

SURVEY MODEL-MODEL Pencarian Informasi Rekam Medik Elektronik

Abstract

The rapid growth of medical records is an issue that must be anticipated. To handle the phenomenon of information overload in medical record information, it is necessary to do study in depth to develop a model of medical record information filtering that effectively supports the quality improvement of information search process recommendations. Various researches related to the search for medical information have been widely used, including developing research with concentration on novelty and diversity, using fuzzy ontology, tensor based factor, considering intentions / intention of users when searching and approach by combining phrase-based search with existing concept mapping tools MetaMap and ULMS Metathesaurus data sources.

Keywords : *Information Retrieval, Electronic Medical Record*

Abstrak

Pertumbuhan jumlah data rekam medik yang pesat, menjadi masalah tersendiri yang harus diantisipasi. Untuk menangani fenomena information overload dalam informasi rekam medis, perlu studi yang mendalam untuk dapat mengembangkan model filtering informasi rekam medik yang secara efektif mendukung peningkatan kualitas rekomendasi proses pencarian informasi. Berbagai penelitian terkait pencarian informasi medis telah banyak dilakukan, diantaranya mengembangkan penelitian dengan konsentrasi pada kebaruan dan keberagaman, menggunakan fuzzy ontology, berbasis factor tensor, memepertimbangkan niatan/intention pengguna ketika melakukan pencarian serta pendekatan dengan menggabungkan pencarian berbasis frasa dengan alat pemetaan konsep yang ada menggunakan MetaMap dan sumber data ULMS Metathesaurus.

Kata Kunci : Pencarian Informasi, Rekam Medik Elektronik

1. PENDAHULUAN

Institusi pendidikan kedokteran mengalami perubahan trend yang salah satunya dipengaruhi oleh keberadaan teknologi informasi dan komunikasi (*Information and communication technology/ICT*). Perkembangan teknologi informasi dapat dimanfaatkan untuk membantu institusi pendidikan kedokteran dalam menjamin kelancaran proses pendidikan dan pencapaian kompetensi, seperti yang termuat dalam Standar Pendidikan Profesi Dokter Indonesia. Peran teknologi informasi tersebut diantaranya yaitu (Herman RB, Sukarya WS, Rasmin M, Soebono H, Yuniadi Y, Soemitro D, 2012):

- Untuk mengembangkan sistem informasi akademik, pengembangan pangkalan data, dan telekonferensi.
- Jaringan internet dengan *bandwidth* yang memadai untuk menunjang proses pembelajaran.
- Tersedia perpustakaan elektronik untuk mengakses *e-book* dan *e-journal*.

Beberapa studi menunjukkan bahwa sumber informasi yang biasa dipakai oleh tenaga kesehatan untuk menangani permasalahan kesehatan adalah bersumber dari kolega, buku referensi dan mengambil literatur melalui pustakawan (Quintana, 1998). Akses informasi yang relevan dan terbaru menjadi penting untuk meningkatkan kompetensi calon dokter. Data rekam medik pasien yang merupakan tindakan penanganan permasalahan kesehatan pasien oleh kolega dokter, menjadi menarik untuk dipertimbangkan sebagai bahan informasi pembelajaran bagi calon dokter.

Pemanfaatan sistem rekam medik elektronik semakin berkembang pada berbagai lembaga kesehatan. Berdasarkan survey, 82% dokter terindikasi telah memakai atau akan memakai sistem pendataan rekam medik elektronik(Li et al., 2015).

Sistem pendataan rekam medik elektronik semakin berkembang dalam dunia medis. Keuntungan dari penerapan teknologi informasi dalam pendataan rekam medis pasien, membuat perkembangan sistem rekam medik elektronik semakin masif. *Accreditation Council for Graduate Medical Education* (ACGME) memberikan rekomendasi kepada institusi pendidikan medis untuk mendayagunakan Rekam Medik Elektronik untuk mendukung capaian pembelajaran(Tierney et al., 2013).

Pengembangan sistem *Clinical decision support* (CDS) dalam sistem rekam medik elektronik memiliki potensi untuk meningkatkan pemahaman dan pengetahuan serta menjadi pedoman pembelajaran bagi calon dokter. *Point-of-care education* yang diakses via CDS memungkinkan akses data medik terkini dan relevan untuk mendukung diagnosa dan rencana keperawatan calon dokter(Tierney et al., 2013).

Pertumbuhan jumlah data rekam medik yang terus bertambah setiap harinya, menjadi masalah tersendiri yang harus diantisipasi. Penggunaan sekunder data rekam medik elektronik bergantung pada kemampuan untuk mendapatkan informasi yang lengkap dan akurat terkait populasi pasien yang diinginkan. Pencarian data medis menjadi masalah teknis yang penting untuk mendapatkan informasi pasien secara cepat dan efektif(Li et al., 2015). Untuk menangani fenomena *information overload* dalam informasi rekam medis, perlu studi yang mendalam untuk dapat mengembangkan model filtering informasi rekam medik yang secara efektif mendukung peningkatan kualitas pembelajaran calon dokter.

Teknologi pencarian medis sangat penting bagi pengguna untuk dapat menemukan informasi yang bermanfaat dan relevan dengan cepat dan efektif dari kumpulan data medis dan klinis yang sangat besar. Karena kompleksitas istilah medis, metode pencarian informasi tradisional tidak dapat sepenuhnya digunakan untuk memenuhi permintaan kueri dan mengeksplorasi pengetahuan semantik potensial dalam dokumen medis(Kang et al., 2017).

Banyak peneliti yang sudah melakukan kajian umum teknis sistem pencarian informasi dan menemukan kenyataan bahwa sistem pencarian informasi tradisional tidak sepenuhnya dapat menyelesaikan permasalahan pencarian informasi medis. Hal ini diperkuat dengan kenyataan bahwa industri masih menghadapi kesulitan khususnya dalam mengelompokkan informasi yang relevan berdasarkan riwayat permintaan pengguna. Alasan pokok terkait permasalahan tersebut adalah disebabkan oleh proses pencarian informasi medis yang belum sempurna dan sesuai harapan(El-Bathly et al., 2011).

