskan profesi, nama sekolah dan data lainnya dalam bahasa Inggris.

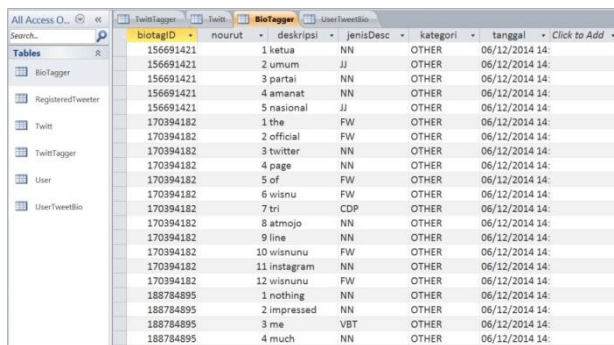


Gambar 3. Bio Twitter pada akun @anismatta.

3. Pengenalan Entitas Menggunakan Perangkat Lunak *IndonesiaNETagger*

Uji coba pertama penggunaan perangkat lunak pengenalan entitas dengan nama *IndonesiaNETagger*, untuk melakukan pengenalan entitas pada data Bio yang terdapat pada relasi *UserTweetBio*, dengan hasil uji coba adalah *IndonesiaNETagger* mengalami kegagalan proses. Langkah penanganan kesalahan proses dilakukan dengan cara menghilangkan semua karakter khusus dan merubah semua huruf pada kata menjadi huruf kecil (*lower case*).

Setelah dilakukan perubahan, ujicoba diulang kembali dengan hasil proses pengenalan berjalan dengan baik, akan tetapi semua kata dikenali sebagai entitas *Other*, seperti terlihat pada Gambar 4.



biotagID	nourut	deskripsi	jenisDesc	kategori	tanggal
156691421	1	ketua	NN	OTHER	06/12/2014 14:
156691421	2	umum	JJ	OTHER	06/12/2014 14:
156691421	3	partai	NN	OTHER	06/12/2014 14:
156691421	4	amanat	NN	OTHER	06/12/2014 14:
156691421	5	nasional	JJ	OTHER	06/12/2014 14:
170394182	1	the	FW	OTHER	06/12/2014 14:
170394182	2	official	FW	OTHER	06/12/2014 14:
170394182	3	twitter	NN	OTHER	06/12/2014 14:
170394182	4	page	NN	OTHER	06/12/2014 14:
170394182	5	of	FW	OTHER	06/12/2014 14:
170394182	6	wisnu	FW	OTHER	06/12/2014 14:
170394182	7	tri	COP	OTHER	06/12/2014 14:
170394182	8	atmojo	NN	OTHER	06/12/2014 14:
170394182	9	line	NN	OTHER	06/12/2014 14:
170394182	10	wisnu	FW	OTHER	06/12/2014 14:
170394182	11	instagram	NN	OTHER	06/12/2014 14:
170394182	12	wisnu	FW	OTHER	06/12/2014 14:
188784895	1	nothing	NN	OTHER	06/12/2014 14:
188784895	2	impressed	NN	OTHER	06/12/2014 14:
188784895	3	me	VBT	OTHER	06/12/2014 14:
188784895	4	much	NN	OTHER	06/12/2014 14:

Gambar 4. Hasil ujicoba menggunakan sumber data *UserTweetBio*.

Jika semua kata dikenali sebagai entitas *Other*, maka dapat dikatakan bahwa proses pengenalan tidak berhasil, karena proses pengenalan dengan menggunakan *IndonesiaNETagger* harus berhasil mengenali 5 kategori entitas, yaitu *Person*, *Organization*, *Location*, *Position* dan *Other*.

Ujicoba berikutnya menggunakan data korpus *BioTwitter.txt*, yang dibuat secara manual dan belum mengalami perubahan. Ujicoba ini juga mengalami kegagalan proses ketika dilakukan proses pengenalan dengan *IndonesiaNETagger*. Untuk mengetahui sebab kegagalan, dilakukan ujicoba berulang secara manual, selain mengidentifikasi berbagai potensi kegagalan pengenalan, juga dilakukan perubahan pada pada korpus untuk menghindari kegagalan proses.

4. Identifikasi Kesalahan Pengenalan

Berdasarkan pada kegagalan yang terjadi pada proses pengenalan entitas, dilakukan eksperimen dengan menggunakan korpus (*BioTwitter.txt*) yang telah dibuat seperti terlihat pada Gambar 3, ujicoba

mengalami kegagalan proses pengenalan. Kegagalan proses bukan disebabkan oleh perangkat lunak *IndonesiaNETagger*, akan tetapi karena file teks yang digunakan berisi data dengan format penulisan seperti tertulis pada *Bio Twitter*.

Format penulisan *Bio Twitter* yang memberikan kebebasan kepada pemilik untuk menuliskan sesuai dengan keinginan mereka, merupakan identifikasi awal yang menyebabkan kegagalan proses, seperti pada contoh berikut ini:

@fadlizon Waki Ketua DPR-RI; Wakil Ketua Umum DPP Partai @Gerindra; Sekjen DPN @HKTI; @FadliZonLibrary; Dewan Redaksi HORISON; Ketua @ILUNIFIBUI

Untuk mengetahui kesalahan yang menyebabkan kegagalan proses, isi korpus direvisi dan diuji kembali berulang sampai tidak terjadi kegagalan proses. Dari pengujian ini diperoleh beberapa format yang tidak dapat dieksekusi oleh *IndonesiaNETagger*, diantaranya:

- Kata yang terdapat huruf kapital yang diapit oleh huruf non kapital, seperti *PressCode*
- Penggunaan *double hyphen* (--) atau *dash* (-), *slash* (/), *hashtag* (#), dan *pipe* (|)
- Penggunaan karakter @ yang diikuti dengan kata, seperti @tangandiatas
- Kata yang mengandung angka setelah huruf, seperti S2, S3
- Penulisan URL (<http://SekolahMonyet.com>)

Hal yang sangat menarik pada pengujian ini tidak dikenalnya kata *group*, yang menyebabkan proses NER mengalami kegagalan. Hal ini menjadi unik, karena beberapa kata dalam bahasa Inggris tetap dikenali sebagai entiti walaupun tidak sesuai.

Berdasarkan hasil ujicoba, isi korpus direvisi secara manual, yaitu dengan menghilangkan dan memperbaiki beberapa kata yang menjadi sebab kesalahan, misal pada kata S3 diganti menjadi doktor, kata *group* dirubah menjadi grup, dan menghilangkan karakter khusus dengan hasil revisi seperti pada file *BioTwitterCoba.txt*.

Ujicoba kembali dilakukan dengan menggunakan korpus baru yang telah disesuaikan, dan perangkat lunak tidak mengalami kegagalan proses. Berdasarkan hasil dari proses identifikasi kesalahan ini, beberapa bentuk penulisan yang terdapat pada *Bio Twitter* akan dijadikan sebagai dasar dalam proses formalisasi data.

