

PENGGUNAAN POLA BAHASA ALAMI DALAM PENGETAHUAN AL QURAN

Dadad Zainal Musadad¹, Zen Munawar², Novianti Indah Putri³

1. Manajemen Informatika, Politeknik LP3I Bandung
2. Manajemen Informatika, Politeknik LP3I Bandung
3. Teknik Informatika, Fakultas Teknologi Informasi Universitas Bale Bandung

ABSTRACT

There are many studies that apply computational and natural language processing (NLP) techniques to help overcome some of the problems in increasing knowledge or gaining new insights into learning the Quran. Increasing knowledge of the Quran is a task that requires challenge and motivation because it is different from the grammar that is used by humans. An arrangement that has the characteristics of the text makes work that intends to search for meaning, with many meanings linguistic and has many meanings, it is not easy to do without using other sources such as Hadith. This study explains the improvement from part and make use of nl patterns in increasing English dictionary knowledge of the Quran text by using nlp techniques. The result was carried out the metrics of true positive, false positive and false negative, output in line with expectations. The final section outlines future research directions in the Arabic language literature, including the new application of sentiment based on natural language concepts.

Key Word: ontology learning, pattern extraction, concepts, relations, rules

ABSTRAK

Ada banyak penelitian yang menerapkan teknik komputasi dan natural language processing (NLP) untuk membantu mengatasi beberapa masalah dalam meningkatkan pengetahuan atau mendapatkan wawasan baru ke dalam pembelajaran Al Quran. Meningkatkan pengetahuan terhadap Al Quran adalah adalah tugas yang membutuhkan tantangan dan motivasi karena Al Quran berbeda dengan tatabahasa yang biasa digunakan manusia. Pengaturan yang mempunyai ciri khas dari membuat tugas untuk mengakses maknanya secara langsung, yang kaya akan linguistik dan makna yang berlapis-lapis, sulit dilakukan tanpa menggunakan sumber-sumber lain seperti Hadits. Penelitian ini menjelaskan peningkatan lapisan pertama dan penggunaan pola bahasa alami dalam meningkatkan pengetahuan teks terjemahan Al Quran bahasa Inggris dengan menggunakan teknik pemrosesan bahasa alami. Evaluasi dilakukan dengan menggunakan metrik positif benar, positif false dan negatif false dan telah diperoleh hasil yang menggembirakan. Bagian terakhir menguraikan arah penelitian masa depan dalam literatur bahasa Arab, termasuk penerapan baru dari sentimen berbasis konsep bahasa alami.

Kata Kunci: pembelajaran ontologi, ringkas pola, konsep, relasi, aturan

1. PENDAHULUAN

Ringkasan informasi otomatis dari sumber tidak terstruktur telah memberikan kesempatan untuk menanyakan, mengatur, dan menganalisis data untuk menghasilkan basis data semantik formal yang bersih. Bidang ringkasan informasi berawal dari pemrosesan bahasa alami, di mana dorongan utamanya berasal dari upaya untuk meringkas entitas nama seperti nama orang,

tempat, organisasi, dan waktu. Ketika masyarakat merasa lebih mudah untuk mengakses data terstruktur dan tidak terstruktur, hal itu menandai adanya aplikasi baru ringkasan terstruktur. Saat ini, ada minat untuk mengubah pengetahuan dalam publikasi ilmiah menjadi database struktural dan semantik, dan dalam memanfaatkan Internet untuk penyelidikan pencarian fakta semantik dan terstruktur. Saat ini, sejumlah besar data yang dikumpulkan dan dihasilkan

setiap hari menawarkan berbagai peluang analitis bagi organisasi untuk mengungkap informasi yang bermanfaat untuk operasinya [1]. Akibatnya, berbagai komunitas peneliti yang berbeda membawa teknik dari pembelajaran mesin, database, pengambilan informasi, dan linguistik komputasi untuk berbagai aspek masalah ringkas informasi. Karena teknik komputasi untuk memahami dan meringkas pengetahuan Al Quran, maka teknik pembelajaran ontologi digunakan untuk meringkas pengetahuan ini. Dalam lima belas tahun terakhir, para sarjana telah menulis buku-buku yang menyoroti berbagai linguistik, gaya bahasa, ilmiah, retorik, dan banyak penemuan pengetahuan tersembunyi di berbagai bidang lain dari Al Quran. Jelas, selama periode waktu ini, para ulama ini telah mengandalkan pengetahuan pribadi dan keakraban mereka dengan Al Quran karena tidak ada alat komputasi yang tersedia saat itu. Meskipun analisis komputasi Al Quran akan sangat bermanfaat dalam membantu umat Islam memahami Al Quran, namun hanya ada sedikit analisis yang dilakukan terhadapnya. Al Quran perlu disusun dengan beberapa cara khusus, yang melibatkan struktur bahasa yang dapat mengungkapkan makna yang berbeda di sepanjang zaman [2] karena Al Quran juga dikenal sebagai "pengetahuan yang hidup" (tentu saja mengandung pengetahuan tentang masa lalu, sekarang dan masa depan.)