Penelitian ini mereview berbagai pendekatan dalam pencarian informasi, sehingga dapat dijelaskan karakteristik dari masing-masing pendekatan tersebut untuk dapat disesuaikan dengan kebutuhan pencarian informasi medis, khususnya dalam kebutuhan pencarian informasi medis sebagai sumber pengetahuan dan pembelajaran medis bagi calon dokter. Hasil penelitian yang digunakan untuk review adalah hasil penelitian yang dipublikasikan dalam jurnal penelitian yang terbit dari tahun 2011 sampai tahun 2017

2. ASPEK-ASPEK PENTING DALAM PENCARIAN INFORMASI MEDIS

Sistem pencarian yang digunakan dalam kumpulan dokumen tradisional umumnya menggunakan pendekatan berbasis konten/*content-based approach* seperti *vector space model*, *latent semantic indexing* maupun *non-negative matrix factorization*. Pendekatan ini digunakan untuk mengukur kesamaan antara kueri dan dokumen untuk merekomendasikan dokumen yang paling relevan. Pendekatan ini sulit diterapkan untuk menangani kompleksitas dalam terminologi medis(Li et al., 2015).

Pencarian informasi adalah teknologi atau proses menemukan informasi yang relevan sesuai dengan permintaan pengguna dari kumpulan koleksi dokumen. Hal ini utamanya didasarkan pada metode mencocokkan kesamaan kata kunci atau kesamaan konsep dalam proses

pencarian informasi, seperti Model Boolean, *Vector space model* dan Model Probabilistik, yang kurang cocok digunakan oleh profesional dan untuk pencarian informasi medis yang kompleks meskipun metode analisis berbasis semantik telah banyak diperkenalkan untuk pencarian informasi (Kang et al., 2017).

Penerapan skenario komputasi pencocokan kemiripan rekam medik elektronik meliputi kecocokan latar belakang sosial pasien, sketsa penyakit, deteksi efek samping obat, kualitas layanan medis, dll. Dibandingkan dengan perhitungan kesamaan teks secara umum, dimana perhitungan kemiripan rekam medis elektronik memiliki kesulitan dan tantangan yang unik. Salah satu bentuk tantangannya adalah bahwa rekam medis elektronik mengandung sejumlah besar terminologi biomedis, yang harus diperhatikan untuk menghitung kesamaan/kemiripan secara akurat. Di sisi lain, catatan medis elektronik sangat sensitif terhadap ungkapan negatif, yang harus diperhatikan untuk melakukan perhitungan yang akurat (Zhang et al., 2017).

Sistem perhitungan kemiripan rekam medis elektronik yang lengkap harus mencakup tiga bagian berikut: sistem informasi rumah sakit, *data warehouse* yang dibangun untuk mendukung analisis rekam medis elektronik, dan algoritma untuk menghitung kemiripan data rekam medik elektronik (Zhang et al., 2017). Data masukan untuk algoritma dapat berupa kumpulan dokumen rekam medis elektronik yang semi terstruktur. Informasi pasien yang digunakan meliputi biodata pasien, riwayat, keluhan, diagnosis, catatan kemajuan, catatan bedah dan informasi lainnya. Outputnya adalah kemiripan dua rekam medis elektronik dari serangkaian kumpulan dokumen rekam medis elektronik.

Isu utama dalam pencarian informasi medis adalah adanya keragaman pengguna layanan ini. Secara umum pengguna dapat memiliki perbedaan kategori informasi yang dibutuhkan, perbedaan tingkat pengetahuan medis dan perbedaan kemampuan bahasa (Goeuriot et al., 2016).

3. PERKEMBANGAN MODEL PENCARIAN INFORMASI MEDIS

Pencarian informasi berbasis kesamaan semantic telah banyak dilakukan dalam penelitian pencarian informasi (Kang et al., 2017) (Hliaoutakis et al., n.d.). Meskipun begitu, perkembangan penelitian dalam bidang pencarian informasi telah mengalami banyak kemajuan. Berikut disampaikan beberapa penelitian yang memperhatikan hal-hal lain selain kesamaan semantic untuk dilakukan dalam proses pencarian informasi rekam medis.

a. *Diversity Aware Information Retrieval* (Li et al., 2015)

Model penelitian ini dapat dikelompokkan ke dalam kategori pencarian medis semantik dengan menganggap faktor relevansi dan kebaruan sebagai aspek penting untuk perankingan dokumen rekam medis (Li et al., 2015). Pada penelitian ini Peneliti mengusulkan pendekatan baru untuk menghasilkan keragaman hasil pencarian informasi rekam medis dengan mempertimbangkan relevansi dan kebaruan untuk perankingan dokumen rekam medis. Pendekatan ini menggabungkan pencarian informasi berbasis semantic dan keberagaman hasil pencarian untuk mengatasi ambiguitas dalam pencarian informasi medis.

Berbeda dengan strategi diversifikasi yang ada, dimana banyak mengandalkan data log kueri, pendekatan yang diusulkan ini menggunakan ontologi medis yang berisi banyak pengetahuan medis untuk mengurangi ambiguitas kueri asli dengan membentuk beberapa sub-kueri. Setiap sub-kueri mewakili satu aspek dari maksud yang tersirat dari kueri asli. Berdasarkan aspek yang dimodelkan sub-kueri, diberikan sebuah strategi baru yang mengeksplorasi disambiguitas kueri untuk menghasilkan keragaman pencarian medis.

Berdasarkan domain ontologi medis yang ada, dicari dasar semantik dari kueri dan digunakan untuk memodelkan aspek-aspek yang mendasari kueri tersebut. Kemudian berdasarkan aspek kueri yang berhasil dimodelkan, diusulkan strategi diversifikasi baru, yang tidak hanya mempertimbangkan aspek kepentingan tetapi juga aspek kemiripan untuk melakukan perankingan informasi rekam medis.