5. Formalisasi Data

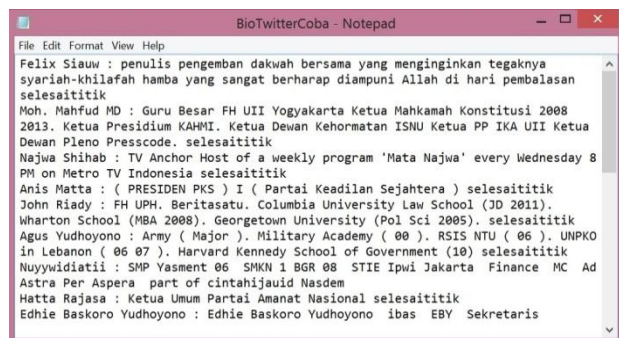
Berdasarkan pada hasil identifikasi kesalahan, dilakukan proses formalisasi data yang bertujuan untuk merubah format asli dari *Bio Twitter* menjadi

format yang dapat dieksekusi oleh perangkat *IndonesiaNetagger*. Perubahan ini dilakukan secara otomatis, untuk memenuhi kebutuhan tersebut disiapkan satu program untuk digunakan pada proses formalisasi data.

Perubahan format yang dilakukan pada proses formalisasi data, meliputi proses penghapusan karakter khusus yang terdapat pada *Bio Twitter*. Perubahan yang dilakukan adalah mengganti setiap karakter khusus dengan spasi, dan penggunaan *regular expression* (regex) untuk menangani kesalahan yang disebabkan adanya huruf besar yang diapit oleh huruf kecil pada sebuah kata, seperti terlihat pada penggalan program di bawah ini.

```
String isifile =
FileHelper.openDoc(jTextFieldCorpus.getText());
isifile = isifile.replaceAll("@", " ");
isifile = isifile.replaceAll("#", " ");
isifile = isifile.replaceAll("/", " ");
isifile = isifile.replaceAll(".", " ");
isifile = isifile.replaceAll("\\|+", " ");
for (String string : isi) {
if (string.matches("[A-Za-z]+[0-9]+")) {
String[] s =
string.split("(?<=\\d) (?=\\d) | (?<=\\d) (?=\\d)");
;
string = "";
for (String splitted : s) {
string += splitted;
string += " ";
}
}
System.out.println(string+"huruf diikuti
angka");
}
if (string.matches("[0-9]+[A-Za-z]+")) {
System.out.println(string+"huruf diikuti
angka");
}
if (string.matches("[A-Z]+[a-z]+")) {2,99}} {
string = string.charAt(0) +
string.toLowerCase().substring(1,
string.length());
System.out.println(string+"huruf alay");
}
```

Hasil dari perubahan ini dijadikan sebagai korpus baru dalam format teks dengan nama (*BioTwitterCoba.txt*) seperti terlihat pada Gambar 5, sebagai sumber data pada tahap pengujian akhir.



Gambar 5. Hasil formalisasi data

6. Pengujian Akhir

Proses pengenalan entitas menggunakan korpus yang telah diformalisasi berhasil dilakukan, dengan hasil pengenalan entitas *Person* sebesar 71% dikenali dengan benar. Sebuah kata akan dikenali sebagai entitas *Person* jika kata tersebut merupakan sebuah nama. Keberhasilan ini dihitung berdasarkan jumlah kata nama yang dikenali dengan benar dibagi dengan jumlah kata nama.

Sedangkan jumlah kata yang dikenali dengan benar sebagai entitas *Organization*, sebesar 50% dari 36 kata yang berindikasi nama organisasi, seperti kata Universitas, Partai, DPR dan sebagainya. Seperti pada kalimat Partai Keadilan Sejahtera, pengenalan yang dibenar sebagai entitas *Organization* hanya pada kata Partai dan Keadilan, sedangkan kata Sejahtera dikenali sebagai entitas *Person*.

Hasil uji coba yang terlihat pada Tabel 1, merupakan hasil pengenalan kata yang menunjukkan lokasi sebagai entitas *Location*. Kata yang dikenali dengan benar sebagai entitas *Location* sebesar 50 %. Umumnya kata yang terdapat pada korpus adalah nama negara dan kota.

Tabel 1. Hasil pengenalan entitas Location

Kata	POS	NER
Yogyakarta	NNP	LOCATION-B
Indonesia	NNP	LOCATION-B
Columbia	NN	OTHER
Lebanon	NNP	LOCATION-B
Jakarta	NNP	LOCATION-B
Bekasi	NNP	PERSON-B
Indonesia	NNP	OTHER
China	NNP	PERSON-I
Beijing	NNP	LOCATION-B
China	NNP	PERSON-B

Tabel 2 menunjukkan hasil pengujian pengenalan entitas untuk kata yang menunjukkan posisi atau jabatan atau entitas *Position*. Keberhasilan pengenalan entitas hanya sebesar 20%, hal ini dikarenakan adanya perbedaan pengenalan antara jabatan dan pekerjaan. Seperti pada kata Dosen atau Professor yang dikenali sebagai entitas *Other* dan *Person*. Karena kata *Dosen* tidak dikenali sebagai jabatan tetapi sebagai pekerjaan atau profesi, sedangkan kata *Profesor* lebih memberi makna sebagai gelar yang melekat pada nama seseorang sehingga dikenali sebagai entitas *Person*.

Tabel 2 Hasil pengenalan entitas Position

Kata	POS	NER
Penulis	NN	OTHER
Guru	NN	OTHER
Ketua	NN	POSITION-B
Ketua	NN	OTHER
PRESIDEN	NN	POSITION-B
Umum	JJ	POSITION-I
Sekretaris	NN	POSITION-B
Jenderal	NN	POSITION-I
Bupati	NNP	OTHER
Konsultan	NN	OTHER
Rektor	NN	PERSON-I
Sekjen	NN	OTHER
Penulis	NN	OTHER
Aktor	NN	OTHER
Sutradara	NN	OTHER
Ketua	NN	PERSON-I
Jurnalis	NN	OTHER
Mahasiswa	NN	PERSON-I
Dosen	NN	OTHER
Professor	NNP	PERSON-I

7. Kesimpulan dan Penelitian Selanjutnya

Data yang terdapat pada *Bio Twitter* dapat digunakan sebagai sumber data untuk mengisi atribut *trust scope*, yang akan digunakan pada penentuan tingkat kepercayaan sumber informasi. Penggunaan *Bio Twitter* sebagai sampel dalam pengujian ini dikarenakan format penulisan pada *Bio Twitter* tidak menggunakan format yang baku dan setiap elemen data tidak dipisahkan sebagai atribut tersendiri seperti pada media sosial *Facebook*, *Google+* dan *LinkedIn*

Sehingga format yang tidak terstruktur ini dapat dijadikan dasar dalam menentukan aturan yang akan digunakan ketika dilakukan proses pengenalan entitas. *IndonesiaNETagger* sebagai perangkat lunak dengan fungsi untuk mengenali entitas untuk data dalam Bahasa Indonesia dapat digunakan dalam proses pengenalan data *Bio Twitter*, walaupun keberhasilan masih belum maksimal untuk 4 kategori entitas.