Saat ini, teknik dari bidang teknologi semantik menawarkan pendekatan lain dalam mengatur pengetahuan Al Quran dalam informasi yang lebih bermakna. Dengan demikian, ontologi adalah kandidat yang baik untuk mewakili pengetahuan Al Quran yang kompleks dan juga dalam mendukung perluasan pengetahuan dalam Al Quran. Dengan demikian, pembangunan ontologi konten Al Quran akan memberikan konseptualisasi bersama dari beberapa domain yang mungkin dapat dikomunikasikan antara manusia dan sistem. Ini merupakan kosakata khusus yang digunakan untuk menggambarkan model dunia tertentu, di samping serangkaian asumsi eksplisit yang mempertimbangkan makna yang ditujukan dari kata-kata dalam

kosakata tersebut. Baik kosakata dan asumsi membantu orang dan sistem untuk memahami kesimpulan umum setelah berkomunikasi.

Penelitian ini bertujuan untuk menemukan dan mengusulkan aturan dan pola yang dapat digunakan dalam meringkas pengetahuan Al Quran yang berbasis teks terjemahan bahasa Inggris.

Penelitian ini telah memotivasi dan mudah-mudahan akan memungkinkan peneliti lain yang tertarik dengan Al Quran untuk sedekat mungkin dengan pemahaman makna yang dimaksudkan melalui penciptaan hierarki dan hubungan ontologis. Karena pengguna ontologi Al Quran sebagian besar adalah siswa dan peneliti Al Quran, maka fokus besar penelitian ini ditempatkan pada penciptaan ontologi secara keseluruhan, tetapi berfokus pada subjek tertentu yaitu ayat-ayat yang berhubungan dengan shalat

2. HAL TERBARU DARI TEORI YANG ADA

Al-Qur'an terdiri dari 77784 kata token dan 19287 jenis kata (kata-kata berbeda) [2], yang diurutkan dalam bab dan ayat. Umat Islam percaya bahwa format data asli diucapkan dalam bahasa Arab Klasik, yang kemudian ditangkap dengan tepat dalam sistem transkripsi yang canggih. Akses ke Al Quran secara tradisional melalui teks. Banyak Muslim belajar menghafal dan melafalkan kumpulan data kata demi kata. Akses ke pengetahuan, kebijaksanaan dan hukum yang mendasari membutuhkan interpretasi dan kesimpulan. Dunia saat ini tidak lepas dari peran data karena semua dibangun di atas sebuah fondasi data [3].

Banyak pengetahuan dikodekan melalui penggunaan kata-kata halus, tata bahasa, kiasan, tautan, dan referensi silang. Selama lebih dari seribu tahun, para sarjana telah berusaha untuk mengekstrak pengetahuan dan hukum dari teks, dan telah membangun tafsir atau kumpulan analisis, interpretasi, dan rantai inferensi yang jauh lebih besar [2], [3], [4]. Sarjana telah menulis banyak buku untuk stilistika, ilmiah, retorik dan banyak penemuan tersembunyi

dari Al-Qur'an di berbagai bidang lainnya. Tentunya, semua pengetahuan ini didasarkan pada latar belakang pengetahuan mereka tentang Al Quran dan pemahaman mereka tentang isi Al Quran. Al Quran dicirikan oleh kemampuannya untuk menyimpan informasi yang luas dalam sebuah ayat yang tidak terstruktur dan tersebar - namun terkait secara konseptual.

Ilmu Komputer dan Kecerdasan Buatan menyajikan kesempatan untuk menganalisis ulang data teks, untuk mengekstrak dan menangkap pengetahuan yang mendasarinya melalui representasi dan penalaran pengetahuan yang tepat, dan untuk memungkinkan inferensi dan kueri otomatis dan objektif [4].

Sebagian besar sistem yang dihasilkan komputer bersifat eksplisit dan merupakan representasi simbolis untuk pengetahuan terkait tentang domain tertentu. Pada prinsipnya, dapat digunakan kembali di seluruh sistem. Pada dasarnya, komputer adalah mesin yang memanipulasi simbol dan mereka membutuhkan instruksi yang jelas tentang bagaimana memanipulasi simbol dengan cara yang berarti. Ketika merepresentasikan pengetahuan secara simbolis dengan cara yang dapat diproses oleh komputer, muncul pertanyaan tentang simbol apa yang harus digunakan dan untuk apa mereka berdiri. Dengan demikian, ontologi sebagai representasi pengetahuan dipandang tepat untuk tujuan ini. Jenis ontologi ini mampu menyatakan apa yang penting bagi domain yang bersangkutan, dan juga menentukan hubungannya. Penelitian menyelidiki apakah ada efek signifikan dalam keakuratan model prediktif [5]. Dalam konteks sistem berbasis pengetahuan, ontologi yang mendasari pada dasarnya akan memberi tahu kita simbol mana yang diperlukan dan bagaimana mereka harus diinterpretasikan. Pada tingkat logis, interpretasi kemudian dapat dibatasi berdasarkan ontologi dengan melakukan aksiomatisasi simbol secara tepat [6]. Teknik yang paling menjanjikan dalam pembuatan ontologi adalah kombinasi dari pendekatan berbasis aturan menggunakan *natural*

language processing dan pendekatan pembelajaran mesin [7]. Certainty factor adalah metode untuk mengelola ketidakpastian dalam sistem berbasis aturan [8].

