

## ABSTRAK

### **SISTEM TANYA JAWAB MEDIS BERBASIS *CASE BASED REASONING* MENGGUNAKAN *SEMANTIC ROLE LABELING***

Oleh

**Wiwin Suwarningsih**

**NIM : 33213014**

**(Program Studi Doktor Teknik Elektro dan Informatika)**

Sistem tanya jawab (*Question Answering System=QAS*) adalah sistem yang menerima masukan berupa pertanyaan dalam bahasa alami dan secara otomatis memberikan keluaran berupa jawaban singkat dalam bahasa alami pula. Permasalahan mendasar pada sistem tanya jawab adalah bagaimana meningkatkan kemampuan untuk klasifikasi pertanyaan, perumusan untuk pencarian jawaban yang tepat, dan penanganan akuisisi pengetahuan baru. Metoda yang digunakan pada penelitian disertasi ini berbasis CBR (*Case Based Reasoning*) dengan menggunakan SRL (*Semantic Role Labelling*). Hal ini dilakukan karena sebuah paradigma dominan dibidang medis adalah *Evidence-Based Medicine* (EBM) yang mengacu pada penggunaan bukti terbaik yang diperoleh dari penelitian ilmiah. Sehingga diperlukan suatu cara untuk pencarian bukti terbaik dengan menggunakan pelabelan peran semantik secara otomatis dalam pendeteksian dan pengenalan entitas medis. Kontribusi pada disertasi ini adalah (i) Menggabungkan pendekatan klasifikasi pertanyaan berdasarkan EAT (*Expected Answer Type*) dengan transformasi kalimat tanya ke dalam pola template berbasis semantik; (ii) Transformasi bahasa alami ke elemen PICO (*Problem, Intervention, Compare, Outcome*) untuk pencarian jawaban yang tepat; (iii) Menerapkan metoda *Question Generation* (QG) dalam proses akuisisi pengetahuan baru untuk menghasilkan pola pasangan pertanyaan-jawaban (QA-pairs).

Penggabungan metoda klasifikasi pertanyaan dengan transformasi bahasa alami ke dalam bentuk elemen PPPICCOODTQ adalah pendekatan yang diusulkan untuk mencari dan mendapatkan jawaban yang tepat. Proses transformasi yang dilakukan menggunakan pelabelan peran semantik (SRL) secara otomatis untuk setiap kata atau frase kunci. Hasil dari transformasi ini akan dimanfaatkan untuk klasifikasi berdasarkan EAT dan digunakan untuk pendefinisian pola template pada setiap kalimat tanya. Langkah-langkah yang dilakukan untuk mencapai tujuan tersebut adalah memanfaatkan skema CBR yang terdiri dari 4 tahap kerja, yaitu *retrieve, reuse, revise* dan *retain*. Tahap *retrieve* yaitu mencari dan mengambil kasus yang mirip dari basis data kasus. Dua tahap pendekatan untuk pencarian kasus yang mirip telah diusulkan, yaitu: (i) Pencarian berbasis entitas adalah pencarian dengan membuat variasi sintaks dari kalimat tanya. (ii) Pencarian berbasis *bag-of-word* adalah membuat struktur pohon dari kalimat tanya dan menentukan relasi antar simpul di pohon tersebut. Akurasi yang dicapai oleh pendekatan yang diusulkan

mencapai 88,64%. Tahap *reuse*, *revise* dan *retain* dalam disertasi ini mencoba menggabungkan tiga tahap tersebut menjadi skema baru yang digunakan untuk memilah jawaban yang tepat. Pemilahan jawaban berdasarkan pencocokan pola berbasis SRL dengan elemen PICO. Skema yang diusulkan ini mencapai akurasi 88,33%.

Pola pasangan kalimat pertanyaan-jawaban untuk domain medis dihasilkan dari proses pembangkitan pertanyaan. Pembangkitan pertanyaan ini adalah cara untuk menambah pengetahuan baru pada sistem tanya jawab medis. Cara kerja dari pembangkitan pertanyaan adalah melakukan transformasi kalimat deskriptif menjadi kalimat interogatif dengan membuat pola template berbasis semantik. Dasar pemikirannya adalah ketika ada sebuah pertanyaan, sistem dapat dengan mudah memberikan jawaban karena setiap pasangan pertanyaan-jawaban sudah tersedia dalam basisdata sistem. Hasil dari proses transformasi ini adalah pola template pasangan pertanyaan-jawaban sebanyak 80 pasang pola. Pasangan kalimat pertanyaan-jawaban yang telah dihasilkan berdasarkan pola tersebut adalah 157 pasang kalimat. Akurasi yang dicapai oleh pendekatan yang diusulkan mencapai 85,71%.

Kata kunci: QA-medis, CBR, SRL, pembangkitan pertanyaan, elemen  
PPPICCOODTQ

## **ABSTRACT**

### **CASE BASED REASONING MEDICAL QUESTION ANSWERING SYSTEM USING SEMANTIC ROLE LABELING**

By

**Wiwin Suwarningsih**

**NIM : 33213014**

**(Doctoral Study Program of Electrical Engineering and  
Informatic)**

Question Answering System (QAS) is a system that accepts an input in the form of questions in a natural language and automatically provides an output in the form of short answers in a natural language. The fundamental problems in the system are regarding the optimization of the question of classification, detection of similarity for the formulation of the right answer, ambiguous handling of the knowledge base and the acquisition of expert knowledge. This dissertation research proposes an approach by combining CBR (Case Based Reasoning) and SRL (Semantic Role Labeling) for the medical debriefing system and resolve the fundamental problems in the question and answer system with contributions to (i) combining for classification approach for the question by EAT (Expected Answer Type) with transformation of natural language element to PICO (Problem, Intervention, Compare, Outcome) to search for answers; (ii) the phrase of the question-answer pair for the medical domain from the generation process questions.

The combining between classification of questions and transformation of natural language into a form element PPPICCOODTQ is a proposed approach to search and obtain a right answer. Some steps taken to achieve these objectives are to utilize CBR scheme that consists of four stages of work: retrieve, reuse, revise and retain. The retrieving phase includes searching and retrieving the similar cases of database instances. A two-stage approach for the search of similar cases has been proposed: (i) the entity-based search as a searching by varying the syntax of the sentence asked. (ii) Search-based bag-of-words making the tree structure of interrogative sentence and determining the relationships among the node in a tree. Accuracy achieved by the proposed approach reaches 88.64%. The phase of reusing, revising and retaining in this dissertation attempts to combine these three steps into a new scheme used to sort out the right answer. Sorting answers are based on pattern matching-based PICO SRL elements. The proposed scheme achieves the accuracy of 88.33%.

Some questions-answering pairs (QA-pairs) pattern for the medical domain are resulted from the questions generation process. The generation of these questions

is how to add new knowledge to the medical question and answer system. The workings of the generation question are to transform the descriptive sentence into interrogative sentences to create a pattern-based template semantics. The rationale is when there is a question, the system can easily provide some answers for each question-answer pair already available in the system database. The results of this transformation process is a template pattern of question-answer pair as many as 80 pairs of patterns. Couple sentence-answer questions that have been generated based on the patterns are 157 pairs of sentences. Accuracy achieved by the proposed approach could reach 85.71%.

Keywords: QA-medical, CBR, SRL, generating questions, PPPICCOODTQ elements