

Untuk mengatasi kesulitan tersebut, dibutuhkan suatu mekanisme yang memungkinkan komputer memahami makna informasi yang dicari. Dengan kata lain, dibutuhkan suatu cara agar informasi dalam suatu dokumen Web dapat dibaca dan dipahami oleh mesin (*machine understandable*). Web dengan kemampuan demikian, seolah-olah memiliki kecerdasan yang sanggup memberikan jawaban yang tepat terhadap pertanyaan atau kebutuhan para penggunanya.

2. TINJAUAN PUSTAKA

Sebuah sistem QA, menerima *query* dalam bentuk pertanyaan dengan bahasa alami, mencari jawaban pada sekumpulan dokumen atau pada sebuah basis pengetahuan, mengekstraknya dan kemudian memformulasikan jawaban yang ringkas (Moldovan & Surdeanu, 2003). Umumnya sistem QA terdiri atas tiga modul utama, yakni *question processing*, *document retrieval* dan *answer processing*. Kebanyakan sistem QA mengelompokkan pertanyaan berdasarkan jenis pertanyaannya (Cooper & Ruger, 2000; Moldovan & Surdeanu, 2003; Perez-Coutino *et al*, 2004; Gunawan & Lovina, 2006; Wijono *et al*, 2006; August, 2007; Kangavari *et al*, 2008). Jika jenis pertanyaan dapat ditentukan maka jenis jawabannya dapat ditentukan pula. Dimisalkan, jenis pertanyaannya adalah "Siapa..." , maka jawaban yang diinginkan adalah orang atau organisasi. Jika pertanyaannya "Kapan..." jawaban yang diinginkan adalah waktu atau tanggal.

Web dengan milyaran informasi yang sangat beragam dan tak terstruktur dipandang sebagai sumber informasi yang bernilai. Walaupun saat ini tersedia banyak mesin pencari, namun mereka tidak mampu memberikan informasi yang spesifik yang diinginkan pengguna. Pemanfaatan teknologi QA pada web bertujuan untuk mengatasi masalah tersebut. Teknologi QA diharapkan dapat menjadi antarmuka yang lebih intuitif untuk memformulasikan pertanyaan dan memberikan jawaban dalam bahasa alami daripada mengembalikan sekumpulan dokumen web yang terurut berdasarkan ranking (Moldovan & Surdeanu, 2003; Perez-Coutino *et al*, 2004; McGuinness, 2004; Lopez *et al*, 2005).

Penelitian-pelitan yang terkait dengan sistem QA pada SW telah banyak dilakukan. Katz *et al* (2002) menyebutkan bahwa terdapat peluang sinerjik antara teknologi bahasa alami dan SW, yakni sebuah sistem QA yang mampu memberikan informasi yang relevan dari sebuah basis pengetahuan berbasis ontologi dalam menanggapi *query* yang berikan oleh pengguna dalam bahasa alami.

Ide ini diwujudkan dengan mengadopsi *triple-based data model* (misalnya RDF) sebagai basis pengetahuan pada sistem QA (Katz *et al*, 2002; Lopez *et al*, 2005; Lopez, *et al*, 2006; Litkowski, 2003). Hal ini didasarkan pada pertimbangan bahwa terdapat kemungkinan untuk merepresentasikan sebuah *query* berbasis bahasa alami ke dalam bentuk *triple*, yang dalam hal ini berbentuk subyek, predikat dan obyek dari sebuah kalimat. Sementara pemodelan data dalam SW dengan menggunakan RDF (*Resource Description Framework*) juga menyatakan sebuah *statement* dalam bentuk *triple*: *resources*, *properties*, dan *value*.

Untuk mentransformasikan pertanyaan bahasa alami ke sebuah bentuk *query* formal digunakan metoda-metoda yang diadopsi dari teknologi *Natural Language Processing* (NLP), *Information Retrieval* (IR) dan *Information Extraction* (IE). Beberapa metoda yang sering digunakan adalah *named-entity recognition* dan *entity relation recognition*. Dalam kaitannya dengan representasi pengetahuan dalam sebuah ontologi, *named-entity* dapat dipandang sebagai sebuah *instance* atau kelas atau *value* dari sebuah properti dan *entity relation* dapat dipandang sebagai sebuah properti.