Ontologi idealnya terdiri dari kelas (konsep), relasi (subkelas kelas, baik taksonomi atau non-taksonomi), slot (fitur, atribut, peran atau properti), nilai batasan (segi, kardinalitas, jenis, ruang lingkup) dan contoh (individu, objek atau entitas) [9], [6]. Sistem rekomendasi diperlukan karena sebelumnya terdapat kelemahan pada sistem berbasis konten [10]. Selain itu, ontologi juga dapat diklasifikasikan berdasarkan struktur dan subjek konseptualisasi, yaitu isinya. Sowa [11] membaginya menjadi dua tipe, yaitu Tipe 1 dan Tipe 2. Tipe 1 adalah ontologi Terminologi / Leksikal, dimana sebuah ontologi yang konsep dan relasinya tidak sepenuhnya ditentukan oleh aksioma dan definisi yang menentukan kondisi perlu dan cukup dari penggunaannya. Secara komersial, dapat disebut sebagai kegiatan yang berusaha menciptakan transaksi yang panjang antara perusahaan dan individu

[12]. Konsep mungkin sebagian ditentukan oleh relasi seperti subtype / supertype atau sebagian / keseluruhan, yang menentukan posisi relatif konsep terhadap satu sama lain, tetapi tidak sepenuhnya mendefinisikannya. Tipe 2 adalah Axiomatized / Formal ontology di mana sebuah ontologi terminologi yang konsep dan relasinya memiliki aksioma dan definisi terkait yang dinyatakan dalam logika atau dalam beberapa bahasa berorientasi komputer yang dapat secara otomatis diterjemahkan ke dalam logika.

Di sisi lain, klasifikasi [13] Guarino adalah: (i), ontologi tingkat atas di mana ia menjelaskan konsep-konsep yang sangat umum yang tidak bergantung pada masalah atau domain tertentu. (ii), domain dan ontologi Tugas di mana mereka menggambarkan, masing-masing, kosakata (konseptualisasi) yang terkait dengan domain generik dengan menghususkan istilah yang diperkenalkan dalam ontologi tingkat atas. (iii) ontologi aplikasi dimana ontologi ini berisi semua definisi yang diperlukan untuk memodelkan pengetahuan yang dibutuhkan untuk aplikasi

tertentu dan Akhirnya, ontologi representasi yang menggambarkan konseptualisasi tentang formalisme representasi pengetahuan yang dikemukakan oleh [14].

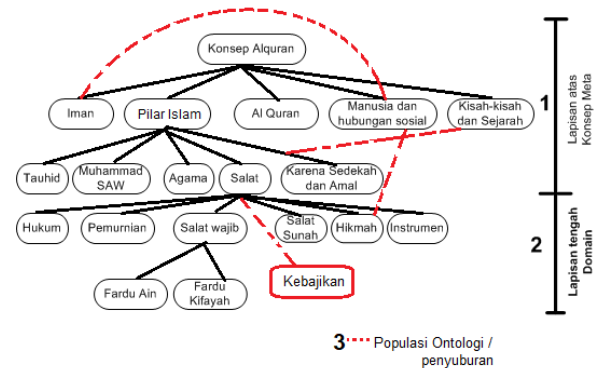
Saat ini, ketersediaan teknik komputasi untuk pemahaman pengetahuan Al Quran masih terbatas. [15] misalnya menggunakan Ringkas Entitas Nama dari Teks Al Quran untuk membentuk ontologi Al Quran, yang selanjutnya dapat diklasifikasikan sebagai tipe 1 dan (ii). [16] menyajikan desain dan implementasi model ontologis dan hasil aplikasinya terdapat pada kosakata “Kata benda waktu” dari Al Quran (tipe 2 dan (i), (ii) & (iii)). Qurany Explorer [17] adalah korpus Al Quran yang ditambah dengan ontologi atau indeks konsep-konsep kunci, dilakukan secara manual dan diambil dari 'Mushaf Al Tajweed' (Al Tajweed adalah nama lain untuk Al Quran). Ini berisi indeks hierarki atau ontologi yang komprehensif dari hampir 1200 konsep dalam Al Quran (tipe 1 dan (ii)). [18] menggunakan teknik klasifikasi otomatis dari ayat-ayat Al Quran berdasarkan surah tertentu bab) sesuai dengan klasifikasi yang dibuat oleh Ulama (tipe 1 dan (ii)). Namun demikian, dasar ontologi ilmu pengetahuan Islam dalam mengklasifikasikan topik-topik utama dalam Al Quran telah dibangun puluhan tahun sebelumnya.

Seluruh topik diklasifikasikan secara manual oleh para ahli domain berdasarkan pemahaman mereka tentang konteks Al Quran. Pengetahuan ini dapat digunakan sebagai pengetahuan umum sebagai pedoman dalam konstruksi ontologi otomatis komputer berbasis Ilmu Islam.