Pendekatan yang diusulkan untuk *diversifikasi-aware* dalam pencarian informasi rekam medis mencakup dua langkah, yaitu:

- a) Pemahaman kueri untuk menemukan aspek tersirat dari kueri asli dan membentuk beberapa sub-kueri turunan yang memodelkan aspek yang berbeda dari kueri asli. Rincian proses pemahaman kueri berisi tiga sub langkah sebagai berikut:
 - i) Transformasi kueri. Sub-langkah ini melakukan dua fungsi, yaitu identifikasi frase kata kunci dan perluasan kata kunci. Secara formal, setelah proses langkah ini, setiap frase kata kunci $k_i \in q = \{k_1, k_2, \dots, k_m\}$, dimana kunci k bagian dari kueri q, diperluas ke kumpulan kata kunci $S_i = \{k_{i1}, k_{i2}, \dots, k_{ipi}\}$.
 - ii) Pemetaan konsep kandidat. Setelah m himpunan perluasan kata kunci $S_i (i = 1, \dots, m)$ terbentuk, sub-langkah ini memetakan kata kunci masukan pengguna ke serangkaian konsep kandidat yang didefinisikan dalam ontologi medis. Untuk setiap frase kata kunci $k_i \in q = \{k_1, k_2, \dots, k_m\}$ kita bisa menemukan satu set simpul konsep kandidat $C_i = \{c_{i1}, c_{i2}, \dots, c_{ini}\}$ $C_i = \{c_{i1}, c_{i2}, \dots, c_{ini}\}$, dengan rating nilai $R_i = \{r_{i1}, r_{i2}, \dots, r_{ini}\}$ $R_i = \{r_{i1}, r_{i2}, \dots, r_{ini}\}$. Diasumsikan $match(a,b)$ merupakan bentuk prosentase frase b muncul dalam frase kata kunci a, dimana r_{ij} merupakan $Max_{k \in S_i}$. konsep c_{ij} dipilih sebagai kandidat satu hanya jika $r_{ij} > 0,75$, di mana 0,75 dipilih untuk menyeimbangkan akurasi pencarian dan menghitung kompleksitas untuk pembangkitan sub-kueri berikutnya.
 - iii) Membuat turunan kueri. Sub langkah ini membuat sebuah daftar turunan dari kueri untuk memodelkan berbagai aspek dari q. Ide dasarnya adalah untuk mengekstrak sebuah daftar dari subgrafik dari ontology medis, di mana masing-masing sub grafik meliputi satu calon konsep dari setiap frase kata kunci $k_i \in q = \{k_1, k_2, \dots, k_m\}$ $k_i \in q = \{k_1, k_2, \dots, k_m\}$. Kemudian setiap sub-grafik diubah sebagai sub-query yang mewakili satu aspek potensial dari interpretasi kebutuhan informasi pengguna.
- b) *Diversifikasi-aware medical search*, keberagaman hasil pencarian dapat dinyatakan sebagai hasil transaksi antara factor *relevance* dan *Novelty*. Model ini untuk menghasilkan keberagaman hasil pencarian melalui proses perangkingan ulang S dari basis rangking berdasarkan turunan kueri dan daftar dokumen yang ada. Proses perangkingan ulang sebagaimana langkah berikut :
 - i) Menentukan dokumen peringkat pertama. Dokumen peringkat 1 teratas dalam dasar perangkingan dipilih untuk diletakkan pada posisi pertama dari daftar peringkat S, karena ini adalah yang paling relevan untuk q.
 - ii) Urutkan sisa dokumen dalam peringkat dasar. Setelah memilih dokumen peringkat pertama, dilakukan pemilihan dokumen (*local best document*) untuk diletakkan pada urutan selanjutnya. Dengan formulasi perhitungan:

$$LB(d, S) = \delta \cdot relevance(d, q) + (1 - \delta) \cdot novelty(d, S) \quad (1)$$

$$novelty(d, S) = \prod_{d_j \in S} novel(d, d_j) \quad novelty(d, S) = \prod_{d_j \in S} novel(d, d_j)$$

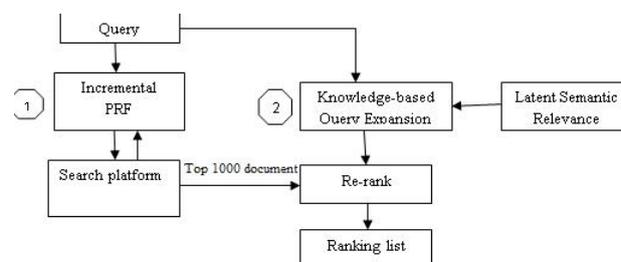
$$novelty(d, S) = \prod_{d_j \in S} novel(d, d_j) \quad (2)$$

$$novel(d, d_i) = \sum_{q_j \in Q} rs(q_j) \cdot \prod_{m \in [1, j-1]} (1 - (q_j, q_m)) \cdot rel(d, q_j) \cdot (1 - rel(d, d_i) \cdot rel(d_i, q_j)) \quad (3)$$

Manfaat penerapan pendekatan *diversity-aware* untuk pencarian medis yaitu: meningkatkan penggunaan sekunder dari rekam medis, meningkatkan kualitas perawatan medis dan menyeimbangkan penggunaan sumber daya medis. Beberapa hal yang harus diperhatikan yaitu pentingnya bekerja sama dengan pengguna, domain pengetahuan medis, mengulangi proses pengindekan untuk mengurangi waktu respon, dan menangani masalah privasi.

b. *Tensor Factorization* (Wang et al., 2017)

Pendekatan ini mengembangkan sistem pencarian informasi medis yang disempurnakan secara semantik, dengan menggunakan strategi perluasan kueri dua tahap (seperti yang ditunjukkan pada Gambar 1) untuk mengintegrasikan umpan balik relevansi pseudo dan perluasan kueri berbasis pengetahuan untuk meningkatkan kinerja pengambilan dokumen yang relevan dari kueri.



