International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015)pp34217-34223

Discovery Indonesian Medical Question-Answering Pairs Pattern With Question Generation

Wiwin Suwarningsih[1,2], Iping Supriana[3], Ayu Purwarianti[4]

[1,3,4]School of Electronic Engineering and Informatics, Bandung Institute of Technology
[2]Research Center for Informatics, Indonesian Institute of Science
[1,2]wiwin.suwarningsih@students.itb.ac.id, [3]iping@informatika.org, [4]ayu@informatika.org

Abstract
Question generating refers to the task of automatically generating questions from various inputs such as raw text, database, or semantic representation. This paperconcerns with the developmentof the novel question generating methods for Indonesian medical question answering. It begins by presenting an overview of the main methods and followed by showing the architecture of the proposed system in detail. Furthermore, this paper proposed a novel template-based approach to question a generation by combining a number of semantic roles through a method of generating both general and domain-specific questions. It was more focused on Question Generating with a discourse task involving the following three steps: (1) content selection of sentence (Problem, Intervention, Comparison and Outcome (PICO) categories), (2) question type identification, and (3) question answering construction. It is argued then that many ways in which the sentences of the answer to a question can be formulated and acquired from resources containing a large number of semantically related sentences.This approach achievedF-measure values of 0.88, precision 0,84 and recall 0,86 for active sentence.Finally, it can be concluded that the pattern matching criteria of the training set and semantic role labelling based on PICO frame can be reproducible with minimal expert intervention.

Keywords : medical question generating, question-answering pairs, semantic transformation, PICO frame.

Introduction

Question generating(QG) refers to an important component in dialogue systems, virtual environments, and learning technologies such as Intelligent Tutoring Systems, inquiry-based environments, and instructional games[7][9].Question Generating system can be helpful in Closed-domain Question Answering (QA) systems such as medical QA, clinical QA and biomedical QA[6][9]. Some closed-domain QA systems use a number of predefined (sometimes hand-writing) question-answer pairs to provide QA services. Employing a QG approach to such systems could be ported to other domains with little or even no effort[6].Somepreviousresearched on QG such as Lindberg [7] have ever been conducted by introducing a novel template-based approach to automatically generates natural language questions to support online learning. Heilman and Noah [2], addressed the challenge of automatically generating questions from reading materials for educational practice and assessment.

Another research created a framework for question generationin composing a number of general-purpose rules to transform some declarative sentences into questions. It was modular in that the existing NLP tools could be leveraged, and included a statistical component for scoring questions based on features of the input, output, and transformations performed[3]. Yao [8], proposed a novel approach based entirely on semantics for QG with Minimal Recursion Semantics (MRS), a meta-level semantic representation withan emphasis on a scope underspecification. Through the English Resource Grammar and varioustools from the DELPH-IN community, a natural language sentence can be interpretedas an MRS structure by parsing in which this,structure can be realized as a naturallanguage sentence through generation.

Based on these studies, we in this study aimed to create a system for shallow question generation (QG) that can take as an input of a medical article of text, focused on QG with a discourse task involving the following three steps: (1)content selectionof sentence (Problem, Intervention, Comparison and Outcome (PICO) categories), (2) question type identification, and (3) question answeringconstruction.

Overall, this paper is aimed to provide following contributions:

1) Converting Indonesian natural language to frame PICO for content selection and question type identification Indonesian medical sentence.
2) Mappingsemantic role labelling using PICOframeto develop semanticrules and provide a good set of question answering (QA) pair patterns.
3) As an application, we show thatF-measure values of 0.88, precision 0.84 and recall 0.86 for active sentence.

The rest of the paper is organized as follows: Section 2 describes related work on Question generation, followed by Section 3 presenting the details of the proposed method. Furthermore, Section 4 shows the evaluation and results and Section 5 describes the experiment. Finally, the conclusion and future directions are presented in Section 6.