Berdasarkan hasil ujicoba ini, beberapa penelitian lanjutan sedang dikerjakan, diantaranya adalah pembuatan aturan (*rule*) yang akan digunakan untuk mengenali 8 entitas sesuai dengan atribut *trust scope*. Pembuatan aturan terdiri dari 2 jenis, yaitu jenis pertama pembuatan aturan dengan memanfaatkan hasil pengenalan entitas dengan 4 kategori yang telah tersedia.

Jenis pertama ini akan memanfaatkan entitas *Organization* dalam menentukan entitas Tempat Pendidikan, Tempat Bekerja, dan Komunitas dimana ketiga entitas ini adalah atribut *trust scope*. Sedangkan entitas *Position* dapat dimanfaatkan dalam penentuan entitas Jabatan dan Profesi.

Entitas *Location* dapat juga dimanfaatkan dalam penentuan Tempat Pendidikan, Tempat Bekerja, dan Jabatan.

Sedangkan aturan kedua akan dibuat untuk dapat langsung menentukan entitas sesuai dengan atribut *trust scope*, dan penelitian ini sedang dikerjakan untuk 3 entitas pertama yaitu Pendidikan, Tempat Pendidikan, dan Pekerjaan.

Daftar Pustaka

- [1] Y. Gil and V. Ratnakar, "Trusting Information Sources One Citizen at a Time," *Proceeding First Int. Semant. Web Conf.*, 2002.
- [2] S. Javanmardi and C. V. Lopes, "Modeling Trust in Collaborative Information Systems," *Evolution (N. Y.)*, 2007.
- [3] V. Tundjungsari, J. E. Istiyanto, E. Winarko, and R. Wardoyo, "A Reputation based Trust Model to Seek Judgment in Participatory Group Decision Making," *Int. Conf. Distrib. Framew. Multimed. Appl.*, 2010.
- [4] J. Matysiewicz, "Consumer trust – challenge for e-healthcare," *Management*, pp. 337–342, 2009.
- [5] L. Wen, P. Lingdi, L. Kuijun, and C. Xiaoping, "Trust Model of Users' behavior in Trustworthy Internet *," *Wase Int. Conf. Inf. Eng.*, pp. 403–406, 2009.
- [6] Q. X. Bo Zhang, Yang Xiang, "Trust and Reputation based Model Selection Mechanism for Decision-making," *Second Int. Conf. Networks Secur. Wirel. Commun. Trust. Comput.*, pp. 14–17, 2010.
- [7] S. I. Singh and S. K. Sinha, "A New Trust Model based on Social Characteristics and Reputation Mechanisms using Best Local prediction Selection Approach," *Int. Conf. New Trends Inf. Serv. Sci.*, 2009.
- [8] E. Marchetti, T. A. Faedo, L. Schilders, and S. Winfield, "Scenario-based testing applied in two real contexts : Healthcare and Employability," 2011.
- [9] K. Thirunarayan, P. Anantharam, C. A. Henson, and A. P. Sheth, "Some Trust Issues in Social Networks and Sensor Networks," *IEEE*, vol. 978–1–4244, pp. 573–580, 2010.
- [10] K. Jung and Y. Lee, "Autonomic Trust Extraction for Trustworthy Service Discovery in Urban Computing," *2009 Eighth IEEE Int. Conf. Dependable, Auton. Secur. Comput.*, vol. 978–0–7695, pp. 502–507, Dec. 2009.
- [11] A. Ekbal, S. Saha, and D. Singh, "Ensemble based Active Annotation for Named Entity Recognition," pp. 331–334, 2012.
- [12] X. Su, S. Mo, H. Wang, and X. Zhang, "Discovering Significant Persons , Locations and Organizations through Named Entity Ranking," pp. 328–331, 2012.
- [13] S. Wang and J. Feng, "A FRAMEWORK FOR ANALYZING THE ' INFORMATION BEARING CAPABILITY ' OF AN INFORMATION SYSTEM," *Mach. Learn.*, no. August, pp. 19–22, 2007.

- [14] M. L. Khodra and P. Ayu, "Ekstraksi Informasi Transaksi Online pada Twitter," *Cybermatika*, vol. 1, no. 1, 2013.
- [15] X. Wu, Z. Wu, J. Jia, L. Cai, and C. Science, "ADAPTIVE NAMED ENTITY RECOGNITION BASED ON CONDITIONAL RANDOM FIELDS WITH AUTOMATIC UPDATED DYNAMIC GAZETTEERS Tsinghua National Laboratory for Information Science and Technology (TNList)," pp. 363–367, 2012.
- [16] A. F. Wicaksono and P. Ayu, "HMM Based Part-of-Speech Tagger for Bahasa Indonesia."

CIELab Color Moments: Alternative Descriptors for LANDSAT Images Classification System

Momen Warna CIELab: Deskriptor Alternatif untuk Sistem Klasifikasi Citra LANDSAT

Retno Kusumaningrum¹, Hisar Maruli Manurung², Aniati Murni Arymurthy²

¹Department of Informatics, Universitas Diponegoro, Jl. Prof. Soedarto, SH, Tembalang, Semarang, 50275

²Faculty of Computer Science, Universitas Indonesia, Kampus UI Depok 16424

Email: retno_ilkom@undip.ac.id, maruli@cs.ui.ac.id, aniati@cs.ui.ac.id

Abstract

This study compares the image classification system based on normalized difference vegetation index (NDVI) and Latent Dirichlet Allocation (LDA) using CIELab color moments as image descriptors. It was implemented for LANDSAT images classification by evaluating the accuracy values of classification systems. The aim of this study is to evaluate whether the CIELab color moments can be used as an alternatif descriptor replacing NDVI when it is implemented using LDA-based classification model. The result shows that the LDA-based image classification system using CIELab color moments provides better performance accuracy than the NDVI-based image classification system, i.e 87.43% and 86.25% for LDA-based and NDVI-based respectively. Therefore, we conclude that the CIELab color moments which are implemented under the LDA-based image classification system can be assigned as alternative image descriptors for the remote sensing image classification systems with the limited data availability, especially when the data only available in true color composite images.

Keywords: Normalized Difference Vegetation Index, CIELab, color moments, Latent Dirichlet Allocation, LANDSAT, remote sensing image classification

Abstrak

Penelitian ini membandingkan 2 jenis sistem klasifikasi citra, yaitu sistem berbasis Normalized Difference Vegetation Index (NDVI) serta sistem berbasis Latent Dirichlet Allocation (LDA) menggunakan momen warna CIELab sebagai deskriptornya. Penelitian tersebut diimplementasikan untuk klasifikasi citra LANDSAT dengan mengevaluasi nilai akurasi dari kedua jenis sistem tersebut. Tujuan dari studi ini adalah untuk mengevaluasi apakah momen warna CIELab dapat digunakan sebagai deskriptor alternatif dari NDVI ketika diterapkan menggunakan model klasifikasi berbasis LDA. Hasil eksperimen menunjukkan bahwa sistem klasifikasi berbasis LDA menggunakan momen warna CIELab memberikan kinerja akurasi yang lebih baik dibandingkan sistem klasifikasi berbasis NDVI, dengan perbandingan nilai akurasi sebesar 87,43% untuk model klasifikasi berbasis LDA dan 86,25% untuk model klasifikasi berbasis NDVI. Oleh karena itu, dapat disimpulkan bahwa momen warna CIELab yang diimplementasikan pada sistem klasifikasi berbasis LDA dapat digunakan sebagai alternatif deskriptor untuk sistem klasifikasi citra penginderaan jauh pada saat data yang tersedia terbatas, khususnya ketika hanya citra *true color composite* yang tersedia.