Gambar1. Pendekatan Faktor Tensor

Pertama, pendekatan ini mengusulkan pendekatan umpan balik relevan pseudo (*incremental PRF*) untuk perluasan kueri untuk mendapatkan daftar peringkat awal dokumen yang diambil. Kedua, mengembangkan metode perluasan kueri berbasis pengetahuan menggunakan model relevansi laten semantik baru. Metode yang diusulkan akan melakukan perangkingan ulang dokumen yang diambil oleh *incremental PRF* pada tahap pertama.

- 1) *Incremental Pseudo Relevance Feedback (incremental PRF)*. Metode populer umpan balik *pseudo-relevance* mengasumsikan bahwa dokumen yang teratas yang diambil adalah relevan dengan kueri sehingga semua terminologi dari dokumen ini digunakan untuk perluasan pencarian. Untuk memilih terminologi perluasan yang sesuai, diusulkan pendekatan heuristik, yang secara iteratif menggunakan skor nilai yang diberikan oleh sistem pencarian *Solr* sebagai indikator untuk meningkatkan nilai kemiripan antara kueri dan dokumen. Strategi inkremental digunakan untuk mengevaluasi kumpulan terminologi perluasan yang sesuai dan mengendalikan proses perluasan dengan ambang batas untuk mengurangi risiko kueri mengambang.
- 2) Model relevansi laten semantik. Setelah melakukan *incremental PRF*, sistem akan mendapatkan daftar 1000 dokumen teratas yang terkait dengan kueri. Pada bagian ini, diperkenalkan perluasan kueri berbasis pengetahuan dengan model relevansi laten semantik, yang mampu memberikan jalur perluasan yang optimal di bawah pengaturan yang tertentu, dan menggunakan kueri yang diperluas untuk melakukan perangkingan ulang pada 1000 dokumen untuk mendapatkan hasil akhir.
 - i) Representasi semantic berbasis Tensor. tensor menyediakan representasi alami untuk fitur-fitur heterogen seperti dependensi antara pasangan konsep dengan

hubungan semantik yang berbeda. Dalam kasus penelitian ini dipakai bentuk hubungan antara tipe semantik query, semantic relasi dan tipe semantic dokumen. Hubungan semantik yang berbeda memiliki dampak yang berbeda terhadap penilaian relevansi, dan hubungan semantik yang *multiple* membuat model alami tensor orde tiga.

Hubungan 3elemen (tipe semantik kueri, hubungan semantik, dan tipe semantik dokumen) diformulasikan sebagai berikut :

$$f(q_i, s_j, d_k) = \sum_{r=1}^D q_{ir} s_{jr} d_{kr} \quad (4)$$

$f(q_i, s_j, d_k) = x_{ijk}$, lalu digunakan untuk menilai relevansi semantik dari kueri dan dokumen yang sesuai.

$$f_o(q_i, s_j, d_k) = \frac{C_r(q_i, s_j, d_k)}{C_f(q_i, s_j, d_k)} \quad (5)$$

$C_f(q_i, s_j, d_k)$ adalah jumlah (q_i, s_j, d_k) dari keseluruhan himpunan pelatihan, dan $C_r(q_i, s_j, d_k)$ adalah jumlah tripel (q_i, s_j, d_k) yang diamati. Nilai untuk masing-masing tripel berkisar antara 0 sampai 1.

- ii) Faktorisasi Tensor. Metode *Alternating Least Squares method for CP decomposition* (CP-ALS) menghasilkan kinerja yang baik untuk memprediksi hubungan semantik dalam subset UMLS. Untuk itu digunakan metode CP-ALS untuk faktorisasi tensor dalam sistem yang dikembangkan. Dalam penelitian ini, digunakan metode serupa dengan basis metode CP-ALS dan algoritma optimasi heuristik. Pertama, nilai mean dari masukan yang diketahui digunakan sebagai nilai awal untuk estimasi masukan yang hilang. Kedua, memasukkan fungsi kesalahan tertimbang untuk estimasi peringkat.

$$f_w(A, B, C) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \{w_{x_{ijk}} (x_{ijk} - x'_{ijk})\}^2 \quad (6)$$

Dimana $w_{x_{ijk}} = 1$ jika x_{ijk} adalah masukan yang diketahui dan

$w_{x_{ijk}} = 0$ jika x_{ijk} adalah masukan yang hilang.

x'_{ijk} adalah nilai masukan dari x .

- iii) Perluasan kueri berbasis pengetahuan. Faktorisasi tensor order tinggi biasanya menimbulkan permasalahan komputasi yang rumit. Dalam sistem ini, faktorisasi hanya perlu dihitung sekali untuk memperkirakan skor nilai tiga asosiasi semantik. Oleh karena itu, tidak akan mempengaruhi efisiensi untuk pencarian informasi berdasarkan query yang spesifik. Pada metode perluasan yang diusulkan, kompleksitas meningkat secara linier sesuai dengan ambang iterasi. Skor untuk tiga asosiasi semantic, termasuk yang tidak teramati, dapat dihitung dengan mengambil hasil *inner product* dari vektor fitur laten yang terkait. Tiga asosiasi semantik ini memungkinkan perluasan jalur pencarian menggunakan relasi semantik dan constraint tipe semantik untuk membentuk konsep dalam jaringan semantik. Untuk setiap kueri, gabungan tiga asosiasi semantik teratas yang sesuai dengan konsep kueri ini akan digunakan sebagai perluasan jalur untuk mengidentifikasi perluasan konsep.

Percobaan dengan dataset TREC menunjukkan keefektifan sistem yang diusulkan. Sistem yang diusulkan memiliki potensi untuk disesuaikan dengan aplikasi pembelajaran komputer dan medis informatika lainnya, seperti sistem rekomendasi, pembelajaran ontologi, bioinformatika, dan lain-lain.

c. *Intention-Based Information Retrieval*(Cesar et al., 2016)

Penelitian ini bertujuan untuk mencari teknik-teknik baru dalam pencarian informasi dengan memperhatikan faktor niat/*intention* dari pengguna saat mereka membuat ataupun mencari informasi medis. Model ini mendefinisikan metode pencarian dengan mengandalkan anotasi makna dan kategori niat dalam deskripsi tekstual bahasa alami yang ada dalam data rekam medis. Teknik ini mengeksplorasi definisi terminologi dari teks medis berdasarkan sistem organisasi pengetahuan medis/*medical knowledge organization system* yang mencakup segala jenis model representasi pengetahuan seperti klasifikasi, kamus dan ontologi. Usulan model ini menggabungkan anotasi makna dengan jenis niatan yang terungkap dalam bagian dokumen rekam medik elektronik.