Related Work

In common, the following issues have been addressed in question generation such as Yao *et al.*[1]thatpresented a question generation system based on the semantic rewriting approach. They here obtained a principle way of generating questions that avoided the ad-hoc manipulation of syntactic structures. In addition, they were able to use an independently developed parser and generator for the analysis and generation stage. The generator typically proposesseveral different surface realizations of a given input in view of its extensive grammatical coverage.Heilman and Smith [2], meanwhile, focused on question generation (QG) for the creation of educational materials for reading practice and assessment. Their objective was to generate some fact-based questions about the content of a given article. Lindberg *et al.*[4], introduced a sophisticatedtemplate-based approach that incorporatedsome semantic role labels into a system that automatically generated natural languagequestions to support online learning.The task of question generation (QG) from text can be broadly divided into three (not entirely disjoint)categories: syntax-based, semantics-based, and template-based. The systems in the syntactic category frequently use the elements of semantics and vice-versa.

Another related work concerned with the automatic generation of factual WH questions[5]. The system created an automated system that could take an input of a text and produced the output questions for assessing a reader's knowledge of the information in the text. The questions could then be presented to a teacherwho could select and revise the ones judged to be useful. Husam*et al.*[6]considered an automatic Sentence-to-Question generation task in which a sentence given, the Question Generation (QG) system generated a set ofquestions for which the sentence contains, implies, or needsanswers. To facilitate the question generation task, they built some elementary sentences from the input complexsentences using a syntactic parser. A named entity recognizer and a part of speech tagger have been applied oneach of these sentences to encode any important information. They then classified the sentences based on their subject,verb, object and preposition to determine the possible type of questions to be generated.Lindberg[7] introduced a novel template-based approach to question generation that combined some semantic roles with a method of generating both general and domain-specific questions. They evaluated their approach that was mindful of the context in which the generated questions were goingto be used. This evaluation then showedthe approach to be effective in generating pedagogically-useful questions. Bednaric and kovacs[9] developed an automated question generation with a more flexible and open frameworkfor the Hungarian language. The focus of their work was on the investigation how effective the application of generalsoft computing stochastic methods in the processing of semantic and grammatical problem domains was.

Proposed Method

The idea of the discovery of QA-pairs pattern method is to automatically transformation from a declarative sentence to interrogative ones. The method used for the manufacture of transformation rules in this paperwas the result of utilizing the output sentence dependencies between NE. It is assumed that the sentences with a dependency relationship are going to share the semantic role raised in the same or related sentences. It has been motivated by a number of researches in which the lexical semantic hypotheses were on the word behaviour, especially regarding the expression and interpretation of rules to a large extent determined by the meaning. For each slot in the pattern of semantic templates, it was performed by examining the role of semantic and adjusting it with the transformation rules. The role of the text was in accordance with the transformation rules, altered according to slots extracted and put into words the question. The flowchart of this study is illustrated in Figure.1.

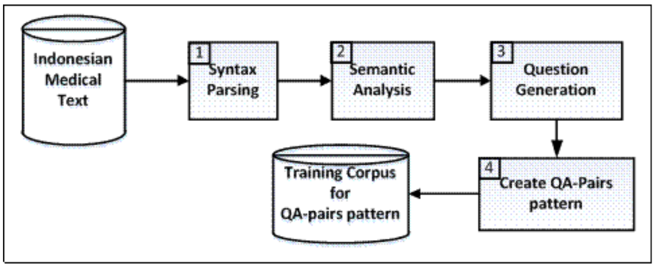

Figure.1. Flowchart of QA pairs pattern discoveryfor Indonesian medical question answering system

Evaluation and Results

QG from text is of a vital importance for self-directed learning[10]. It has been shown [11][12] that questioning is an effective method of helping learners to learn better.Many previous works have examined QG from single sentences, butthistechnique has almost exclusively more focused on generating factoid questions. Factoids are questions requiring the learner to recall some facts explicitly as stated in the source text[12].

Content Selection of Sentence

Selecting the content of sentence was performed through three stages: firstly parsing the syntax to obtain Parts of Speech (POS) tagging to determine whether each word as a noun, verb, or preposition. Here, the list of POS for medical domain used the results of research as conducted by Alfan [13]. Secondly, it was through syntactic parsing to identify the type of phrase contained in the sentence; and thirdly, through semantic analysis to process sentences tree that has been affixed to the semantic rules to generate representations of meaning in the form of semantic rules. Semantic rules used in this paper were based on the frame PICO (see Table 1.).