Kata kunci: Normalized Difference Vegetation Index, CIELab, momen warna, Latent Dirichlet Allocation, LANDSAT, klasifikasi citra penginderaan jauh

1. Introduction

The success value of image classification process is influenced by the suitability of the used features and the domain of images to be classified. In remote sensing images, spectral values are widely used for remote sensing images

classification systems [1], [2], [3], [4], [5], [6], [7]. This is because the spectral values of different materials reflect and absorb differently at different wave lengths. Thus, the objects on the earth surfaces can be easily differentiated by their spectral reflectance signatures.

Vegetation can be easily distinguished from other types of land cover by the high NIR (near infrared) but generally low visible reflectance. One of discriminative features for distinguishing

between vegetation area and non-vegetation area is NDVI (normalized difference vegetation index) which employs near infrared band and red (visible) band [2], [6], [8], [9].

However, the NDVI value can only be extracted when we have multispectral images; in particular, the availability of false color composite images. There are some types of these composite images. A common false color composite image for displaying a LANDSAT multispectral image is as follows:

- R = Band 4 (NIR Band)
- G = Band 3 (Red Band)
- B = Band 2 (Green Band)

Therefore, we propose the use of CIELab color space as an alternative feature of NDVI due to the limitation of data availability. In this study, we compare the accuracy value of an image classification system which employs CIELab color space in LDA-based image classification system and NDVI-based image classification system. The selection of LDA as a classification model since this model provides good performance for classifying land cover in another remote sensing image, IKONOS, compared to the other traditional classification models such as C4.5 and Naïve Bayes tree [10]. These systems are implemented for LANDSAT images.

This paper is organized as follows. In section 2, we will review the LDA-based classification model. Section 3 outlines the NDVI and CIELab color space. The experimental scenario and results is described in section 4. The conclusion of entire study will be explained in section 5.

2. LDA Based Classification Model

This section reviews LDA in general and then describes its implementation for image classification model.

2.1. LDA in General

LDA is a probabilistic latent space model in which each document is represented as a random mixture over a set of latent topics and each topic is represented as a distribution over a predefined and fixed vocabulary [11]. LDA can be seen in two ways, i.e. generative process and inference process. In this study, we implement LDA as inference process since we use this model for image classification process. It means that we implement LDA as inverse process of generative model, i.e. we learn a set of words in the corpus in order to identify the word distribution for each topic (ϕ) and topic proportion for each

document (θ). A corpus consists of D documents where each document consists of N words. In order to overcome the existence of complex coupling latent variables which leads the LDA inference into an intractable process, we can implement the inference process using approximate posterior inference algorithms such as mean field variational methods [11], Gibbs sampling [12], and collapsed variational inference [13].

In this study, we implement Gibbs sampling algorithm. This algorithm updates each of the variables respectively based on the conditional probability of all other variables, although the joint probability distribution is unknown. The goal of the inference process is to compute the probability of topic z given word w for the entire data collection.

$$p(\vec{z}|\vec{w}) = \frac{\prod_{i=1}^W p(w_i, z_i)}{\prod_{i=1}^W \sum_{k=1}^K p(z_i = k, w_i)} \quad (1)$$

Gibb sampling algorithm is used to compute the probability of topic z_i being assigned to word token w_i given all other topic assignments to all other words, full conditional probability $p(z_i = k | \vec{z}_{-i}, \vec{w})$. This probability is used to simulate $p(\vec{z}|\vec{w})$.

$$p(z_i = k | \vec{z}_{-i}, \vec{w}, \vec{\alpha}, \vec{\beta}) = \frac{p(\vec{z}, \vec{w} | \vec{\alpha}, \vec{\beta})}{p(\vec{z}_{-i}, \vec{w}_{-i} | \vec{\alpha}, \vec{\beta})} \quad (2)$$

$$= \frac{n_{k,-i}^{(t)} + \beta_t}{\sum_{t=1}^V n_{k,-i}^{(t)} + \beta_t} \cdot \frac{n_{d,-i}^{(k)} + \alpha_k}{\left[\sum_{k=1}^K n_{d,-i}^{(k)} + \alpha_k \right] - 1} \quad (3)$$

By the definition, we can compute the value of ϕ_k and θ_d as follows

$$\phi_k = p(w = t | z = k) = \frac{n_{t,k} + \beta_t}{\sum_{t=1}^V n_{k,t} + \beta_t} \quad (4)$$

$$\theta_d = p(z = k | d) = \frac{n_{d,k} + \alpha_k}{\sum_{k=1}^K n_{d,k} + \alpha_k} \quad (5)$$

Where $n_{t,k}$ is the number of words w equals to t that are assigned as topic z equals to k , and $n_{d,k}$ is the number of words in document d that are assigned as topic z equals to k .

2.2. LDA for Image Classification

As mentioned before, we implement LDA as inference process for image classification. There are two main processes in LDA-based image classification, i.e. training process and testing process. The aim of training process is to get the estimator for classification process, while the testing process is to classify and segment an image into vegetation and non vegetation area. Training process consists of five steps, i.e.:

2.2.1. Building the training data

Since the final purpose of our task is to classify an image into vegetation and non vegetation area, then we build the training data which consist of 2 set of images. The first set is sample of vegetation images and the second set is sample of non vegetation images. Each set consists of 30 images in the form of 16×16 pixels image patch. These image patches are subsequently called as visual document.

2.2.2. Extracting visual features

In this study, we employ two image features including NDVI and CIELab color moments. These features were extracted over 4×4 pixels image patch which is called as visual words. The detail explanation of those features can be seen in the next section.

2.2.3. Generating visual vocabulary

The visual vocabulary is built using Gaussian mixture clustering from Bouman [14] and integrated visual vocabulary (IVV) as an approach. This approach is built as the union of all centroids obtained from the separate quantization process of each class.

2.2.4. Representing each visual document as Bag of Visual Words (BoVW)

Based on the visual vocabulary from previous step, we use K-nearest neighbor to represent each visual document as BoVW.

2.2.5. Learning data using LDA model

LDA learns the training data, a set of visual words, and the given input of hyperparameters $\alpha = 50/K$ and $\beta = 0.01$, number of topics K , and number of iterations T , in order to get the words distribution for each topic and topic proportions for each document. Subsequently, the topic proportion for each class can be computed as a harmonic mean of topic proportion for each document of all visual documents in the respective class.