3. METODE

Dalam membangun ontologi yang relatif lengkap untuk ontologi Pengetahuan Islam, teknologi semantik dan teknik pemrosesan bahasa alami telah digabungkan dalam melakukan penciptaan ontologi (semi) otomatis untuk teks terjemahan Al Quran. Secara tentatif, konstruksi ontologi ini didasarkan pada tiga lapis pendekatan. Lapisan pertama adalah konsep Meta dari Al Quran (konsep super), yang didasarkan pada teks Indeks Al Quran yang diperkenalkan oleh [2]. Lapisan kedua adalah ontologi domain yang berfokus pada ontologi

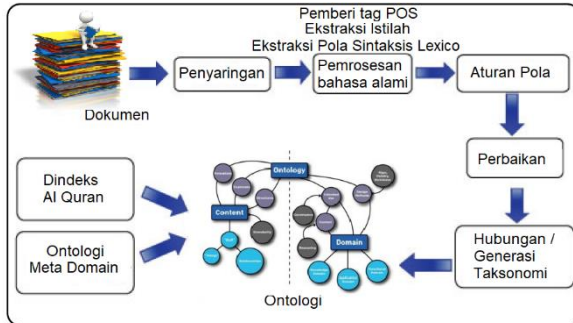
Salah (dirinci dalam [19]). Ini adalah proses semi-otomatis yang melibatkan insinyur ontologi dan pakar domain. Lapisan kedua akan menjadi lapisan bawah. Lapisan terakhir atau lapisan ketiga adalah populasi ontologi yang digunakan untuk memperkaya dan menjembatani ontologi yang dibangun pada lapisan pertama dan kedua (ini direpresentasikan sebagai garis putus-putus pada Gambar 1). Gambar 1 menunjukkan tingkatan konstruksi ontologi.



Gambar 3.1. Lapisan Ontologi Al Quran

Untuk membangun populasi ontologi yang didasarkan pada Teks Terjemahan Al Quran Inggris, langkah-langkah yang terlibat ditunjukkan pada Gambar 2. Penyaringan, yang merupakan proses untuk menyaring teks dan semua informasi dalam tanda kurung, dihapus (sementara untuk tujuan penguraian yang mungkin mengganggu teks fokus), serta mengubah semua kata ganti dengan huruf kapital seperti 'Kami', 'Kami', 'Dia', 'Tuhan', 'Saya' menjadi Allah dan 'Hai orang yang beriman' menjadi Mukminun.

Pada tahap ini, fokus hanya tertuju pada teks terjemahan tanpa melihat makna dan penjelasan lain tentang teks aslinya. Kemudian diikuti dengan proses *natural language processing* yang merupakan penandaan *part-of-speech*, penguraian sintaksis, dan istilah ringkas yang cocok dengan pola yang diidentifikasi. Langkah selanjutnya, fokus ditempatkan pada istilah dan informasi dalam tanda kurung untuk mengekstrak lebih banyak informasi dan untuk mengatasi ambiguitas dalam beberapa teks. Proses ini meneliti sinonim, definisi dan penjelasan istilah.



Gambar 3.2. Proses Ringkas Populasi Ontologi

Untuk mendeteksi hubungan taksonomi, pertimbangan dibuat berdasarkan yang didefinisikan oleh [20] dan yang didasarkan pada frase nomina yang diperkenalkan oleh [21]. Beberapa pengujian telah dilakukan untuk mengetahui pola mana yang berpotensi diperoleh. Selanjutnya, cara di mana kedua jenis pola dapat digabungkan dalam meningkatkan kinerja diusulkan. Selain itu, diusulkan pula pola yang dapat dijadikan sebagai relasi taksonomi.

Seperti disebutkan sebelumnya, sebagian besar peneliti saat ini akan menerapkan pola Hearst [20] dalam menentukan dan memiswahi hiponim dari korpus. [22] telah menggunakan pola Hearst dalam melakukan perbandingan dalam meringkas pola dengan menggunakan teks Terjemahan Al Quran Yusuf Ali. Penelitiannya didasarkan pada pola Hearst berikut:

- H1: <CONCEPT> seperti <INSTANCE>
- H2: seperti <CONCEPT> sebagai <INSTANCE>
- H3: <KONSEP>, (terutama | termasuk) <INSTANCE>
- H4: <INSTANCE> (dan | atau) <CONCEPT> lainnya

Dari pengamatan dan penelitiannya, pola Hearst [20] gagal menemukan hasil yang cocok pada ayat-ayat shalat pada teks Terjemahan Al Quran untuk pola H1, H2 dan H4 dan hanya 1 kecocokan untuk pola

H3. Kemudian, dia menggunakan versi pendekatan [6] yang disempurnakan menggunakan pola berikut:

- C1: <INSTANCE> <CONCEPT>
- C2: <CONCEPT> <INSTANCE>
- C3: <INSTANCE>, (a | lain) <CONCEPT>
- C4: <INSTANCE> adalah <CONCEPT>

Dari percobaan, ditemukan 6 kecocokan untuk pola C1 dan C2 yang

dikenal dengan kata benda majemuk. Untuk percobaan di ringkas yang sama dilakukan berdasarkan pola di atas, dengan sedikit modifikasi dilakukan pada pola itu sendiri. Percobaan menghasilkan hasil yang sama untuk H1, H2, H3 dan H4, seperti yang disebutkan dalam [22]. Sedangkan untuk apposition dan copula, aturannya diperluas dengan menambahkan elemen lain yang masih sesuai dengan definisi istilah tersebut (apposition (C3, menggunakan 'a' sebagai penghubung) dan copula (C4)) seperti pada [6] aturan kopula yang diperluas, di mana dia baru saja memfokuskan pada kata kunci 'is', sedangkan eksperimen ini menggunakan aturan berikut:

$\backslash w + \{BE(D)\} \backslash w \implies \text{adalah / are / was / were (verb-to-be)}$

Sedangkan untuk aposisi, aturan tambahan Cimiano menggunakan kata kunci 'a', tetapi dalam percobaan ini, aturan tersebut telah diubah dengan menambahkan pola berikut:

Berdasarkan definisi oleh [23], [24] dan uraian pola C1 dan C2 Cimiano [6], beberapa perbaikan pada pola telah dilakukan, yang juga didasarkan pada tentang definisi kata benda majemuk di [25]. Pola berikut membuat kombinasi dua atau lebih kata benda / kata sifat untuk membentuk satu kata benda, di mana ia dapat diringkas menjadi empat pola, di mana N1 adalah subkonsep dari N0 dan juga merupakan istilah tambahan yang berhubungan dengan kata benda:

- i. N1N0
- ii. N0'sN1
- iii. NP→JJ, NP0
- iv. NP→NP0 (PP NP *)
- v. Kata Benda di mana N0 adalah bagian-dari N1N0 / N0 dari N1

Misalnya, "doa pemakaman" adalah bagian dari "berdoa". Pola lainnya adalah gaya aposisi. Pola yang menangkap intuisi ini terlihat sebagai berikut:

APPOSITION: NP0 a / sebagai NP1

Aposisi juga dapat diturunkan dari dua elemen, biasanya frasa kata benda, yang ditempatkan berdampingan, seperti dt NP0 dt NP1 (dt = penentu).

Ini menunjukkan bahwa:

NP1 adalah-a NP0

Copula adalah pola lain yang bisa digunakan. Ini adalah kata yang digunakan untuk menghubungkan subjek kalimat dengan predikat, di mana predikat tersebut terdiri dari kata sifat dan kata benda, sehingga kata kerja penghubungnya adalah "ada". Ini mungkin cara paling eksplisit untuk mengungkapkan bahwa entitas tertentu adalah turunan dari konsep tertentu dengan kata kerja 'menjadi', di mana istilah umumnya adalah

COPULA: NP0 calon verba NP1

Ini menunjukkan bahwa:

NP1 adalah-a NP0 (NP0 dapat berupa NP atau Adv→NP; PP→NP).

Dari pengamatan terhadap standar emas yang telah ditetapkan oleh para ahli domain ditemukan bahwa terdapat pola yang dapat dibangun untuk membangun hubungan taksonomi. Berikut adalah pola-pola yang telah ditentukan.

Tiga pola yang digunakan dalam mengidentifikasi bagian-dari hubungan antara konsep-konsep tersebut dirujuk oleh dua istilah dalam teks. Untuk mengidentifikasi frasa yang mewakili konsep, selanjutnya P0, dan frasa yang mewakili subkonsep tersebut, selanjutnya Pn (di mana n = 1... .n), frasa tersebut dapat berupa VP atau NP. NP0 merepresentasikan superconcept. Teknik penguraian dangkal (Stanford Parser) digunakan berdasarkan pencocokan ekspresi reguler di atas tag bagian ucapan untuk mengidentifikasi frasa seperti yang dijelaskan di atas.

Pola 1: NP0 [(yaitu P1, P2...Pn)]
→ P1, P2...Pn PartOf NP0

Penggunaan tanda kurung siku []: Digunakan untuk membuat bagian kalimat teks lebih lengkap atau untuk menguraikan konsep secara mendetail. Jika kata 'ie' muncul, itu menunjukkan contoh atau bagian dari komponen ke NP yang disebutkan sebelumnya.

Pola 2: NP0 {:,!} (P1, dan) *, dan Pn

Simbol ':' atau '!' Setelah NP menunjukkan hubungan 'bagian dari' atau 'komponen' NP. Ringkas ini didasarkan pada analisis teks Al Quran, tentang penggunaan tanda seru.

Dimana P * PartOf NP0. Pembatas

didasarkan pada pola berulang yang sama seperti ", (koma)", "dan", kata kerja dalam frasa dan kombinasi pola tersebut. Banyak ayat Al Quran juga menerapkan pola ini.

Ini menunjukkan itu

Jika setelah "kurung kuadrat terbuka ["
yaitu

jika NP0 [P1] * lalu

P1, P2... Pn PartOf NP0

Operator koma dan konjungsi akan menjadi pembatas pernyataan tersebut. Pola-pola ini dijelaskan secara rinci di [19]. Tabel 1 menunjukkan ringkasan dari ringkas pola untuk hubungan 'is-a' dan 'part-of'.