For the purposes of selecting the sentence content in this paper, the conversion of natural language to frame PICO was conducted. For this, a shape-based classification PICO frames were developed by adjusting the needs of the medical domain Indonesia. The development of PICO frames into PPPICCOODTQ (Problem, Population, Patient, Intervention, Compare, Control, Outcome, Organs, Drug, Time and Quantity). The backgrounds why we did the conversion of natural language medical domain Indonesia to frame the PICO were as follows:(i) Frame PICO is merely an organizing structure conducive to the type of clinical or medical questions [14][15][16]; (ii) Frame PICO is an essential element for use in model-based semantic patterns in medical questions [17]; (iii) Frame PICO can be used to obtainpotential answers derived from the identification of named entities and correspondencebetween semantic roles [18][19][20]; and (iv) Identification of candidate answers are more flexible by using of the identification of the frame-based semantic role PICO [21][22][23][24].

Table1. PICOframe-basedsemanticrules

Semantics Rule	Specification
λxλy population (x, y)	y is a kind of disease of x
λxλy patient (x, y)	y is the patient of x
λxλy problem (x, y)	y is the problem of x
λxλy intervention (x, y)	y is the interference of x
λxλy compare (x, y)	y is a comparison of x
λxλy control (x, y)	y is the control of x
λxλy outcome (x, y)	y is the result of x
λxλy organs (x, y)	y is this part of the body x
λxλy drug (x, y)	y is a drug that is used on the x
λxλy time (x, y)	y is the time required from x
λxλy quantity (x,y)	y is a measure of x

For more details, we have attempted to provide a picture of the process of parsing and semantic analysis using the sentence (1).

Peningkatan sistem imun menahanserangan virus(1)
(English : Improved immune system resists virus attacks)

The stages of decomposition for sentence (1) included:

- POS formed is an
  peningkatan/N
  sistem/N
  imun/JJ
  menahan/VB
  serangan/N
  virus/N

- Identification of the phrase: *sistem imun*is an adjective phrase or JJP and *serangan virus* are noun phrases or NP. Thus, labelling sentence examples-1 became:
  Peningkatan/N
  [*sistem*/N *imun*/JJ]/JJP
  menahan/VB
  [*serangan*/N *virus*/N]/NP

- Establishment of a parse tree with frame-based semantic analysis labelling PICO for sentence (1) (see Figure 1).The use of lexical semantic embedding rules is listed in Table 2.

Table 2. Rules are attaching for Lexical Semantics with PICO frame-based

No	POS label	lexical semantic
1.	N, JJ	λx object(x,A)
2.	VB	λxλy intervention(x,y)
3.	NP	λx object(x, A) ^ λxλycompare(x,y) |
		λx object(x, A) ^ λxλy control(x,y) |
		λx object(x, A) ^ λxλy outcome(x,y) |
		λxλycompare(x,y) |
		λxλy control(x,y) |
		λxλy outcome(x,y) |
		λxλy organs(x,y) |
		λxλy drug(x,y) | λxλy problem(x,y)
4.	JJP	λxλy comparison(x,y)
5.	VBI	λxλy intervention(x,y) ^ λxλy problem(x,y)
		λxλy intervention(x,y) ^ λxλy population(x,y)
6.	VBT	λxλy intervention(x,y) ^ λxλy problem(x,y)^ λxλy patient(x,y)
		λxλy intervention(x,y) ^ λxλy population(x,y)^ λxλy patient(x,y)

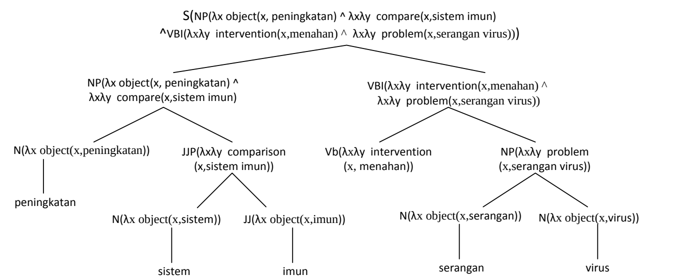

Figure.2. Parse tree, with PICO frame-based semantic labeling

The establishment of a parse tree with semantic labelling was required to define the phrase or keywords contained in the sentence (see Figure 2). Furthermore, keyword phrase was used for the extraction of information (including the manufacture of extraction rules and relations between phrases or keywords) in the sentence.
The process of information extraction using the rules basedon PICOframe and NE relations contained is shownin Figure3. Furthermore, the result of the extraction of information for the sentence in example 1 can be seen in Figure 4.