In testing process, the same steps as training process are applied including the feature extraction and BoVW representation process. The visual documents were generated using overlapping layout. Based on BoVW for each visual document, we can calculate the value of topic proportion for each visual document. The final task is to use Kullback-Leibler Divergence (KLD) [15] for comparing similarity of distribution between topic proportion each visual document in testing data and the estimator value, that is, topic proportion of each class from training data. The smallest value of KLD shows the most similar distribution. Majority vote was used to fuse the final classification results for each pixel which were obtained for each visual document class.

3. NDVI and CIELab Color Moments

As mentioned before, this study proposes the used of CIELab color moment as an alternative feature of NDVI for distinguishing vegetation and non vegetation area. The following sub section explains the NDVI and CIELab color moment in brief.

3.1. NDVI

NDVI was derived using formula as follows:

$$NDVI = \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}} \quad (6)$$

Where R_{NIR} is reflectance in the near infrared band and R_{RED} is reflectance in the red band. In theory, NDVI value has range between -1.0 and +1.0. Non vegetation areas such as clouds, barren, water, snow and ice give negative NDVI values, whereas positive values of NDVI indicate vegetated area [16].

3.2. NDVI

CIELab is a color space developed by CIE (Commission Internationale de L'Eclairage). This color space is linear to the visual perception and close to human vision [17]. CIELab consists of 3 color coordinates, i.e.:

- L^* coordinates

It represents the lightness which shows grey level when $a^* = 0$ and $b^* = 0$. This coordinate has scale from 0 to 100, where 0 represents black and 100 represents white.

- a^* coordinates

This coordinate represents the red/green sensation and has scale from -127 to +127, where -

127 represents pure green and +127 represents pure red

- b^* coordinates

It represents yellow/blue sensation and also has the scale from -127 to +127, where -127 means pure blue and +127 means pure yellow.

The conversion of this color space requires intermediate conversion, that is, the conversion process of RGB image into XYZ color space. Based on the XYZ color space, the next step is to convert the XYZ color space into CIELab color space. Formally, it can be defined as follows:

$$L^* = \begin{cases} 116 \left(\frac{Y}{Y_n} \right)^{1/3} - 16, & \text{if } t > 0.008856 \\ 903.3 \left(\frac{Y}{Y_n} \right), & \text{if } t \leq 0.008856 \end{cases} \quad (7)$$

$$a^* = 500 \left(f \left(\frac{X}{X_n} \right) - f \left(\frac{Y}{Y_n} \right) \right) \quad (8)$$

$$b^* = 200 \left(f \left(\frac{Y}{Y_n} \right) - f \left(\frac{Z}{Z_n} \right) \right) \quad (9)$$

Where X, Y, Z are pixel values of each channel in XYZ color space. And the other variables are as follows:

$$f(t) = \begin{cases} t^{1/3}, & \text{if } t > 0.008856 \\ 7.787 * t + \frac{4}{29}, & \text{if } t \leq 0.008856 \end{cases} \quad (10)$$

$$\begin{bmatrix} X_n \\ Y_n \\ Z_n \end{bmatrix} = \begin{bmatrix} 0.950456 \\ 1 \\ 1.088754 \end{bmatrix} \quad (11)$$

In this study, we use color moments as image descriptors including mean, standard deviation, and skewness [18]. While the illustration of CIELab color space is described as follow.

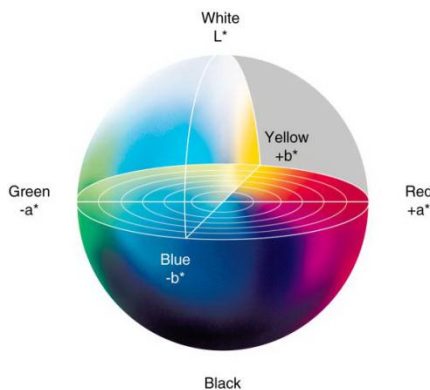


Fig. 1. CIELab Color Space

4. Experiments and Results

4.1. Experimental Setup and Scenario

Multispectral LANDSAT images were used as domain image for image classification. The training data consists of 60 images, including 30 images of vegetation and 30 images of non-vegetation, whereas the testing data consists of an image of 256×256 pixels. The following fig. 2 show the sample of original image.

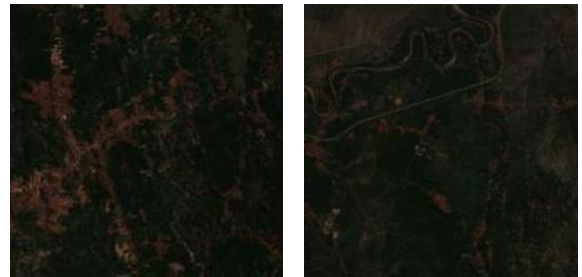


Fig. 2. Original Image and Its Ground Truth Classified Image

The experiment was implemented using Matlab R2010a under Windows 7 – 64 bit and the hardware specifications are as follows:

- Intel Core i5-520M 2.40 GHz
- 8GB of memory (RAM)
- 500GB of hard disk drive

The main task of this study is to compare the accuracy value of LDA-based image classification system using CIELab color moments and NDVI-based image classification system.

4.2. Results and Analysis

The best result of LDA-based image classification system using CIELab color space is obtained for the following variable values, i.e. 5 for the number of topics and 50 for the number of iterations. Figure 3 shows the samples of classified image under the previous setup. Whereas the result of NDVI-based image classification system for the same image is described in fig. 4.

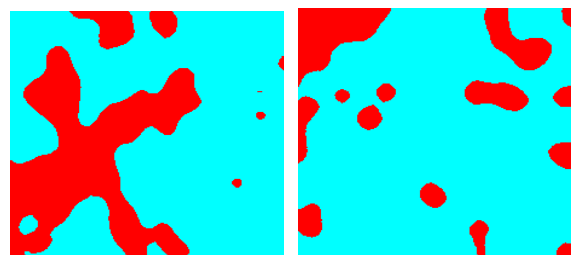


Fig. 3. Result of LDA-based image classification system using CIELab color moments

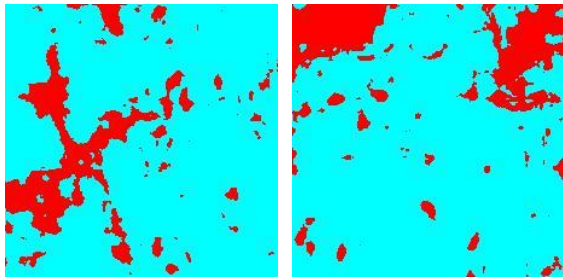


Fig. 4. Result of NDVI-based image classification

Evaluation processes were performed by computing accuracy values for both LDA-based and NDVI-based classification system as shown in table 1. The accuracy values were computed based on ratio between the number of correct classified pixels (pixel with the same label between classified image and ground truth image) and the number of pixels in an image, while overall accuracy value was computed as an average of all testing images accuracy values.

Table 1. Experimental results

No	Overall Accuracy of Image Classification Systems	
	LDA-based and CIELab Color Moments	NDVI-based
1	87.43%	86.25%

Table 1 shows that the LDA-based using CIELab Color Moments gives better result than NDVI-based image classification system. This condition is influenced by the characteristics of CIELab color space which linear to the visual perception and close to human vision.