Kecenderungan penelitian-pelitan QA yang dilakukan saat ini mengarah pada *open domain* QA yang berbasis pada sejumlah besar dokumen pada web. Berbeda dengan kecenderungan tersebut, beberapa penelitian berfokus pada *restricted domain* (Lopez *et al*, 2005; Frank *et al*, 2004; Litkowski, 2003; Gunawan & Lovina, 2006; August, 2007; Cooper & Ruger, 2000; Kangavari *et al*, 2008). Pemilihan *restricted domain* didasarkan pada beberapa alasan, antara lain, *pertama*, eksploitasi informasi pada dokumen web sering dihadapkan

pada masalah reliabilitas informasi tersebut. Dapat saja terjadi bahwa informasi yang diberikan telah kedaluwarsa atau bahkan sepenuhnya salah. *Kedua*, pemanfaatan pengetahuan formal pada *restricted domain* dapat meningkatkan keakuratan sistem QA, karena baik pertanyaan maupun jawabannya dianalisis berdasarkan basis pengetahuan tersebut. *Ketiga*, sangat dimungkinkan bahwa sebuah institusi memiliki dan mengelola basis pengetahuan yang sifatnya terbatas dan hanya dipergunakan dalam lingkup institusi tersebut.

McGuinness (2004) menyebutkan bahwa penggunaan teknologi SW dapat meningkatkan kinerja sebuah sistem QA. Hal itu dapat dilakukan dengan cara memanipulasi konten (basis pengetahuan), memanipulasi *query* atau memanipulasi jawaban. Pada umumnya sistem QA pada web, mengekstrak jawaban dari sekumpulan dokumen yang tidak terstruktur. Pada *restricted domain*, penggunaan basis pengetahuan yang terstruktur sangat dimungkinkan karena ukuran basis pengetahuannya yang cenderung lebih kecil dan stabil (Frank *et al*, 2004) dibandingkan dengan basis pengetahuan pada *open domain*. Dengan basis pengetahuan yang terstruktur (misalnya ontologi), sistem dapat menurunkan lebih banyak makna dan dapat memanfaatkan *domain* dan *range* pada *slot* untuk mengecek konsistensi informasi (McGuinness, 2004).

Sejauh ini terdapat sejumlah penelitian mengenai sistem QA yang menggunakan bahasa Indonesia (Larasati & Manurung, 2007; August, 2007; Mahendra *et al*, 2008). Sebagai bahasa kenegaraan yang resmi, bahasa Indonesia digunakan oleh lebih dari seratus juta orang. Berdasarkan fakta tersebut, penggunaan bahasa Indonesia sebagai bahasa alami dalam sebuah sistem QA patut dipertimbangkan.

3. CARA PENELITIAN

3.1 Analisis dan Rancangan Ontologi

Ontologi film dibangun dengan menggunakan metoda yang dikembangkan oleh Noy dan McGuinness (2001). Langkah pertama yang dilakukan adalah menentukan batasan domain dari ontologi film yang akan dibangun. Sistem QA sederhana yang akan dibangun ini termasuk dalam

kelompok RDQA dengan domainnya adalah informasi film. Yang dimaksudkan dengan informasi film adalah atribut-atribut yang terkait dengan sebuah film, misalnya judul film, durasi, sutradara, aktor dan aktris, dan sebagainya.. Lampiran 1 memperlihatkan hasil dari rancangan ontologi film.

Analisis Pola Kalimat Pertanyaan

Terdapat enam pola kalimat pertanyaan yang dapat diproses oleh sistem, yakni:

1. Pola I: <KT><DP1><OP><DP2><VP>
2. Pola II: <KT><DP1><OP><VP>
3. Pola III: <KT><OP><DP2><VP>
4. Pola IV: <KT><OP><VP>
5. Pola V: <KT><DP1><DP2><VP>
6. Pola VI: <KT><DP1><VP>

<KT> adalah kata tanya yang diijinkan. Terdapat enam kata tanya yang diijinkan yakni: *siapa*, *apakah*, *kapan*, *dimana*, *bagaimana* dan *berapa*. <OP> adalah kata kunci tertentu yang mewakili *object property*. Kata kunci <OP> yang diijinkan tergantung pada kata tanya yang diikutinya. Misalnya kata tanya *siapa* hanya boleh diikuti oleh kata kunci *aktor*, *aktris*, *sutradara* dan sebagainya, tetapi tidak boleh diikuti oleh kata kunci *judul*, *genre*, *penghargaan* dan sebagainya. <DP> merupakan kata kunci yang mewakili *datatype property* tertentu dalam ontologi film. Sedangkan <VP> merupakan nilai dari sebuah *datatype properti*<DP>.