Rule Base on PICO frame:
R1 : (*gejala, tanda-tanda*)[POPULATION] *adalah* [PROBLEM], [PROBLEM], [PROBLEM] .
R2 : [INTERVENTION] [PROBLEM] *akibat* [POPULATION] *gunakan* [DRUG]
R3 : [PATIENT] [POPULATION] *harus* [INTERVENTION] [PROBLEM] *dan* [PROBLEM]
R4 : *Cara* [INTERVENTION] [POPULATION] *adalah* [CONTROL], [CONTROL] dan [CONTROL]

NE Relations :
Rel1 : [POPULATION] *DISEBABKAN-OLEH*[PROBLEM]
Rel2: [PROBLEM] *INDIKASI*[POPULATION]
Rel3: [OUTCOME]*INDIKASI*[POPULATION]
Rel4: [CONTROL]*MENGATASI*[POPULATION]

Figure.3. Rule base on PICO frame and NE relations.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015)pp34217-34223

a. Identification of the sentence for NE:
sistem imun= <COMPARE>
menahan = <INTERVENTION>
serangan virus=<PROBLEM>
b. Put in the rule base extraction. The extraction rule its forms:
Peningkatan[COMPARE] [INTERVENTION] [PROBLEM]
c. Information extraction by establishing relationships between NE
- [COMPARE] ***MENCEGAH*** [PROBLEM]
- *Peningkatan*[COMPARE] ***MENCEGAH*** [PROBLEM]
- [PROBLEM] ***DIATASI-DENGAN****peningkatan* [COMPARE]

Figure.4. InformationExtractionfor Sentence

"*Peningkatan sistem imun menahanserangan virus*"
The extraction of this information was used to identify a role owned by the sentence (1). Furthermore, the role was analysed by the system and classified for the question generation.

Question Type Identification
Question type identification was conducted using the identification of semantic roles held by declarative sentences. The semantics, in this case, acted as a key in the transformation process. The performance of this transformation was based upon transformation rules and the method used for the manufacture of transformation rules in this paper was the result of utilizing the output sentence dependencies between NE. It was then assumed that the sentences with a dependency relationship would share the semantic role raised in the same or related sentences. It was based upon a number of researches in which the lexical semantic hypotheses were on the word behaviour, especially regarding the expression and interpretation of rules to a large extent determined by the meaning.
For each slot in the pattern of semantic templates, the semantic role was examined and adjusted by the transformation rules. The role of the text was in accordance with the rules of transformation and altered based on the slots extracted and put into question. Based on the formed transformation rules, the sentence (1) then could be identified using semantic role (i.e.<COMPARE> = *sistem imun*, <INTERVENTION> = *menahan*, and <PROBLEM> = *serangan virus*) and transformed by the rule:

- **If** System finds role = <COMPARE>
  then question_sentence =***Apa*** +*manfaat*+<COMPARE>?(**What**+are benefits+<COMPARE>?)
- **If** System finds role = <PROBLEM>
  then question_sentence =***Bagaimana***+<INTERVENTION>+<PROBLEM>? (**How**+ <INTERVENTION> + <PROBLEM> ?)
- **If** System finds role = <PROBLEM>
  then question_sentence=***Apa***+*penyebab*+<PROBLEM> ? (**What** + are the causes+ <PROBLEM> ?)

The results of generating questions for sentence (1) can be seen in sentence (2) through (4).

Apa*manfaatsistem imun*?
(Whatare the benefits of immune system?) (2)

Bagaimana *menahan serangan virus?*
(How holds viruses?) (3)

Apa*penyebabserangan virus*?
(What are the causes of virus attacks?) (4)

Sentence (2) through (4) has not produced the answer, for being limited to the transformation of declarative sentences into interrogative ones. To obtain the necessary questions to the pattern matching, templates predicate argument (PA) already have a couple questions answers.

QuestionAnswering Construction.
At this stage, the process of making a template of PAwas used in the filtering process [25]. Based on the results of the manual analysis of sentences used as a source of training, templates PA would be defined (see Table 3) with a predicate <INTERVENTION>, while the other was as an argument.
The use of <INTERVENTION>aspredicatewas based on the following reasons:
(i) a unit of an event or an action and captures was more semantics than keywords, thus more promising that answers could be more precisely extracted[26];
(ii) it described the relations between nouns and used to handle expressions not mediated by verbs[27].