4. HASIL

Untuk perbandingan kinerja ringkas, telah dilakukan perbandingan manual antara teknik pendekatan yang diperoleh dan perbaikannya dengan standar emas. Pendekatan telah diklasifikasikan berdasarkan [6], yang dijelaskan dan didefinisikan dalam Tabel 1, dengan fokus pada pola H1, H2, H3 dan H4. [6] Pola didasarkan pada C1, C2, C3 dan C4. Cimiano ++ adalah pola perluasan yang didasarkan pada [6], tetapi dengan modifikasi pada pola untuk meningkatkan jumlah ringkas dalam definisi yang dijelaskan. Qpattern sebenarnya adalah pola baru yang ditemukan berdasarkan Teks Terjemahan Al Quran, yang cocok sebagai relasi taksonomi. Sifat hubungan taksonomi pada tiap pola terlihat kinerja taksonomi yang diringkas untuk masing-masing pola cukup baik, kecuali untuk pola Hearst dimana temuan ini sejalan dengan [6] yang menyebutkan bahwa pola ini jarang muncul dan sebagian besar kata kunci yang terkait melalui relasi 'is-a' tidak muncul dalam pola gaya Hearst.

Diperoleh persentase ringkas taksonomi berdasarkan pendekatan dan peningkatan saat ini. Jumlah total relasi taksonomi sebesar 130, Hearst true positif 1 dengan 0,8%, false positif 0%, false negatif 0%; Cimiano true positif 55 dengan 43,2%, false positif 14 dengan 10,8% , false negatif 3,8%; Cimiano++ true positif 56 dengan 43,1%, false positif 10 dengan 7,7% , false negatif 13 dengan 10%; QPattern true positif 18 dengan 13,8%, false positif 2 dengan 1,5% , false negatif 3 dengan 2,3%;

Kamus yang dapat dibaca mesin adalah yang terbaik dalam menerapkan pola ini [6]. [26] juga menyebutkan bahwa pola Hearst jarang terjadi, dan cakupan pola ini rendah bahkan dalam koleksi dokumen yang sangat besar. Meskipun jumlah taksonomi yang diringkas dengan pola Cimiano menunjukkan hasil yang cukup baik dengan 0,42% true positif dari keseluruhan hubungan taksonomi, namun masih ada ruang untuk perbaikan. Penggunaan kata benda majemuk (C1 dan C2) oleh [6], memberikan kontribusi terbesar pada ringkas taksonomi atau hierarki konsep.

Peningkatan 0,57% true positif pada Cimiano ++ dan QPattern menunjukkan bahwa, amandemen harus dilakukan berdasarkan proses ringkas seperti yang diilustrasikan oleh para ahli domain. Hal ini juga sejalan dengan saran yang dibuat oleh [6], dimana ringkas tidak hanya berfokus pada Noun dalam suatu pola, tetapi juga untuk menempatkan pertimbangan pada frase nomina yang lebih mendekati itu sendiri. Misalnya, "daerah bersejarah utama Ordino adalah desa yang menawan", jika didasarkan pada pola leksiko-sintaksis akan diturunkan sebagai: is-a (Ordino, desa), di mana itu pasti salah; ringkas yang tepat harus is-a (kawasan bersejarah utama Ordino, desa menawan). Hal tersebut menunjukkan bahwa perbaikan harus dilakukan dengan mempertimbangkan NP yang lebih dekat yang melibatkan NP→DET NP; Kata benda majemuk; N * (Kata sifat N *) *; N * (PP N *) *, dalam meringkas pola pada taksonomi atau hirarki konsep.

Deskripsi Logika Representasi Pengetahuan untuk Ringkas Ayat-ayat Shalat. Dalam proses pemetaan dengan ringkas pengetahuan berdasarkan pola yang telah ditentukan sebelumnya, dilakukan beberapa modifikasi. Beberapa aturan perlu ditambahkan untuk memenuhi persyaratan definisi konsep. Ini melibatkan seluruh konsep kandidat, yang melibatkan "orang yang" dan "orang yang" yang berarti "orang". Jika ada verba yang terlibat dalam frase tersebut, maka verba tersebut menjadi relasi antar konsep. Misalnya konsep AlBirr.

Ada beberapa pengertian yang berbeda untuk konsep ini, yaitu "AlBirr adalah AsSabirin" dan juga "Muttaqun". Pada saat yang sama, makna lainnya adalah AlBirr adalah orang yang

beriman kepada Allah, Hari Akhir, Malaikat, Kitab, Nabi dan memberikan hartanya, meskipun cinta untuk itu, kepada kerabat, kepada anak yatim, dan Al-Masakin (orang miskin), dan kepada musafir, dan kepada mereka yang meminta, dan untuk membebaskan budak, melakukan As-Salat (Iqamat-as-Salat), dan memberikan Zakat, dan siapa yang memenuhi mereka ketika mereka membuat (AlBaqarah, 177). Gambar 3 menjelaskan

aturan yang dihasilkan untuk mewakili ilmu yang diekstrak dari ayat-ayat salat.

5. DISKUSI

Tidak diragukan lagi bahwa beberapa kelemahan mungkin terjadi dalam ringkas pola. Ada tiga poin utama yang diidentifikasi mengalami masalah untuk mencapai presisi dan penarikan 100%.

Analisis anaphora (berdasarkan referensi bersama) diperlukan untuk meringkas informasi tentang konteks dan hubungan dalam ayat-ayat Al Quran. Ini karena dalam gold standard, para ahli domain akan mengekstrak informasi dan pola berdasarkan pengetahuan mereka tentang Al Quran. Misalnya, berdasarkan Gambar 4 di bawah ini, bagian terakhir dari ayat tersebut menyebutkan bahwa "... mereka akan mendapatkan pahala mereka dengan Tuhan mereka". 'Mereka' mengacu pada siapa?