3.2 Analisis dan Rancangan Proses Sistem

Sistem QA yang akan dibangun terdiri atas dua modul utama, yakni modul *Question Processing* (QP) dan modul *Query Retrieval* (QR). Modul QP bertugas menentukan validitas kalimat dan kata kunci berdasarkan pertanyaan yang dimasukkan pengguna. Modul QR bertugas menentukan variabel *query*, *statementquery* dan melakukan *query* jawaban ke ontologi serta memberikan jawaban kepada pengguna. Lampiran 2 memperlihatkan Diagram Alir Data (DAD) Tingkat 1 untuk sistem QA

3.3 Analisis dan Rancangan Modul QP

Modul QP melakukan dua tugas utama yakni menentukan validitas kalimat pertanyaan, dan

menentukan kata kunci yang terdapat dalam kalimat pertanyaan yang dimasukkan pengguna. Untuk menentukan validitas pertanyaan, hal pertama yang dilakukan adalah menghilangkan semua kata-kata yang termasuk dalam *stopwords* dari kalimat pertanyaan. Setelah itu kalimat pertanyaan akan dibagi (*parsing*) dalam satuan-satuan kata. Kemudian akan dilakukan pengecekan dari kata pertama. Kata pertama harus berupa kata tanya. Kata tanya yang dapat dikenali oleh sistem adalah kata: *siapa, apakah, bagaimana, kapan, dimana* dan *berapa*. Kata berikutnya ditentukan berdasarkan kata tanya yang telah ditemukan. Lampiran 3 merupakan *flowchart* untuk menentukan validitas kalimat pertanyaan.

Setelah kalimat pertanyaan dinyatakan valid, langkah selanjutnya adalah menentukan kata-kata kunci, termasuk kata kunci tambahan. Kata-kata kunci tersebut akan digunakan pada modul QR untuk menentukan variabel *query* dan *statement-statementquery*. Lampiran 4 merupakan *flowchart* untuk menambahkan kata kunci yang bersesuaian.

3.4 Analisis dan Rancangan Modul QR

Modul QR melakukan tiga tugas utama yakni, menentukan variabel *query*, menentukan *statement-statement query* dan melakukan *query* jawaban. Variabel *query* berfungsi untuk menampung jawaban pertanyaan yang diperoleh dari ontologi untuk kemudian diteruskan ke pengguna. Variabel *query* ditentukan berdasarkan kata kunci yang mewakili *datatype property*.

Statement-statement query ditentukan berdasarkan kata-kata kunci yang diperoleh dari modul QP. Sebuah *statement* terdiri atas tiga bagian yakni *subject*, *predicate* dan *object* (S, P, O). Setiap kata kunci mewakili properti tertentu dalam ontologi. Properti mewakili *predicate* sebuah *statement*, sedangkan *domain* dan *range* dari properti tersebut mewakili *subject* dan *object* sebuah *statement*. Lampiran 5 memperlihatkan daftar kata kunci dan *statement* yang bersesuaian.

4. HASIL DAN PEMBAHASAN

Ontologi diimplementasikan dalam bahasa OWL DL. Untuk membangun OWL digunakan perangkat lunak *Protege-OWL editor*. Sistem QA diimplementasikan dengan menggunakan JSP sebagai antarmuka sistem dengan pengguna, dan JENA sebagai antarmuka sistem dengan ontologi. Implementasi sistem meliputi implementasi antarmuka pengguna, antarmuka ontologi dan implementasi modul QP dan modul QR.

Pertanyaan-pertanyaan yang digunakan untuk pengujian sistem adalah pertanyaan-pertanyaan faktual yang telah dikumpulkan sebelumnya melalui email. Evaluasi dilakukan dengan melihat apakah sistem mampu menentukan validitas kalimat secara tepat, dan apakah kalimat yang valid mampu diproses oleh sistem secara tepat.

Hasil pengujian menunjukkan bahwa sistem berhasil menentukan validitas kalimat pertanyaan yang dimasukkan pengguna sebesar 96,7%. Kegagalan sistem terjadi saat menentukan validitas kalimat pertanyaan nomor 14 dan 15. Pertanyaan-pertanyaan tersebut dianggap valid oleh sistem.

Pada pertanyaan nomor 14: *Siapa aktor yang berperan sebagai Jack Dawson dalam film Titanic?*, sistem menemukan kata kunci *nama* dan *aktor*. Sehingga pola yang ditemukan adalah “KTOP”. Pola ini dianggap valid oleh sistem. Demikian pula pada pertanyaan nomor 15: *Siapa nama pemeran james bond di film James Bond Casino Royale?*, sistem menemukan kata kunci *nama* dan *pemeran*. Jawaban yang diberikan adalah seluruh nama pemeran James Bond. Hal ini terjadi karena sisa kalimat pertanyaan pada bagian setelah kata kunci tokoh ditemukan dianggap sebagai nilai dari sebuah *datatype property* (VP). Permasalahan tersebut dapat diatasi dengan menambahkan pola-pola kalimat yang mampu dikenali sistem.