Table 3. PA template for question answering pair

Question Template	Answering Template
Apa+*manfaat* +<COMPARE>?	<INTERVENSION> + <PROBLEM>
Apa+*manfaat* +<CONTROL>?	<INTERVENSION> + <PROBLEM>
Apa+*manfaat* +<CONTROL>?	<INTERVENSION> + <POPULATION>
Apa+*manfaat* +<CONTROL>?	<INTERVENSION> + <OUTCOME>
Apa+*indikasi*+<OUTCOME >?	<PROBLEM>, <PROBLEM>
Apa+*penyebab*+<PROBLE M>?	<POPULATION>
Apa+*penyebab*+<POPULAT ION>?	<PROBLEM>, <PROBLEM>,<PROBLEM>
Bagaimana+<INTERVENS ION>+<PROBLEM>?	<COMPARE>
Bagaimana+<INTERVENS ION>+<PROBLEM>?	<CONTROL>
Bagaimana+<INTERVENS ION>+<POPULATION?	<COMPARE>
Bagaimana+<INTERVENS ION>+<POPULATION?	<CONTROL>
Bagaimana+<PROBLEM>+ *dapat dihindari*?	*Lakukan*<CONTROL>,<CO NTROL>*dan*CONTROL>

The generation of sentence (1) using the transformation rules of the sentence was not to produce an answer. As a consequence, it was deemed important to have a way to bring the answer. In this research, it was conducted by making the process of comparison against the template PA which has been formed as shownin Table 3. Sentence (2) through (4) was processed using pattern matching. If the pattern was found and fit, the pattern of responses would be displayed. If the pattern was not found, the process would be terminated then.

Phase filtering functions to sort a number of interrogative sentences taken from the transformation stage sentence. The sorting process was performed using a template pattern matching in PA with the sentence pattern transformation results. The algorithms used for pattern matching were modified from Deterministic Finite Automata (DFA) method as created by Knuth *et al*. [28]. Figure 5 presents the concept of the algorithm for pattern matching.

```
Algorithm for pattern matching & found of answering
Input: Sentence question PICO pattern
Output: The pattern of the sentence answers found
Method: Each state of each pattern PICO and each node has multiple
      labels (such as POS tagging, NER)

// Create a pointer to the next state
Int j = 0
Int p = 0

// identification for PICO in the database table = d.pico
// identification to query PICO of sentence = q.pico

// Matching process
For (int i = 0; i <N; i ++) // N is the length of the pattern
      If (q.pos (i) = d.pos (j) then // compare position PICO with
                                     the pattern of PICO
            Next [j] = next i];
            j ++; // pattern match then copy and increment
      Else
            Next [j] = x + 1;
      x = next [x]; // do the opposite pattern does not match
Endif

// If the pattern corresponding questions and found, show patterns of
response
If (j = p) then
      print "Answering pattern";
Endif
Endfor
```

Figure 5. Algorithm for pattern matching & found of answering

The workings of this algorithm are to compare two databases that contain a pattern PICO and PICO query from an input sentence. The pattern is read one by one and compared. If the pattern is found equal, it is recorded and does to the next string up with unreadable input pattern and reaches that mark a question (?). Conversely, if the pattern does not match the pattern, it is ignored then.

Based on the concept of the pattern matching algorithm, each sentence (2) to (4) has an answer as in sentence (5) through (7).

Question :*Apamanfaatsistem imun*?
(What are the benefits of the immune system?) (2)
Pattern Answer: <INTERVENTION> + <PROBLEM>

Answer:*menahan serangan virus*
(to resist virus attacks) (5)

Sentence (3) produced two answers, those are<COMPARE> or <CONTROL>. However, asthe source of the sentence (1) found no role <CONTROL>, then for role <CONTROL>was ignored.

Question:*Bagaimana menahan serangan virus*?
(How to resist a virus attack?) (3)
Pattern Answer : <COMPARE>
Answer: *sistem imun* (immune system) (6)

From the pattern of the sentence (4), the answer <POPULATION> was produced, since the source of the sentence (1) found no role <POPULATION>, the sentence (4) was the ignored and did not include the generation of questions for the sentence (1).