Menurut para ahli, jawabannya adalah "Mukminun mendapat pahala"; "mereka yang melakukan tindakan benar mendapatkan pahala"; "mereka yang melakukan shalat mendapatkan pahala"; "mereka yang memberi sedekah mendapatkan pahala" (didefinisikan dalam standar emas oleh para ahli domain). Namun, analisis referensi bersama tidak tercakup dalam penelitian ini.

Sungguh mereka yang beriman, dan melakukan perbuatan benar, dan melakukan shalat (iqamat-as-salat), dan memberi zakat, mereka akan mendapatkan pahala mereka dengan Tuhan mereka. Pada mereka tidak akan ada rasa takut, mereka juga tidak akan berduka.



Gambar 5.1. Surat Al-Baqarah ayat 277

ii. Alasan lainnya adalah 5% dari jawaban yang diekstrak sering kali disebabkan oleh kesalahan pada saat penguraian seperti ".. dan melaksanakan As-Salat (Iqamat-as-Salat) .."

Kesalahan dalam penguraian ini akan meningkatkan hasil dalam pola ringkas 'kata benda majemuk'. Kesalahan lainnya adalah karena masalah dalam mengidentifikasi kepala kata benda majemuk. Menurut [24]: Artinya salah satu kata yang membentuk senyawa tersebut dominan secara sintaksis. Dalam bahasa Inggris kepala biasanya adalah benda di sebelah kanan kata majemuk. Sifat sintaksis head diteruskan ke seluruh senyawa. Jadi, jika kita memiliki kata majemuk seperti kursi malas yang terdiri dari kata sifat easy dan kata benda chair, secara sintaksis seluruh kata adalah kata benda. "

Kebanyakan kata benda majemuk bahasa Inggris adalah frase kata benda (frase nominal) yang menyertakan kata benda yang dimodifikasi oleh kata sifat atau kata benda atributif. Biasanya kepala diidentifikasi sebagai berikut (lihat Tabel 1):

Tabel 5.1. Kepala Kata Benda Senyawa

Pengubah	Kepala	Contoh
kata benda	kata benda	Kalender Islam
kata sifat	kata benda	Jalan yang benar
kata kerja	kata benda	pemecah gelombang
preposisi	kata benda	neraka
kata benda	kata sifat	Muhammad damai

Namun, cukup sulit untuk mengidentifikasi kata benda atributif menggunakan parser. Misalnya, dalam frase "Messenger Muhammad", kata benda tambahannya "Messenger" memodifikasi kata benda "Muhammad". Seharusnya, hasilnya harus mendapatkan Utusan sebagai Kepala dan Muhammad sebagai pengubah (isA (Utusan Muhammad, Utusan) dan bukan A (Utusan Muhammad, Muhammad)). Dalam penelitian ini, fokus ditempatkan hanya pada pola ringkas "kata

benda, kata benda" dan "kata sifat, kata benda".

6. KESIMPULAN

Karya ini menyajikan pendekatan generasi otomatis konsep / contoh ontologi dan hubungan dari kumpulan dokumen tidak terstruktur yang dikenal sebagai Al Quran. Pendekatan yang disajikan dirangsang berdasarkan kombinasi teknik pemrosesan bahasa alami, Ringkas Informasi (IE) dan teknik Penambangan Teks. Berdasarkan sistem IE tradisional, penulis telah menerapkan dan mendefinisikan aturan tata bahasa dan ringkas untuk mendapatkan ontologi dari konsep / contoh dan relasi.

Sistem kemudian mencoba untuk membentuk konsep / kejadian parsial yang benar dan relasi dengan mengambil kata dan entitas dari teks dan menggabungkannya untuk membentuk yang benar dan lengkap. Di sini, pola baru disajikan untuk ringkas teks Terjemahan Al Quran, dengan tujuan secara otomatis mendorong konsep / contoh dan ringkas relasi dari teks dokumen. Pendekatan tersebut dievaluasi dengan membandingkan konsep dan relasi yang dihasilkan dengan standar emas yang telah diidentifikasi oleh para ahli domain.

Kontribusi dan kesimpulan dapat diringkas sebagai berikut, telah menghasilkan pola / aturan baru yang sesuai untuk Teks Al Quran, yang didasarkan pada pola dasar yang diperkenalkan oleh peneliti sebelumnya. Dari penelitian ini, juga mengidentifikasi kesulitan-kesulitannya dalam meringkas informasi yang melibatkan referensi bersama dan gaya Al Quran, yang memiliki pergeseran tata bahasa dan metonimi.