Penelitian ini mengusulkan untuk memformalkan algoritma peringkat dalam melakukan pengurutan hasil pencarian sesuai dengan string input pencarian pengguna yang mengekspresikan niatnya. Penelitian ini menunjukkan gambaran niatan yang ada dalam kueri dieksplorasi dengan anotasi niatan yang ada dalam rekam medis elektronik. Ekplorasi ini melibatkan banyak dimensi yang merepresentasikan kelas niatan dalam data rekam medis elektronik pasien. Basis teori yang digunakan untuk mengidentifikasi niatan meliputi teori perilaku bicara dan semiotic dan teori/kajian lain yang menyediakan metode tepat guna untuk mengidentifikasi dan membentuk struktur aspek yang terkait dengan identifikasi niatan dalam perilaku komunikasi.

Algoritma perangkaian dalam model ini dilakukan dengan tujuh langkah sebagai berikut:

(Sumber Masukan berupa Repositori rekam medis elektronik pasien, kueri dan domain *Knowledge Organization System* (KOS). Sumber masukan Ini terdiri dari repositori rekam medis elektronik dengan meta-data yang ditentukan dengan konsep dan anotasi illocutions, kueri spesifik dan domain ontologi yang terdapat dalam KOS.)

1 - *Syntactic retrieval*: Langkah ini mengambil dari repositori rekam medis elektronik untuk semua dokumen memunculkan kata kunci yang ditentukan oleh kueri pengguna. Hasil output diwakili oleh set SYNrel. Urutan dasar hasil pada tahap ini mencerminkan metode yang diterapkan pada platform pencarian sintaksis yang digunakan.

2 – *Keyword search in KOS* : memeriksa kemunculan kata kunci di semua konten KOS (seperti dalam UMLS atau domain ontologi). Langkah ini mendeteksi dan memilih konsep termasuk member label mereka dan menunjukkan sinonim kata kunci. Pendekatan ini juga memungkinkan mengeksplorasi konsep super dan sub dari KOS.

3 - *Semantic retrieval*: Langkah ini mengambil semua dokumen dari rekam medis elektronik yang memiliki label, sinonim dan terminologi lain yang sesuai dengan konsep hasil proses sebelumnya. Model ini mengeksplorasi platform pencarian sintaksis. Hasil output diwakili oleh set SEMrel.

4 - *Union*: Langkah ini menggabungkan himpunan SYNrel dan SEMrel. Hasil output diwakili oleh himpunan REL1.

5 – *Illocution-based Filtering* : Pada tahap ini, algoritma menyaring semua data rekam medis elektronik dari REL1 di mana kata kunci, label konsep atau istilah yang dipilih (sinonim dan istilah terkait) muncul di dalam fragmen teks yang diberi anotasi dengan masukan dari jenis *Illocution* dan dimensi nilai. Lebih khusus lagi, algoritma mempertimbangkan parameter berikut:

<term> +Dm <dimension> +VI <value>

<term> untuk kata kunci, label konsep, sinonim atau terminologi terminologi terkait, sementara <dimensi> menunjukkan jenis dimensi *Illocution* yang menunjukkan anotasi *Illocution* dalam fragmen teks. Parameter <value> menggambarkan nilai kuantitatif atau ekspresi yang mewakili dimensi dalam domain. Misalnya kata kunci "arthralgia", dengan dimensi waktu dan nilai "two days". Hasil keluaran dari filter ditunjukkan oleh himpunan REL2.

6 - *Ordering*: Jika beberapa rekam medis elektronik ada dalam REL2, algoritma akan mengurutkan dokumen sesuai dengan kecocokan antara istilah dan *Illocution* yang ditemukan pada tahap penyaringan. Pertama, algoritma memberi prioritas pada dokumen yang menyediakan kata kunci masukan persis dari kueri yang ditemukan dalam anotasi *Illocution*. Kedua, algoritma tersebut menempatkan laporan medis di mana label konsep dan sinonim ditemukan dalam anotasi *Illocution*. Akhirnya, semua terminologi yang sesuai ontologis yang terkait ditempatkan dalam daftar urutan. Dokumen yang berisi *Illocution* yang sesuai dengan kata kunci, label atau sinonim dari konsep memiliki relevansi yang lebih tinggi.

7 - *Priority by occurrence*: Algoritma memperbaiki urutan dokumen di REL2 sesuai dengan jumlah kemunculan istilah. Ini berarti bahwa dokumen dengan lebih banyak memunculkan kata kunci, label atau sinonim yang sesuai memiliki prioritas tertinggi dalam algoritma peringkat.

Hasil Keluaran: urutan hasil pencarian dokumen yang relevan.

Penelitian ini mendefinisikan algoritma perankingan untuk menentukan urutan hasil sesuai dengan jenis niat dari kueri pengguna dan dimensi yang berbeda dimana *Illocution* diklasifikasikan. Pendekatan yang diusulkan menggabungkan anotasi makna dengan elemen niat dalam dokumen teks dari repositori rekam medis elektronik. Penelitian ini mengungkapkan potensi metode dalam skenario pencarian informasi rekam medis elektronik, sesuai dengan perilaku algoritma dan hasil rekam medik.

d. *Fuzzy Ontology-Based Medical Information Retrieval*(Besbes, 2016)

Pendekatan ini bertujuan melakukan pencarian informasi yang efektif dengan memperluas pemahamannya tentang konsep medis yang ambigu, yang biasanya terjadi karena pengguna kurang terbiasa dengan istilah medis dan kesulitan dalam menunjukkan kebutuhan mereka. Pendekatan pengambilan informasi medis berbasis ontology fuzzy terdiri dari tiga komponen:

- a) Membangun ontology medis personal.