5. Conclusion

In this study, we proposed the used of CIELab color moment which is implemented under the LDA-based classification system for LANDSAT images as an alternative image descriptor to overcome the limitation of data availability when data only available in true color composite image. The classification systems are implemented to determine the vegetation and non vegetation area. The results show that the used of CIELab color space under the LDA-based image classification system outperforms the traditional NDVI-based classification system by about 1.18%.

References

- [1] J. Li, "Hierarchical Land Cover Information Retrieval in Object-Oriented Remote Sensing Image Databases with Native Queries," in *Proceedings of the 45th Annual Southeast Regional Conference*, 2007, pp. 467–472.
- [2] N. Durand, G. Forestier, P. Ganc, O. Boussa, and A. Puissant, "Ontology-based Object Recognition for Remote Sensing Image Interpretation," in *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence*, 2007, pp. 472–479.
- [3] C. Li, L. Khan, B. Thuraisingham, M. Husain, S. Chen, and F. Qiu, "Geospatial data mining for national security: Land cover classification and semantic grouping," in *Proceedings of IEEE Conference on Intelligence and Security Informatics 2007*, 2007, pp. 254–261.
- [4] T. Liu, K. Muramatsu, D. Motomasa, P. Li, and L. Zhang, "Region-based Image Retrieval Using Semantic Mining," *Doshisha Univ. world wide Bus. Rev.*, vol. 10, no. 2, pp. 97–104, 2009.
- [5] N. Ruan, N. Huang, and W. Hong, "Semantic-based image retrieval in remote sensing archive: An ontology approach," in *IEEE Geoscience and Remote Sensing Symposium*, 2006, pp. 2903–2906.
- [6] T. Liu, L. Zhang, P. Li, and H. Lin, "Remotely sensed image retrieval based on region-level semantic mining," *EURASIP J. Image Video Process.*, vol. 2012, no. 1, p. 4, 2012.
- [7] W. Yi, H. Tang, and Y. Chen, "An Object-Oriented Semantic Clustering Algorithm for High-Resolution Remote Sensing Images Using the Aspect Model," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 522–526, 2011.
- [8] K. C. Tan, H. S. Lim, M. Z. MatJafri, and K. Abdullah, "Landsat data to evaluate urban expansion and determine land use/land cover changes in Penang Island, Malaysia," *Environ. Earth Sci.*, vol. 60, no. 7, pp. 1509–1521, Sep. 2009.
- [9] A. Ghorbani, A. M. Mossivand, and A. E. Ouri, "Utility of the Normalised Difference Vegetation Index (NDVI) for land / canopy cover mapping in Khalkhal County (Iran)," *Ann. Biol. Res.*, vol. 3, no. 12, pp. 5494–5503, 2012.
- [10] R. Kusumaningrum, H. Wei, R. Manurung, and A. Murni, "Integrated visual vocabulary in latent Dirichlet allocation-based scene classification for IKONOS image," *J. Appl. Remote Sens.*, vol. 8, no. 1, pp. 083690–1 s/d 083690–18, Jan. 2014.
- [11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
- [12] T. L. Griffiths and M. Steyvers, "A probabilistic approach to semantic representation," in *Proc. 24th Annual Conference of the Cognitive Science Society*, 2002, pp. 381–386.
- [13] Y. W. Teh, D. Newman, and M. Welling, "A Collapsed Variational Bayesian Inference

- Algorithm for Latent Dirichlet Allocation,” in *Proc. 20th Annual Conference on Neural Information Processing Systems*, 2006, pp. 1353–1360.
- [14] C. A. Bouman, “CLUSTER: An Unsupervised Algorithm for Modeling Gaussian Mixtures Developed by:,” 2005. [Online]. Available: <https://engineering.purdue.edu/~bouman/software/cluster/manual.pdf>. [Accessed: 22-Feb-2012].
- [15] G. Heinrich, “Parameter estimation for text analysis,” 2009.
- [16] F. Yuan and M. E. Bauer, “Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery,” *Remote Sens. Environ.*, vol. 106, no. 3, pp. 375–386, Feb. 2007.
- [17] A. Ford and A. Roberts, “Colour Space Conversions,” 1998. [Online]. Available: <http://www.poynton.com/PDFs/coloureq.pdf>. [Accessed: 22-Jun-2011].
- [18] M. Stricker and M. Orengo, “Similarity of Color Images,” in *Proc. SPIE 2420, Storage and Retrieval for Image and Video Databases III*, 1995, pp. 381–392.

Indeks Penulis

INKOM Volume 8, 2014

- Ade Ramdan, 61
 Pendeteksian Gerakan Menggunakan Transduser Ultrasound dengan Metoda Pembandingan, 8(2), 61
- Adi Nurhadiyatna
 Parallel Algorithms for Spatial Rainfall Distribution, 8(1), 29, 61
- Adi Nurhadiyatna, 29, 61
- Agus Subekti
 Metode Adaptive Beamforming untuk Cognitive Radio, 8(1), 53, 61
- Agus Subekti, 53, 61
- Andon Insani
 Pengembangan Perangkat Lunak Akuisisi Data dan Kontrol Difraktometer Serbuk Resolusi Tinggi, 8(1), 45, 61
- Andon Insani, 45, 61
- Andriyan Bayu Suksmono
 Peningkatan Kinerja Skema Estimasi Arah Kedatangan Sinyal dengan Compressive Sensing Sparsitas Sudut dan Sampel Multisnap, 8(1), 21, 61
- Andriyan Bayu Suksmono, 8(1), 21, 61
- Arnida L Latifah
 Parallel Algorithms for Spatial Rainfall Distribution, 8(1), 29, 61
- Arnida L Latifah, 29, 61
- Ayu Purwarianti, 69, 103
 Model dan Metoda Arsitektur pada Sistem Tanya Jawab Medis, 8(2), 69
 Pengenalan Entitas User Profile Pada Twitter, 8(2), 103
- Bharoto
 Pengembangan Perangkat Lunak Akuisisi Data dan Kontrol Difraktometer Serbuk Resolusi Tinggi, 8(1), 45, 61
- Bharoto, 45, 61
- Dahmir Dahlan, 83
 Model Pengukuran Proses Bisnis CRM Berbasis eTom dan IT-IL, 8(2), 83
- Dicky Rianto Prajitno, 61
 Pendeteksian Gerakan Menggunakan Transduser Ultrasound dengan Metoda Pembandingan, 8(2), 61
- Edi Kurniawan, 95
 Perbandingan Performa Pelacakan antara Repetitive Controller dan PI Controller, 8(2), 95
- Gendut Suprayitno, 83
 Model Pengukuran Proses Bisnis CRM Berbasis eTom dan IT-IL, 8(2), 83
- Hendra Gunawan
 Peningkatan Kinerja Skema Estimasi Arah Kedatangan Sinyal dengan Compressive Sensing Sparsitas Sudut dan Sampel Multisnap, 8(1), 21, 61
- Hendra Gunawan, 21, 61
- Iping Supriana, 69, 103
 Model dan Metoda Arsitektur pada Sistem Tanya Jawab Medis, 8(2), 69
 Pengenalan Entitas User Profile Pada Twitter, 8(2), 103
- Koredianto Usman
 Peningkatan Kinerja Skema Estimasi Arah Kedatangan Sinyal dengan Compressive Sensing Sparsitas Sudut dan Sampel Multisnap, 21, 61
- Koredianto Usman, 21, 61
- Lihardo Ranjaliba Saragih, 83
 Model Pengukuran Proses Bisnis CRM Berbasis eTom dan IT-IL, 8(2), 83
- Maria Ariesta, 61
 Web-Based Decision Support Systems Application of Stock Recommendation Using Bayesian Methods, 8(1), 1, 61
- Nina Sevani
 Web-Based Decision Support Systems Application of Stock Recommendation Using Bayesian Methods, 8(1), 1, 61
- Nina Sevani, 1, 61
- Oka Mahendra
 Evaluasi Modulasi MFSK untuk Transmisi Data Melalui Kanal Suara GSM, 8(1), 37, 61