Question :*Apapenyebabserangan virus* ?
(What are the causes of virus attack?) (4)
Pattern Answer : <POPULATION>
Answer: null (7)

Experimental

Experimental Data

In our knowledge, we have built our own Indonesian medical sentence and named entity based on PICO frame. We collected Indonesian medical articles fromtwo popular Indonesian sites (*http://health.detik.com/* and *health.kompas.com/konsultasi*) for datain 2014 (after eliminating the similar medical entity, 500 medical sentences for eachclassification andfeatures were obtained).We randomly selected 60% of sentences for testing and 40% for training. We adopted the default regularization parameter and tested few cost-factor values to adjust the rate between Precision and Recall on the validation set.

Experimental Result

To evaluate the proposed method we have used a combination of the different sentences (see Table 4) such as active sentence, passive sentence, complex sentences with the conjunction "*tetapi*" (but), complex sentences with the conjunction "*sedangkan*" (while) and complex sentences with the conjunction "*ketika*"(when). This approach then achieved F-measure values of 0.88, precision 0,84 and recall 0,86 for active sentence.It can be concluded here that the pattern matching criteria of the training set and semantic role labeling based on PICO frame can be reproducible with minimal expert intervention.

Table 4. Result of Experiment with Combination Sentence

Combination of Sentence	F-Measure	Precision	Recall
active sentence	0,88	0,84	0,86
passive sentences	0,85	0,79	0,81
complex sentences with the conjunction "*tetapi*"	0,77	0,81	0,82
complex sentences with the conjunction "*sedangkan*"	0,79	0,76	0,78
complex sentences with the conjunction "*ketika*"	0,78	0,73	0,76

Conclusion and Future Work

A new method to the discovery of Indonesian medical question answering pairs pattern with question generation is presented in this research. This method only uses algorithms in the form of transformation rule and filtering based on the predicate argument. Since the results in this early phase have been found very encouraging, particularly when the very simple question selection scheme was taken into account, the chance of further improvements in recognition accuracy was high. Using question generation one might switch very quickly to transformation sentence and found in the pattern without any needs of a linguist expert.

The research has been conducted regarding that generation of questions still uses the concept-based and rule-based pattern. In fact, there are some disadvantages when using these concepts, including the need to build a knowledge base for managing rules and patterns needed to support the system as well as the need to always update the new knowledge in anytime.

The future machine learning method allows the system to learn patterns of text in which the question and the answer might be not possible, and then convert these patterns into the formula. Formula learning system is supposed to be made at levels where each level builds a set of formulas that are different in quality and quantity of the other levels. The system then provides weight to each formula representing QG query capabilities to produce and find the right answer.

Acknowledgment

We thank anonymous reviewers for their helpful comments.

References

[1]. Xuchen Yao, Gosse Bouma and Yi Zhang, 2012, “Semantics-based Question Generation and Implementation”, In Dialogue and Discourse Journal 3(2), pp. 11-42.

[2]. Michael Heilman, Noah A. Smith, 2010, “Good Question! Statistical Ranking for Question Generation”,Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL , pages 609–617, Los Angeles, California.

[3]. Michael Heilman and Noah A. Smith, 2009, “Question Generation via Overgenerating Transformations and Ranking”, Language Technologies Institute, School of Computer Science, Carnegie Mellon University, CMU-LTI-09-013, www.lti.cs.cmu.edu

[4]. David Lindberg, Fred Popowich, John Nesbit and Phil Winne, 2013, “Generating Natural Language Questions to Support Learning On-Line”, Proceedings of the 14th European Workshop on Natural Language Generation, pages 105–114, Sofia, Bulgaria, August 8-9 2013.

[5]. Michael Heilman, 2011, “Automatic Factual Question Generation from Text”, thesis on Language Technologies Institute, School of Computer Science, Carnegie Mellon University, www.lti.cs.cmu.edu.

[6]. Husam Ali,YlliasChaliSadid, A. Hasan, 2010, “Automatic Question Generation from Sentences”, TALN 2010, Montréal, 19–23 juillet 2010

[7]. David Lindberg, 2013, “Automatic Question Generation from Text for Self-Directed Learning”, Thesis of Master of Science in the School of Computing Science, Faculty of Applied Sciences, Simon Fraser University.