Sebuah ontology fuzzy terpersonalisasi untuk pengguna u didefinisikan sebagai berikut:

$$O_{fu} = \{C; R; A\}$$

Dimana C adalah sebuah himpunan konsep fuzzy, R adalah himpunan dari hubungan fuzzy dan A adalah himpunan aksiom yang ditunjukkan dalam sebuah Bahasa logika.

Mari kita perhatikan sebuah himpunan ontology $S = \{O_{f1}; O_{f2}; \dots; O_{fn}\}$ $S =$

$\{O_{f1}; O_{f2}; \dots; O_{fn}\}$ dimana $O_{f1}; O_{f2}; \dots; O_{fn}$ $O_{f1}; O_{f2}; \dots; O_{fn}$ adalah ontology fuzzy terpersonalisasi.

Komponen ini tersusun dari dua langkah yaitu :

1. Analisa Kueri
2. Pengembangan dan pengayaan ontology
Inputnya berasal dari kueri pengguna dan outputnya adalah sebuah ontology medis terpersonalisasi.

bentuk ontology fuzzy dapat didasarkan pada beberapa dukungan:

- Tekstual database

- Ontologi yang singkat dan padat dan penggunaan fungsi keanggotaan fuzzy
- Model XML dengan mengubah dokumen XML fuzzy dan DTD dalam ontologi fuzzy
- Wordnet untuk menghitung jarak antara konsep atau untuk mengekstrak konsep atau hierarki konsep dll.

b) Kontekstualisasi fungsi fuzzy dari ontology medis personal.

Tujuan dari komponen ini adalah untuk membentuk kontekstualisasi fungsi fuzzy dari ontology medis personal yang diperoleh dari komponen sebelumnya. Proses ini termasuk pemberian nilai keanggotaan dari setiap relasi untuk mendukung konsep yang paling relevan untuk pengguna tertentu. Untuk tujuan itu kami menggunakan profil pengguna dalam komponen ini.

Profil pengguna bisa berbentuk teks, semi terstruktur atau terstruktur. Ekstraksi dari informasi ditunjukkan dalam bentuk formal (basis data, ontology, dll) biasanya dapat dilakukan dengan bahasa kueri (SQL, SPARQL, dll). Bagaimanapun juga, dalam kasus penjelasan tekstual dari pengguna, sebuah analisis diperlukan untuk dilakukan untuk menentukan konteks penelitian pengguna. Dalam pekerjaan ini peneliti melakukan penjelasan profil secara tekstual.

Proses fuzzifikasi kontekstual dilakukan dalam dua tahapan:

- Ekstraksi kontekstual. Kekhususan metode ekstraksi ini terletak pada eksploitasi konsep yang muncul paling banyak dalam bahan medis yang membantu mengekstrak konsep relevan dari kueri pengguna dan profil pengguna.
- Fuzzifikasi kontekstual. Fungsi fuzzification yang diusulkan tergantung pada konteks ctx untuk mendukung konsep yang berhubungan dengan profil dan konsep pivotal c_p, c_p . Nilai keanggotaan dari hubungan antara sebuah konsep c_i, c_i dan konsep pivotal c_p, c_p , ditandai $s_{contextualized}(c_i)$ $s_{contextualized}(c_i)$, diformulasikan dalam persamaan 7:

$$s_{contextualized}(c_i) = \frac{c_{i \text{ rsub } \{i\}} + \{c\} \text{ rsub } \{p\} + ctx}{c_{i \text{ rsub } \{j\}}}$$

(7)

$$s_{contextualized}(c_i) = \frac{\sum_{j=1}^n c_{i \text{ rsub } \{j\}}}{h}$$

c) Reformulasi kueri.

Proses reformulasi dilakukan sbb: untuk masing-masing konsep kandidat dihubungkan dengan konsep pivotal dalam ontology personal fuzzy, jika nilai hubungan keanggotaan lebih besar dari pada nilai batas a , maka konsep kandidat digunakan untuk memperkaya kueri.

Penelitian tersebut memutuskan a pada 0,5 karena setelah berbagai variasi dari nilai ini, menandai bahwa dibawah bobot ini, hubungannya tidak signifikan.

Pendekatan ini secara otomatis mengekstrak konsep penting kueri pengguna dan membuat ontologi medis yang ditentukan untuk pengguna. Kontribusi yang kedua adalah mempertimbangkan fuzzifikasi otomatis dari ontologi yang dipersonalisasi dengan

mempertimbangkan hubungan taksonomi dan non-taksonomi dan profil pengguna. Kontribusi yang ketiga menyangkut integrasi ontologi fuzzy medis ke dalam proses perumusan query, yang didasarkan pada nilai keanggotaan yang terkait dengan semua hubungan yang ada dalam ontologi fuzzy untuk mendukung konsep yang relevan kepada pengguna.

Dalam konteks pencarian informasi, integrasi ontology terbukti efektif dalam meningkatkan ketepatan informasi yang dikembalikan, berkat teknologi pemodelan semantik yang dihadirkan.

e. *Flexible Concept Matching*(Nie dan Shen, 2015)

Pendekatan ini mengusulkan untuk menggabungkan pencarian berbasis frasa dengan alat pemetaan konsep yang ada, yaitu MetaMap. Selain itu, juga memperluas kueri dengan ekspresi sinonim yang tersimpan di ULMS Metathesaurus.

Asumsi yang dibuat dalam penelitian ini adalah bahwa menggabungkan pengambilan berbasis frasa dengan konsep yang diidentifikasi oleh alat (MetaMap) dan sinonimnya dapat menghasilkan hasil yang lebih baik lagi. Untuk menguji asumsi ini, dirancang sebuah pendekatan pencarian yang menggunakan konsep yang diidentifikasi oleh MetaMap, dan juga semua ungkapan alternatif, seperti ungkapan. Ungkapan ini disesuaikan dengan cara yang ketat atau fleksibel. Dengan demikian, dapat memanfaatkan pemetaan konsep dan pencocokan frasa yang fleksibel untuk konsep.