Kegagalan menentukan pola kalimat menyebabkan sistem memberikan jawaban salah. Secara keseluruhan, apabila kalimat tersebut dinyatakan valid, modul QP dan QR mampu memproses pertanyaan dan jawaban dengan baik.

5. KESIMPULAN

Penelitian ini bertujuan untuk membangun sebuah sistem QA sederhana berbasis ontologi sebagai sebuah aplikasi SW. Hasil penelitian menunjukkan bahwa pemanfaatan sistem QA sederhana berbasis ontologi mampu memroses pertanyaan-pertanyaan faktual dan tunggal secara signifikan dengan tingkat keberhasilan mencapai 96,7%. Hasil ini menunjukkan bahwa perpaduan teknologi QA dan teknologi SW dapat menjadi sebuah pilihan baru dalam membangun aplikasi pencarian berbasis web pada domain yang terbatas.

6. DAFTAR PUSTAKA

- [1]. August, S., 2006, Perancangan dan Penerapan Question Answering System pada Alkitab Digital Berbahasa Indonesia dengan Menggunakan Natural Language Processing, *Skripsi*, Fakultas Teknik Industri, Universitas Kristen Petra, Surabaya.
- [2]. Cooper, R.J., dan Ruger, S.M., 2000, A Simple Question Answering System. *Proceedings of the 9th Text REtrieval Conference*.
- [3]. Frank, A., Krieger, H., Xu, F., Uszkoreit, H., Cysmann, B., Jorg, B., Schafer, U., 2005, Querying Structured Knowledge Sources. *Proceedings of AAAI-05 and Workshop on Question Answering in Restricted Domains*.
- [4]. Gunawan dan Lovina, G., 2006, Question Answering System dan Penerapannya pada Alkitab. *Jurnal Informatika*. No. 1, Vol 7, hal 1-9.
- [5]. Kangavari, M.R., Ghandchi, S., dan Golpour, M., 2008, A New Model for Question Answering Systems. *Proceedings of World Academy of Science, Engineering and Technology*, Vol 32. Agustus 2008
- [6]. Katz, B., Lin, J., dan Quan, D., 2002, Natural Language Annotations for the Semantic Web. *Proceedings of the International Conferences on Ontology, Databases, and Applications of Semantics (ODBASE 2002)*, Oktober 2002.
- [7]. Larasati, S.D. dan Manurung, R., 2007, Towards a Semantic Analysis of Bahasa Indonesia for Question Answering. *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics (PACLING 2007)*.
- [8]. Litkowski, K.C., 2003, Question Answering Using XML-Tagged Documents. *Proceedings of the 11th TREC*.
- [9]. Lopez, V., Pasin, M., dan Motta, E., 2005, AquaLog: An Ontology-Portable Question Answering for the Semantic Web. *Proceedings of the 2nd European Semantic Web Conference*.
- [10]. Lopez, V., Motta, E., dan Uren, V., 2006, PowerAqua: Fishing the Semantic Web. *Proceedings of European Semantic Web Conference 2006*.
- [11]. Mahendra, R., Larasati, S.D., dan Manurung, R., 2008, Extending an Indonesian Semantic Analysis-based Question Answering System with Based Linguistic and World Knowledge Axioms. *The 22nd Pacific Asia Conferences on Language Information and Computation (PACLIC22)*.
- [12]. McGuinness, D. L., 2004, Question Answering on the Semantic Web. *IEEE Intelligent Systems* .No. 1, Vol 19, hal 82-85.
- [13]. Moldovan, D. dan Surdeanu, M., 2003, *On The Role of Information Retrieval dan Information Extraction in Question Answering Systems*, Pazienza, M.T., *Information Extraction in the Web Era*. LNAI 2700. Springer, Berlin.
- [14]. Noy, N.F. dan McGuinness, D.L., 2001, *Ontology Development 101: A Guide to Creating Your First Ontology*, http://protege.stanford.edu/publications/ontology_development/ontology101.pdf, diakses tanggal 12/10/2008.
- [15]. Perez-Coutino, M., Solorio, T., Montes-y-Gomez, M., Lopez-Lopez, A., Villasenor-Pineda, L., 2004, *Toward A Document Model for Question Answering Systems*. *Advances in Web Intelligence*. LNCS 3034. Springer, Berlin