Oka Mahendra, 37, 61

Rika Sustika

Evaluasi Modulasi MFSK untuk Transmisi Data
Melalui Kanal Suara GSM, 8(1), 37, 61

Rika Sustika, 37, 61

Suyoto

Studi Perbandingan Teknik Sinkronisasi Waktu
untuk Sistem Mobile OFDM, 8(1), 11, 61

Suyoto, 11, 61

Titin Pramiyati, 103

Pengenalan Entitas User Profile Pada Twitter,
8(2), 103

Wiwin Suwarningsih, 69

Model dan Metoda Arsitektur pada Sistem
Tanya Jawab Medis, 8(2), 69

Pedoman Penulisan Naskah

1. Ruang Lingkup

Jurnal INKOM menerima naskah yang berisi hasil penelitian, pengembangan, dan/atau pemikiran di bidang Informatika, Sistem Kendali, dan Komputer. Naskah harus orisinal dan belum pernah dipublikasikan serta tidak sedang dalam proses publikasi di jurnal/media lain. Setiap naskah yang diterima akan dievaluasi substansinya oleh paling sedikit 2 orang pakar mitra bestari (peer reviewer) sebagai juri dalam bidang yang sesuai. Untuk menjunjung fairness proses penilaian dilakukan hanya pada isi naskah dengan menghilangkan identitas penulis (blind review). Penulis/para penulis bertanggung jawab sepenuhnya terhadap akurasi naskah. Penulis utama bertanggung jawab untuk sebelumnya menyelesaikan ijin penulisan yang berkaitan dengan hasil kerja anggota kelompoknya. Naskah yang diterima dianggap sudah menyelesaikan seluruh kewajiban (clearance) dan ijin reproduksi bila memuat hal-hal yang mengandung hak cipta (copyright) pihak lain.

2. Standar Umum Penulisan

- a. Naskah ditulis dalam bahasa Indonesia atau Bahasa Inggris.
- b. Judul, Abstrak, dan Kata kunci harus ditulis dalam dua bahasa (Indonesia dan Inggris)
- c. Ditulis menggunakan *word processor* (Microsoft Word, Open Office, atau Latex). Naskah diketik dalam 2 kolom (ukuran kertas A4) dengan huruf Times New Roman ukuran 11, rata kanan-kiri. Panjang naskah sekurang - kurangnya 6 halaman, dan tidak lebih dari 10 halaman, tidak termasuk lampiran.
- d. Naskah diawali dengan judul, nama penulis, instansi, alamat surat, dan alamat email untuk korespondensi.
- e. Materi yang akan dicetak, meliputi teks, gambar ilustrasi, dan grafik harus berada dalam area pencetakan yaitu bidang kertas A4 (297mm x 210mm), dengan margin 2cm di semua sisi kertas. Format yang dianjurkan adalah dalam format \LaTeX karena redaksi hanya mengedit makalah dalam format \LaTeX . Namun, redaksi masih dapat menerima format yang lain seperti *word* atau *odt* sesuai dengan template yang redaksi telah sediakan. Jangan menuliskan atau meletakkan sesuatu diluar bidang cetak tersebut. Seluruh teks ditulis dalam format dua kolom dengan jarak antar kolom 1 cm, kecuali bagian abstrak yang dituliskan dalam format satu kolom. Seluruh teks harus rata kiri-kanan. Template ini menggunakan format yang dianjurkan. Untuk mempermudah penulis dalam memformat makalahnya, format ini dapat digunakan sebagai petunjuk atau format dasar penulisan.
- f. Isi naskah setidaknya-tidaknya berisi/menerangkan tentang pendahuluan, metoda, hasil, diskusi, kesimpulan, daftar pustaka. Ucapan terimakasih bila diperlukan dapat dituliskan setelah bagian kesimpulan. Sistematika penulisan mengacu pada Peraturan Kepala LIPI Nomor 04/E/2012 tentang pedoman karya tulis ilmiah.

3. Cara Penulisan Judul

Judul utama (pada halaman pertama) harus dituliskan dengan jarak margin 2cm dari tepi kertas, rata tengah dan dalam huruf Times 16-point, tebal, dengan huruf kapital pada huruf pertama dari kata benda, kata ganti benda, kata kerja, kata sifat, dan kata keterangan; jangan menggunakan huruf kapital pada kata sandang, kata hubung, terkecuali jika judul dimulai dengan kata-kata tersebut. Sisakan satu 11-point baris kosong sesudah judul.

4. Cara Penulisan Nama dan Afiliasi

Nama penulis dan afiliasi diletakkan ditengah dibawah judul. Nama penulis dituliskan dengan huruf Times 12-point, tidak tebal. Afiliasi dan email penulis dituliskan dibawahnya dengan huruf Times 10-point, miring. Penulis yang lebih dari satu orang dituliskan dengan menggunakan *superscript* angka yang merujuk pada masing-masing afiliasi. Sedangkan email cukup dituliskan korespondensi email saja, misal email dari penulis pertama saja.

5. Cara Penulisan Abstrak dan Kata Kunci

Abstrak dalam bahasa Indonesia ditulis dengan rata kiri-kanan dengan inden 0.5cm, sesudah abstrak dalam bahasa Inggris, dengan satu spasi dan satu kolom. Kata Abstrak sebagai judul ditulis dalam huruf Times 11-point, tebal, rata tengah, dengan huruf pertama dikapitalkan. Teks abstrak ditulis dengan huruf Times 10-point, satu spasi, sampai lebih kurang 150 kata. Sesudah abstrak bahasa Indonesia tuliskan kata kunci dari makalah tersebut dalam daftar kata kunci. Kemudian dilanjutkan dengan teks utama makalah.

6. Cara Penulisan Bab (*Heading*)

1. Judul pertama

Sebagai contoh, 1. Pendahuluan, dituliskan dalam huruf Times 11-point, tebal, huruf pertama kata pertama ditulis dengan huruf kapital. Gunakan tanda titik (.) sesudah nomor judul.

1.1. Judul kedua

Sebagaimana judul pertama, judul kedua dituliskan dengan huruf Times 11-point, tebal. Nomor judul terdiri dari dua angka yang dibatasi dengan tanda titik.