Sumber daya dan alat yang digunakan pada pendekatan ini adalah ULMS Metathesaurus dan MetaMap.

a) Penggunaan konsep-konsep sebagai frasa.

Pencocokan frasa telah dirumuskan sebagai model Markov Random Field (MRF), yang telah berhasil digunakan pada pencarian informasi umum. Model ini berisi tiga komponen: skor untuk pencocokan BOW menggunakan model bahasa tradisional, skor untuk pencocokan frasa yang tepat (frasa yang diurutkan) dan skor pencocokan frasa fleksibel (frasa tidak berurutan).

Lebih spesifik lagi, dalam model MRF berurutan, dua istilah yang berdekatan q_1q_2 dalam kueri dianggap sebagai frasa, dimana dua probabilitas berikut didefinisikan untuk pencocokan frasa yang berurutan dan tidak berurutan:

$$I(q_1q_2|D) = \frac{tf_{q_1q_2}(q_1q_2|D) + \mu P(I(q_1q_2)|C)}{|D| + \mu}$$

$$P(uw\delta(q_1q_2)|D) = \frac{tf_{uw\delta}(q_1q_2|D) + \mu P(uw\delta(q_1q_2)|C)}{|D| + \mu} \quad (8)$$

Nilai dokumen untuk pencocokan frasa yang tepat dan fleksibel kemudian didefinisikan sebagai berikut:

$$Score_{Ordered}(Q, D) = \frac{I(q_1q_2|D)}{P} \sum_{n(q_1q_2) \in Q}$$

$$uwg(q_1, q_2 | D) \quad (9)$$

$$Score_{Ordered}(Q, D) = \frac{1}{n} \sum_{(q_1, q_2) \in Q}$$

Skor akhir dengan MRF adalah kombinasi linier dari tiga nilai tersebut sebagai berikut:

$$score_{MRF}(Q, D) = \lambda_T Score_{BOW}(Q, D) + \lambda_O Score_{Ordered}(Q, D) + \lambda_U Score_{Ordered}(Q, D) \quad (10)$$

b) Perluasan kueri.

Karena UMLS Metathesaurus memberi berbagai konsep ekspresi alternatif, dapat dilakukan perluasan kueri dengan mempertimbangkan semua ungkapan alternatif. Pendekatan ini dapat diintegrasikan ke dalam satu skor peringkat sebagai berikut:

$$Score(Q, D) = \lambda_1 Score_{BOW}(Q, D) + \lambda_2 Score_{Phrase_{Exact}}(Q, D) + \lambda_3 Score_{Phrase_{Flex}}(Q, D) + \lambda_4 Score_{Phrase_{Bag}}(Q, D) \quad (11)$$

dimana nilai tersebut didefinisikan seperti di atas dan λ_i adalah parameter untuk disetel, dengan batasan $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$

Kebaruan pendekatan ini terletak pada kombinasi pemetaan konsep, pencocokan konsep fleksibel dan perluasan kueri. Eksperimen pada tiga koleksi uji publik menunjukkan bahwa pendekatan gabungan semacam itu melebihi metode yang ada.

4. ANALISIS MODEL PENCARIAN INFORMASI MEDIS

Model penelitian yang mempertimbangkan keragaman dan kebaruan penting untuk melakukan pengurutan peringkat dalam pencarian informasi. Pendekatan yang diusulkan untuk *diversity-aware* dalam pencarian informasi medis mencakup dua langkah, yaitu pemahaman kueri untuk menemukan aspek tersirat dari kueri asli dan membuat turunan sub-kueri yang sesuai dan langkah kedua yaitu mengeksplorasi beberapa sub-kueri untuk mendiversifikasi hasil pencarian medis.

Manfaat penerapan pendekatan *diversity-aware* untuk pencarian medis yaitu: meningkatkan penggunaan sekunder dari rekam medis, meningkatkan kualitas perawatan medis dan menyeimbangkan penggunaan sumber daya medis. Beberapa hal yang harus diperhatikan yaitu pentingnya bekerja sama dengan pengguna, domain pengetahuan medis, mengulangi proses pengindekan untuk mengurangi waktu respon, dan menangani masalah privasi

Pendekatan faktor tensor mengembangkan sistem pencarian informasi medis yang disempurnakan secara semantik, yang memiliki strategi perluasan kueri dua tahap yaitu incremental PRF dan model laten semantic yang relevan. Percobaan dengan dataset TREC menunjukkan keefektifan sistem yang diusulkan. Sistem yang diusulkan memiliki potensi untuk disesuaikan dengan aplikasi pembelajaran komputer dan medis informatika lainnya, seperti sistem recommender, pembelajaran ontologi, bioinformatika, dan lain-lain.

Penelitian pencarian informasi medis dengan mempertimbangkan niat pengguna mendefinisikan algoritma peringkat untuk menentukan urutan hasil sesuai jenis niat dari kueri pengguna dan dimensi yang berbeda dimana pengkodean diklasifikasikan. Penelitian ini

menggabungkan anotasi makna dengan elemen niat dalam dokumen teks dari repositori rekam medis elektronik. Yang dilakukan dengan 7 langkah yaitu: *Syntactic retrieval*, *Keyword search in KOS*, *semantic retrieval*, *union*, *Illocution-based filtering*, *Ordering*, dan *Priority by occurrence*. Dalam pendekatan yang dilakukan menunjukkan kegunaan pendekatan yang diusulkan untuk menyediakan fitur pencarian informasi yang lebih maju dalam rekam medis elektronik.

Pengambilan informasi berbasis fuzzy ontologi bertujuan melakukan pencarian informasi yang efektif dengan memperluas pemahamannya tentang konsep medis yang ambigu, yang biasanya terjadi karena pengguna kurang terbiasa dengan istilah medis dan kesulitan dalam menunjukkan kebutuhan mereka. Pendekatan pengambilan informasi medis berbasis ontology fuzzy terdiri dari tiga komponen yaitu membangun ontology medis personal, Kontekstualisasi fungsi fuzzy dari ontology medis personal, dan reformulasi kueri. Dalam konteks pencarian informasi, integrasi ontology terbukti efektif dalam meningkatkan ketepatan informasi yang dikembalikan, berkat teknologi pemodelan semantik yang mereka hadirkan.

