

SEBUAH SURVEY KECIL PADA TANTANGAN DAN ARAH PENGEMBANGAN *QUESTION ANSWERING SYSTEM*

Hernawan Sulistyanto¹

¹ Jurusan Teknik Informatika, Fakultas Kominfo, Universitas Muhammadiyah Surakarta
Jl. A Yani Tromol Pos I Pabelan, Surakarta.
*Email: Hernawan.Sulistyanto@ums.ac.id

Abstrak

Search engines menelusur dan mengembalikan dokumen-dokumen yang relevan dengan query berdasarkan kata kunci yang diberikan oleh user. Proses dengan model ini sangat tidak efektif, membuang waktu dan tidak mampu memberikan sebuah jawaban informasi yang pasti kepada user. Oleh karenanya dibutuhkan sebuah sistem yang mampu untuk memberikan informasi berupa jawaban yang pasti dan benar atas pertanyaan yang disampaikan dalam bahasa alami oleh user. Cara terbaik untuk mengatasi permasalahan ini adalah dengan penggunaan sistem tanya jawab (Question Answering System, QAS). Sistem tanya jawab adalah sebuah sistem yang menerima pertanyaan dari user dalam bahasa alami dan kemudian mengembalikan teks pendek sebagai jawabannya. Paper ini disajikan sebagai review kecil terkini atas beberapa pendekatan dan teknik yang telah digunakan untuk mengimplementasikan QAS dalam beragam kategori baik closed domain atau pun open domain-based QAS dan web-based QAS yang mana diharapkan dapat membantu dalam memberikan penanda arah bagi riset di bidang ini

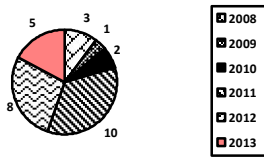
Kata kunci: *search engine, question answering system, query*

1. PENDAHULUAN

QAS adalah sebuah sistem yang mengembalikan teks pendek sebagai sebuah jawaban yang pasti dengan cara mendapatkan kembali jawabannya dari basis datanya untuk sebuah pertanyaan yang ditulis dalam bahasa alami [4]. Intinya QAS disajikan dengan masukan pertanyaan berbahasa alami dan keluaran yang diharapkan adalah jawaban pasti yang teridentifikasi baik dalam suatu teks atau pun dalam fragmen kecil teks yang mengandung jawaban. Pada IR, query dilakukan berdasarkan pada kata-kata kunci masukan dan keluaran yang dihasilkan berupa sebuah daftar dokumen-dokumen yang dianggap relevan dengan query masukan user. Namun sebaliknya QAS mengijinkan user untuk menyampaikan pertanyaan secara langsung pada sistem dengan bahasa alami dan sebuah jawaban selanjutnya akan disediakan oleh QAS dalam bentuk jawaban pendek dan pasti yang telah disarikan dari sumber-sumber dokumen. Sehingga dengan demikian teknologi QAS berbeda dengan teknologi IR umumnya.

QAS telah menjadi sebuah topik riset yang menarik dan bahkan akhir-akhir ini telah menyita perhatian banyak peneliti. Sejumlah riset yang berhubungan dengan perkembangan QAS telah banyak dihantarkan dalam beragam bentuk penyajian oleh para peneliti. Pengembangan QAS dengan pendekatan semantik telah secara intensif diusulkan oleh [12][17][19][20][22] untuk memfasilitasi penggunaan bahasa query atau aplikasi tertentu dalam bahasa alami (*natural language*). Pengembangan QAS dengan pendekatan semantic tidak dapat dipisahkan dengan penggunaan ontology yang mana ontology telah memainkan peranan kunci dalam penyediaan vocabulary bagi pemahaman arti/makna kata-kata yang bersesuaian [19][23] serta menutup gap antara web semantik dengan basis data dalam pemberian jawaban yang dapat terlacak [1]. Sementara itu menurut [13] ontology dapat digunakan untuk mempertinggi derajat kepintaran (*intelligence*) sistem sehingga sistem dapat secara otomatis membangkitkan wilayah pengetahuan baru baik bagi pertanyaan yang lebih komprehensif atau pun bagi jawaban yang bersesuaian. Query oleh user yang tidak dapat ditemukan kembali jawabannya dari ontology akan diperoleh oleh QAS dari dokumen web melalui teknik cross-dokumen [22], *cross-lingual* [5] atau pun dikonversi kedalam Resource Description Framework (RDF) [17]. Jawaban yang dikembalikan ke user oleh QAS biasanya dalam bentuk jawaban pendek dan pasti. Namun oleh [6] jawaban yang diberikan juga dilengkapi dengan keterangan lain sehingga terbangun QAS termodifikasi yang disebut dengan sistem bertipe pertanyaan dunia nyata. Pengusulan dan pengimplementasian beberapa pendekatan yang berbeda-beda secara tidak langsung telah mempertinggi aras perkembangan QAS dalam hal unjuk

