

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:

Gender:

Select Publication type:

Protocol B:

Search MEDLINE/PubMed via PICO

Patient/Problem:

Intervention:

Compare to (leave blank if none):

Outcome (optional):

Age Group:

Gender:

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 penggalan 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]-(R)-[B]
[A]-(R)-[which Y]
Does [A]-(R)-[B]
Why [A]-(R)-[B]
[Which X,Y]-(R)-[B]
[Which X]-(R)-[B,C]
Duration [A]-(precedes)-[B]
Define [A]
Which specific precaution if [A]-(R)-[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.

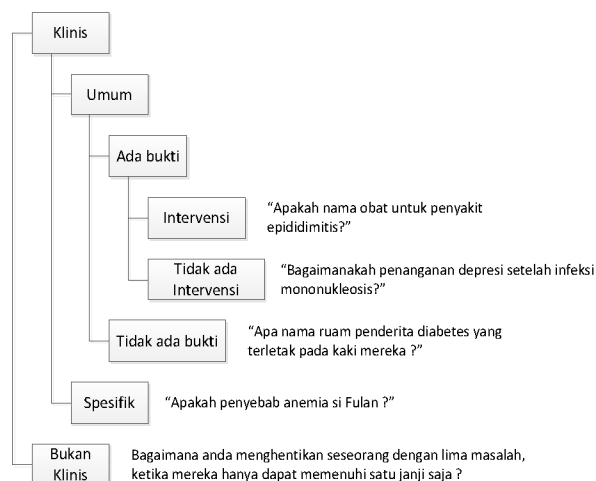
Inti dari sistem QA medis Terol *et al.* adalah modul untuk analisis pertanyaan. Tahap analisis pertanyaan terdiri dari klasifikasi pertanyaan dan analisis pertanyaan. Tahap klasifikasi pertanyaan terdiri dari penurunan bentuk logika pertanyaan, ekstraksi dari bentuk logika kata kerja, NER (*named entity 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 penyusain 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.

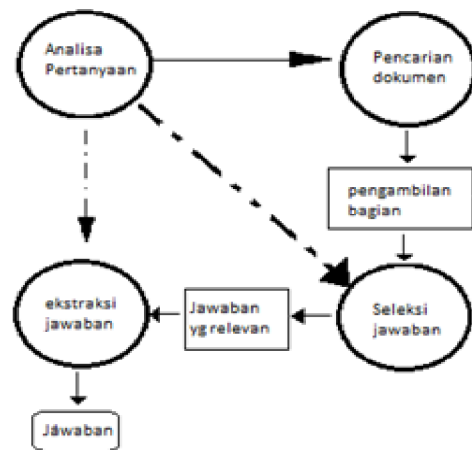
Presentasi jawaban agregat menghilangkan informasi yang berlebihan, secara otomatis untuk menghasilkan ringkasan terstruktur dan menyajikan ringkasan kepada pengguna.

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

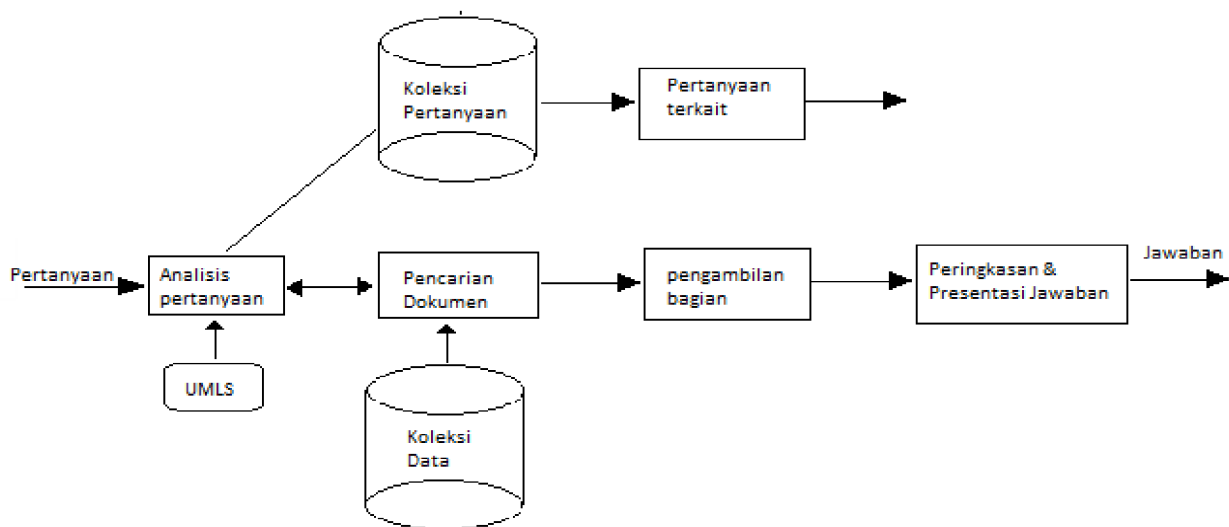
Apakah obat yang dapat diminum untuk kondisi X ? Apakah penyebab dari gejala X ? Apakah indikasi hasil tes menunjukkan situasi X ? Apakah dosis obat X? Bagaimana seharusnya saya mengobati kondisi X ? Bagaimana seharusnya saya mengelola kondisi X ? Apakah penyebab dari temuan fisik X? Apakah penyebab 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.

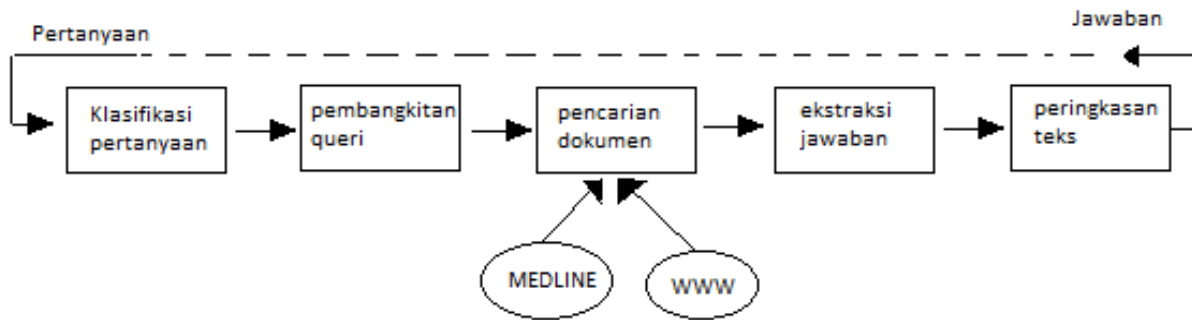
Terol et al. [17] melakukan penelitian dengan menggunakan pendekatan berbasis logika untuk mengadaptasi generik sistem tanya jawab medis. Sistem tanya jawab medis yang dirancang adalah untuk menjawab pertanyaan bahasa alami, dimana pertanyaan yang paling sering berupa pertanyaan taksonomi generik klinis. 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 diatas 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.