[8]. Xuchen Yao, 2010, “Question Generation With Minimal Recursion Semantics”, Master Thesis in Language and Communication Technologies University of Groningen & Saarland University.

[9]. Laszlo Bednarik, Laszlo Kovacs, 2012, “Automated EA-type Question Generation from Annotated Texts”, 7th IEEE International Symposium on Applied ComputationalIntelligence and Informatics 2012, Romania, pp: 191-195.

[10]. Xuchen Yao, Gosse Bouma, Yi Zhang, 2012, “Semantic based Question Generation and Implementation” , Journal of Dialogue and Discourse, volume 3(2), pp : 11-42.

[11]. Per Urlaub, 2012, “Reading strategies and literature instruction: Teaching learners to generate questions to foster literary reading in the second language”, In journal System, volume 40(2), pp : 296-304.

[12]. Noriyuki Iwane, Chunming Gao, Makoto Yoshida, 2013, “Question Generation for Learner Centered Learning”, IEEE 13th International Conference on Advanced Learning Technologies, pp : 330-332.

[13]. Alfan Farizki Wicaksono, Ayu Purwarianti, 2010, "HMM Based Part-of-Speech Tagger for Bahasa Indonesia". Proceeding of the Fourth International MALINDO Workshop (MALINDO2010). Agustus 2010. Jakarta, Indonesia

[14]. H. Yu, 2006, “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.

[15]. H. Yu, C. Sable, 2005[a], “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.

[16]. H. Yu, C. Sable, H.R. Zhu, 2005[b], “Classifying Medical Questions based on an Evidence Taxonomy”, in: Proceedings of the AAAI-05 Workshop Question Answering in Restricted Domains, AAAI, Pittsburgh, PA, USA.

[17]. P. Jacquemart, P. Zweigenbaum, 2003, “Towards a medical question-answering system: a feasibility study”, Stud. Health. Technol. Inform. 95, p: 463–468.

[18]. Y. Niu, G. Hirst, 2004, “Analysis of semantic classes in medical text for question answering”, in: Proceedings of the ACL-2004 Workshop Question Answering in Restricted Domains, ACL, Barcelona, Spain.

[19]. Y. Niu, G. Hirst, G. McArthur, P. Rodriguez-Gianolli, 2003, “Answering clinical questions with role identification”, in: Proceedings of the ACL-2003 Workshop Natural Language Processing in Biomedicine, ACL, Sapporo, Japan, pp: 73–80.

[20]. Y. Niu, X. Zhu, G. Hirst, 2006, “Using outcome polarity in sentence extraction for medical question-answering”, in: Proceedings of the AMIA 2006 Symposium, AMIA, pp : 599-603

[21]. D. Demner-Fushman, B. Few, S.E. Hauser, G. Thoma, 2006[a], “Automatically identifying health outcome information in MEDLINE records”, J. Am. Med. Inform. Assoc. 13, p : 52–60.

[22]. D. Demner-Fushman, J. Lin, 2005, “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.

[23]. D. Demner-Fushman, J. Lin, 2006[b], “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.

[24]. D. Demner-Fushman, J. Lin, 2007, “Answering clinical questions with knowledge-based and statistical techniques”, Comput. Linguist, pp: 3363–103.

[25]. Koichiro Yoshino, Shinsuke Mori, and Tatsuya Kawahara, Language Modeling for Spoken Dialogue System based on Sentence Transformation and Filtering using Predicate-Argument Structures, In *Proc. of SIGDIAL*, pages 59–66, 2012

[26]. Daisuke Kawahara, Nobuhiro Kaji, Sadao Kurohashi, 2003, “Question and Answering System based on Predicate-Argument Matching”, Proceedings of the Third NTCIR Workshop2003, National Institute of Informatics, pp : 150-153.

[27]. Tomohide Shibata, Sadao Kurohashi, Shotaro Kohama, Akihiro Yamamoto, 2013, “Predicate-argument Structure based Textual EntailmentRecognition System of KYOTO Team for NTCIR-10 RITE-2”, Proceedings of the 10th NTCIR Conference, June 18-21, 2013, Tokyo, Japan, pp : 537-544

[28]. Knuth D.E., Morris(Jr) J.H., Pratt V.R., 2001, “Fast pattern matching in strings”, *SIAM Journal on Computing* 6(1):323-350.