kerjanya. Hal yang terpenting dari diskusi *QAS* disini adalah memberikan perhatian pada beberapa tantangan yang sedang muncul, seperti misalnya penggunaan koleksi data selain berbahasa Inggris, pengkombinasian bermacam teknik, dan juga kemungkinan pengusulan pendekatan baru. Kontribusi utama kami dalam paper ini adalah pada pengalamatan kecenderungan arah dari riset *QAS* dan membantu mengidentifikasi beberapa tantangan-tantangan kedepan dalam area ini berdasarkan *survey* pada riset-riset terkini yang terpublikasi antara tahun 2008 sampai 2013.



Gambar 1. Komposisi referensi yang diacu dalam laporan ini berdasarkan tahun publikasi

Secara keseluruhan, paper ini diorganisasikan kedalam seksi-seksi berikut ini. Arsitektur *QAS* terkini akan diuraikan dalam seksi 2. Seksi 3 akan menyajikan arah perkembangan *QAS* saat ini. Tantangan-tantangan yang sedang muncul pada *QAS* secara global dan *QAS* berbahasa Indonesia dihantarkan dalam seksi 4. Akhirnya seksi 5 menyajikan sebuah ringkasan dari area riset *QAS*.

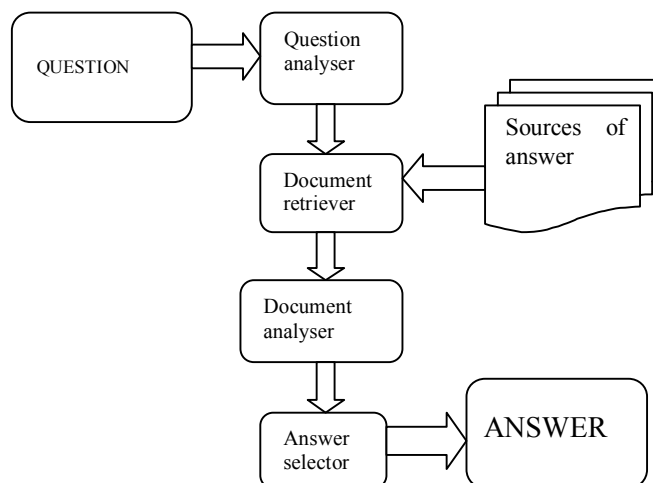
2. ARSITEKTUR UMUM *QAS*

Tugas pokok *QAS* adalah untuk memperoleh kembali sebuah jawaban pasti dan benar dari suatu koleksi dokumen bagi pertanyaan yang diberikan dalam bahasa alami. *QAS* secara *prototype* terdiri atas empat komponen modul, yaitu penganalisa pertanyaan (*question analyzer*), penemu kembali dokumen/informasi (*document retrieval*), penganalisa dokumen (*document analyzer*), dan pemilih jawaban (*answer selection*) sebagaimana disajikan pada gambar 2.