1.1.1. Judul ketiga

Untuk uraian yang lebih panjang dan tidak dapat dituliskan dalam bentuk uraian terurut, digunakan judul ketiga. Judul ketiga menggunakan ukuran huruf yang sama yaitu huruf Times 11-point, tetapi miring. Nomor judul terdiri dari tiga angka yang dibatasi dengan tanda titik. Tidak dianjurkan menggunakan judul hingga tiga tingkatan, sebaiknya hingga Judul kedua saja.

7. Cara Penulisan Text Utama

Ketik teks utama dengan menggunakan huruf Times 11-point, satu spasi. Jangan menggunakan dua spasi. Pastikan teks ditulis dengan rata kiri-kanan. Jangan menambahkan baris kosong di antara paragraf. Istilah dalam bahasa asing (*foreign language*) yang tidak dapat diterjemahkan dalam bahasa utama makalah harus dituliskan dalam huruf miring.

Terdapat dua jenis uraian yaitu: enumerasi dan itemisasi. Untuk enumerasi gunakan digunakan huruf alfabet kecil dengan titik, sebagai contoh:

- a. Uraian yang memiliki aturan pengurutan
- b. Uraian yang terkait dengan uraian lainnya
- c. Uraian yang setiap itemnya akan diacu pada tulisan utama

Sedangkan itemisasi dituliskan dengan *bullet* adalah:

- Uraian yang tidak memiliki aturan pengurutan
- Uraian yang tidak terkait dengan uraian lainnya

8. Cara Penyajian Tabel

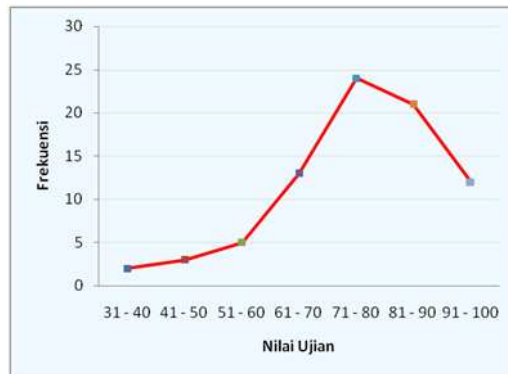
Penyajian tabel harus berada dalam lingkup ukuran A4. Keterangan tabel dituliskan dengan huruf Times 10-point. Keterangan tabel diletakkan sebelum tabel dengan rata kiri. Tabel dibuat tanpa menggunakan garis vertikal. Tabel harus diacu dalam tulisan seperti Tabel 1.

Table 1: Contoh Tabel 1

Header 1	Header 2	Header 4
1	Teks	50%
2	Teks	20%
3	Teks	60%

9. Cara Penyajian Gambar

Penyajian gambar harus berada dalam lingkup ukuran A4. Keterangan gambar dituliskan dengan huruf Times 10-point. Sedangkan pengacuan gambar pada teks menggunakan huruf Times 11-point sesuai dengan teks utama.



Gambar 1: Contoh Gambar

Keterangan gambar diletakkan di bawah, tengah gambar yang dijelaskan. Gambar diletakkan di tengah satu kolom. Jika tidak memungkinkan atau gambar terlalu lebar gambar bisa diletakkan di tengah dalam format dua kolom. Gambar harus diacu dalam tulisan seperti Gambar 1.

10. Cara Penulisan Persamaan (*equation*)

Penulisan formula/persamaan/rumus matematika dapat menggunakan *microsoft equation* apabila penulis menggunakan *Microsoft Word*. Sedangkan apabila penulis menggunakan latex, maka penulis dapat menggunakan penulisan formula standar dalam latex dengan menggunakan paket *amsmath*. Label persamaan ditulis dibagian kanan persamaan menggunakan huruf arabic didalam kurung. Berikut ini adalah contoh penulisan persamaan matematika:

$$G(x, y) = \exp\left(-\frac{x^2 + \gamma^2 y^2}{2\sigma^2}\right) \sin\left(i2\pi \frac{x}{\lambda} + \psi\right) \quad (1)$$

Penulis dapat menggunakan kata "persamaan (1)" apabila akan mengacu pada rumus/formula/persamaan yang memiliki label (1). Label persamaan ditulis berurutan sesuai dengan posisi kemunculan dalam halaman. Berikut ini adalah contoh bagaimana penulis mengacu sebuah persamaan:

"Formula (1) merupakan rumusan Gabor Filter untuk bagian imajiner ..."

11. Cara Penulisan Ucapan Terimakasih

Berikut ini adalah contoh penulisan ucapan terimakasih dalam naskah: Ucapan terima kasih penulis sampaikan kepada Pusat Penelitian Informatika Lembaga Ilmu Pengetahuan Indonesia atas dukungan dana penelitian melalui Daftar Isian Pelaksanaan Anggaran (DIPA) 2012.

12. Cara Penulisan Kutipan dan Daftar Pustaka

Daftar pustaka memuat daftar bacaan yang diacu dalam tulisan utama. Daftar pustaka ditulis dengan metode penulisan kepastakaan IEEE transaction, dengan huruf Times 10-point. Kutipan dalam teks utama yang mengacu kepada daftar pustaka dituliskan dengan angka dalam kurung siku [1]. Jika acuan lebih dari satu, pengacuan ditulis seperti ini ([2, 3]). Daftar rujukan yang dikutip dituliskan pada bagian akhir naskah dengan judul Daftar Pustaka dan diberikan nomor urut sesuai dengan urutan pengutipan pada naskah. Bagian naskah yang mengacu pada satu atau beberapa literatur lain hendaknya mencantumkan nomor urut referensi pada daftar pustaka. Pengacuan acuan pada naskah dengan menggunakan notasi [nomor acuan] seperti: [1] (artikel pada jurnal), [2] (artikel pada prosiding) dan [3] (buku). Berikut ini adalah contoh daftar pustaka:

Daftar Pustaka

- [1] D. Rosiyadi, S.-J. Horng, P. Fan, X. Wang, M. Khan, and Y. Pan, "Copyright protection for e-government document images," *MultiMedia, IEEE*, vol. 19, no. 3, pp. 62–73, 2012.
- [2] A. F. M. Hani, E. Prakasa, H. Nugroho, A. Affandi, and S. Hussein, "Body surface area measurement and soft clustering for pasi area assessment," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 4398–4401.
- [3] D. E. Knuth, *The T_EXbook*. Addison-Wesley, 1984.

13. Template Penulisan Naskah

Template tata penulisan naskah dapat didownload di

<http://jurnal.informatika.lipi.go.id/index.php/inkom/about/submissions#authorGuidelines>

Jurnal INKOM

Pusat Penelitian Informatika
Lembaga Ilmu Pengetahuan Indonesia
Komp. LIPI Gd. 20 Lt. 3
Sangkuriang, Bandung, 40135
Email: jurnal@informatika.lipi.go.id
Telp: +62 22 2504711
Fax: +62 22 2504712
<http://jurnal.informatika.lipi.go.id>