Flexible Concept Matching mengusulkan untuk menggabungkan pencarian berbasis frasa dengan alat pemetaan konsep yang ada, yaitu MetaMap. Selain itu, juga memperluas kueri dengan ekspresi sinonim yang tersimpan di ULMS Metathesaurus. Yang dilakukan dengan dua acara yaitu menggunakan konsep-konsep sebagai frase dan melakukan perluasan kueri. Kebaruan pendekatan ini terletak pada kombinasi pemetaan konsep, pencocokan konsep fleksibel dan perluasan kueri. Eksperimen pada tiga koleksi uji publik menunjukkan bahwa pendekatan gabungan fleksible melebihi metode yang ada, yang mengonfirmasi keuntungan yang diharapkan dari metode yang diajukan.

5. HASIL DAN PEMBAHASAN

Penelitian mengenai pencarian informasi medis telah banyak dilakukan. Setiap pendekatan memiliki kekhususan dalam menyikapi masalah yang timbul dalam proses pengambilan informasi medis. Beberapa diantaranya pendekatan diversity-aware yang berusaha menemukan aspek tersirat dari kueri asli untuk dapat dieksploitasi menjadi beberapa sub kueri sehingga diperoleh diversifikasi hasil pencarian medis. Fuzzy ontology yang berperan membangun ontology personal untuk reformulasi kueri.

Pengambilan informasi medis berbasis faktor tensor yang dapat secara efektif memodelkan dan menggabungkan asosiasi laten semantic. Pendekatan *intention-based* mempertimbangkan niat pengguna ketika melakukan pencarian, dilakukan dengan menggabungkan anotasi makna dengan elemen niat dalam dokumen teks dari repositori. Pendekatan dengan menggabungkan pengambilan berbasis frasa dengan alat pemetaan konsep yang ada dalam hal ini menggunakan alat MetaMap dan sumber data ULMS Metathesaurus. Kebaruan pendekatan ini terletak pada kombinasi pemetaan konsep, pencocokan konsep fleksibel dan perluasan kueri.

KESIMPULAN

Dalam makalah ini, telah membahas berbagai pendekatan untuk melakukan pengambilan informasi medis. Semakin banyaknya jumlah EMR yang digunakan membuat semakin banyak data yang dapat dipergunakan kembali untuk berbagai tujuan demi meningkatkan kualitas pelayanan kesehatan. Banyak pendekatan telah dilakukan dalam melakukan pengambilan informasi medis tersebut. Masing-masing pendekatan memiliki tujuan dan kepentingan sesuai dengan kasus yang dihadapi dan ingin diselesaikan.

DAFTAR PUSTAKA

Besbes, G., 2016, *Fuzzy Ontology-based Medical Information Retrieval*, 178–185,

- Cesar, J., Paulo, S., Bonacin, R., Archer, C.T.I.R. dan Paulo, S., 2016, Intention-based Information Retrieval of Electronic Health Records, *25th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises*, [Online] tersedia di DOI:10.1109/WETICE.2016.56.
- El-Bathy, N., Azar, G., El-Bathy, M. dan Stein, G., 2011, Intelligent information retrieval lifecycle architecture based clustering genetic algorithm using SOA for modern medical industries, *IEEE International Conference on Electro Information Technology*, [Online] tersedia di DOI:10.1109/EIT.2011.5978565.
- Goeriot, L., Jones, G.J.F., Kelly, L., Müller, H. dan Zobel, J., 2016, Medical information retrieval: introduction to the special issue, *Information Retrieval*, [Online] 19 (1-2), 1-5, tersedia di DOI:10.1007/s10791-015-9277-8.
- Herman RB, Sukarya WS, Rasmin M, Soebono H, Yuniadi Y, Soemitro D, dkk., 2012, *Standar pendidikan profesi dokter*, KONSIL KEDOKTERAN INDONESIA.
- Hliaoutakis, A., Varelas, G. dan Petrakis, E.G.M., n.d., MedSearch : A Retrieval System for Medical Information Based on Semantic Similarity, *Computer Engineering*, 2-5,
- Kang, Y., Li, J., Yang, J., Wang, Q. dan Sun, Z., 2017, Semantic Analysis for Enhanced Medical Retrieval, *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*,
- Li, J., Liu, C., Liu, B., Mao, R., Wang, Y., Chen, S., Yang, J.J., Pan, H. dan Wang, Q., 2015, Diversity-aware retrieval of medical records, *Computers in Industry*, [Online] 6981-91, tersedia di DOI:10.1016/j.compind.2014.09.004.
- Nie, J. dan Shen, W., 2015, *Flexible Concept Matching for Medical Information Retrieval*, [Online] tersedia di DOI:10.1109/SMC.2015.332.
- Quintana, Y., 1998, Intelligent medical information filtering, *International Journal of Medical Informatics*, [Online] 51 (2-3), 197-204, tersedia di DOI:10.1016/S1386-5056(98)00115-4.
- Tierney, M.J., Pageler, N.M., Kahana, M., Pantaleoni, J.L. dan Longhurst, C.A., 2013, Medical Education in the Electronic Medical Record (EMR) Era, *Academic Medicine*, [Online] 88 (6), 748-752, tersedia di DOI:10.1097/ACM.0b013e3182905ceb.
- Wang, H., Zhang, Q. dan Yuan, J., 2017, *Semantically Enhanced Medical Information Retrieval System : A Tensor Factorization Based Approach*, 5,
- Zhang, Y., He, Z., Yang, J.J., Wang, Q. dan Li, J., 2017, Re-Structuring and Specific Similarity Computation of Electronic Medical Records, *Proceedings - International Computer Software and Applications Conference*, [Online] 2230-235, tersedia di DOI:10.1109/COMPSAC.2017.200.
-