DAFTAR PUSTAKA

- [1] N. I. Munawar, Zen and Putri, "Keamanan Jaringan Komputer Pada Era Big Data," *J-SIKA/ J. Sist. Inf. Karya Anak Bangsa*, vol. 02, no. 01, pp. 14–20, 2020.
- [2] N. Abbas, "Quran ' Search for a Concept ' Tool and Website Quran ' Search for a

- Concept ' Tool and Website," The University of Leeds, 2017.
- [3] Z. Munawar, B. Siswoyo, and N. S. Herman, "Machine learning approach for analysis of social media," *ADRI Int. Journal. Information. Technol.*, vol. 1, pp. 5–8, 2017.
- [4] E. Atwell *et al.*, "Understanding the Quran: A New Grand Challenge for Computer Science and Artificial Intelligence," *Gd. Challenges Comput. Res. Br. Comput. Soc. Work.*, vol. 1, no. 11, pp. 1829–1841, 2010.
- [5] Z. Munawar, "Penggunaan Profil Media Sosial Untuk Memprediksi Kepribadian," *Temat. - J. Teknol. Inf. dan Komun.*, vol. 4, no. 2 SE-Articles, Dec. 2017.
- [6] P. Cimiano, "Ontology Learning and Population from Text: Algorithms," *Eval. Appl. Springer*, vol. November, 2006.
- [7] D. Maynard, Y. Li, and W. Peters, "NLP techniques for term extraction and ontology population," *Front. Artif. Intell. Appl.*, vol. 167, no. 1, pp. 107–127, 2008.
- [8] N. I. Putri, "Sistem pakar diagnosa tingkat kecanduan gadget pada remaja menggunakan metode Certainty Factor." UIN Sunan Gunung Djati Bandung, 2018.
- [9] P. Buitelaar, P. Cimiano, and B. Magnini, "Ontology Learning from Text: Methods, Evaluation and Applications," *Comput. Linguist.*, vol. 32, no. 4, pp. 569–572, 2006.
- [10] Z. Munawar, N. Suryana, Z. B. Sa'aya, and Y. Herdiana, "Framework With An Approach To The User As An Evaluation For The Recommender Systems," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 2020, pp. 1–5.
- [11] J. Sowa, "Knowledge Representation: Logical, Philosophical, and Computational Foundations," *Brooks Cole Publ. Co*, vol. 5, no. 1, pp. 9.1-9.3, 2004.
- [12] Z. Munawar, "Keamanan Pada E-Commerce Usaha Kecil dan Menengah," *Temat. - J. Teknol. Inf. dan Komun.*, vol. 5, no. 1 SE-Articles, Jun. 2018.
- [13] N. Guarino, "Formal Ontology and Information Systems," *Form. Ontol. Inf. Syst. Proc. 1st Int. Conf.*, vol. 46, no. June, pp. 3–15, 1998.
- [14] G. Van Heijst, A. T. Schreiber, and B. J. Wielinga, "Using explicit ontologies in KBS development," *Int. J. Hum. Comput. Stud.*, vol. 46, no. 2–3, pp. 183–292, 1997.
- [15] K. Dukes, "Ontology of Quranic Concepts," 2020. [Online]. Available: <https://corpus.quran.com/ontology.jsp>. [Accessed: 01-Jan-2021].
- [16] M. Al-Yahya, H. Al-Khalifa, A. Bahanshal, I. Al-Odah, and N. Al-Helwah, "An ontological model for representing semantic lexicons: An application on time nouns in the holy quran," *Arab. J. Sci. Eng.*, vol. 35, no. 2 C, pp. 21–35, 2010.
- [17] N. Abbas and E. Atwell, "Qurany Explorer." [Online]. Available: <http://quranytopics.appspot.com/>. [Accessed: 02-Jan-2021].
- [18] M. N. Al-Kabi, G. Kanaan, R. A.- Shalabi, K. M. O. Nahar, and B. M. B.- Ismail, "Statistical Classifier of the Holy Quran Verses (Fatiha and Yaseen Chapters)," *J. Appl. Sci.*, vol. 5, no. 3, pp. 580–583, 2005.
- [19] S. Saad, N. Salim, and H. Zainal, "Towards Context-Sensitive Domain of Islamic Knowledge Ontology Extraction," *Int. J. Infonomics*, vol. 3, no. 1, pp. 197–206, 2010.
- [20] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," in *COLING '92: Proceedings of the 14th conference on Computational linguistics*, 1992, vol. 2, no. August, pp. 539–545.
- [21] G. Grefenstette, "Short query linguistic expansion techniques: Palliating one-word queries by providing intermediate structure to text," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1299, no. May, pp. 98–114, 1997.
- [22] C. Y. Yong, R. Sudirman, K. M. Chew, and N. Salim, "Comparison of ontology learning techniques for Qur'anic text," in *Proceedings - 2011 International Conference on Future Computer Sciences and Application, ICFCSA 2011*, 2011, pp. 192–196.
- [23] R. Carter and M. Mccarthy, *Cambridge Grammar of English*. Cambridge: Cambridge university press, 2006.

- [24] F. Katamba, *English Words Structure, History, Usage*, 2nd Editio. Routledge, 2005.
- [25] R. Nordquist, "About.com - Grammar & Composition," 2010. [Online]. Available: <http://grammar.about.com/od/c/g/compnounterm.htm>. [Accessed: 01-Jan-2020].
- [26] M. Brunzel and M. Spiliopoulou, "Discovering groups of sibling terms from Web documents with XTREEM-SG," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5383 LNCS, no. January 2008, pp. 126–155, 2008.