2.1 *Question Analyzer*

Secara prinsip terdapat tiga jenis *question analyzer* yang menggunakan pendekatan berbasis dokumen [6], yaitu berbasis leksikal, berbasis sintaksis, dan berbasis semantik. Kata kunci pertanyaan dan jenis jawaban yang diharapkan (*expected answer type (EAT)*) dihasilkan oleh *question analyzer* dalam *QAS* berbasis leksikal [27][28]. Pada *QAS* berbasis sintaksis, *question analyzer* menghasilkan pohon pertanyaan secara sintaksis. Baik kata kunci pertanyaan dan pohon pengurai pertanyaan digunakan untuk memperoleh kembali dokumen yang relevan. Pengklasifikasian pertanyaan menjadi salah satu bagian penting pada pemrediksian jenis entitas jawaban dari sebuah pertanyaan berbahasa alami. Pemetaan sebuah pertanyaan pada sebuah kategori yang telah dikenal yang menspesifikasi jenis entitas jawaban yang diharapkan telah dikerjakan oleh [10]. Selain dari itu, pengidentifikasian jenis semantik pertanyaan juga menjadi langkah penting bagi pemahaman pertanyaan terhadap apa yang ditanyakan. Pada langkah klasifikasi ini pertanyaan diletakkan pada beberapa kategori semantik dan kemudian diamati kata kunci pertanyaannya, seperti kapan mencari tanggal/waktu, dimana untuk lokasi, siapa menanyakan orang, dan lain-lain. Akurasi pengklasifikasian pertanyaan menjadi hal yang sangat penting bagi unjuk kerja sistem secara keseluruhan karena apabila terjadi kesalahan dalam klasifikasi pertanyaan maka akan berakibat pada kinerja pemrosesan selanjutnya. Sekali pertanyaan diklasifikasikan kedalam jenis pertanyaan yang cocok maka akan dapat ditentukan jenis jawaban yang dikehendaki, kemudian mengekstrak kata kunci yang paling relevan dan selanjutnya mereformulasi pertanyaan kedalam multi pertanyaan secara semantik [24]. Hierarki jenis pertanyaan yang berdasar pada jenis jawaban yang dicari dikonstruksi oleh sistem pengklasifikasian pertanyaan dan selanjutnya pertanyaan diletakkan kedalam kategori yang cocok dalam hierarki. Kategorisasi pertanyaan dapat diimplementasikan dalam banyak cara dan pendekatan. Pencocokan pola adalah salah satu cara yang paling sederhana namun sangat efektif. Oleh karena itulah [2] menggunakannya untuk menaikkan efektifitas sistem pengklasifikasian pertanyaan dengan pemeringkatan/perankingan dokumen yang mirip. Sementara itu, algoritma berbasis aturan heuristik membutuhkan penulisan beberapa aturan heuristik secara manual dalam

mengklasifikasi pertanyaan. Pada sisi lain, sebuah pendekatan berbasis pembelajaran mesin (*machine learning (ML)*) dapat secara otomatis membangun sebuah program pengklasifikasi pertanyaan yang berunjuk kerja tinggi karena mempunyai kemampuan mudah beradaptasi pada sebuah domain baru. Hal ini dibuktikan dengan banyaknya riset mengenai klasifikasi pertanyaan berhasil dibangun dengan *Support Vektor Machine (SVM)* [8][11][21].



Gambar 2. Modul-modul QAS

Bagaimanapun *SVM* sebagai sebuah teknik klasifikasi yang menggunakan pendekatan ML mempunyai masalah utama yaitu adanya dimensionalitas yang tinggi pada space fiturnya. Oleh karenanya, berdasarkan masalah yang ada tersebut [10] mengaplikasikan teknik analisis latent semantic yang mengkombinasikan SVM dengan pengklasifikasi berbasis *back-propagation neural network (BPNN)* dan akhirnya berhasil mereduksi space fitur yang besar yang mulanya dihasilkan dalam teknik ML. Masalah serupa dipecahkan pula oleh [23] dengan menggunakan online QAS yang memperkerjakan aturan pencocokan (*matching*). Pada riset mereka, [11] dan [23] telah berhasil tidak hanya meningkatkan efisiensi waktu dalam mengklasifikasi pertanyaan tapi juga meningkatkan akurasi klasifikasi dengan cara menghilangkan fitur-fitur yang berulang. Hasil akhir dari [11] adalah unjuk kerja BPNN lebih baik dari SVM.

2.2 Document Retriever

Penemuan sebuah jawaban dimulai pada langkah mengidentifikasi dokumen-dokumen yang memungkinkan mengandung jawaban. Misi dari perolehan kembali dokumen adalah untuk mengekstrak dokumen yang relevan dari suatu koleksi data. Pada langkah ini dokumen-dokumen dipilah kedalam beberapa bagian dan selanjutnya bagian tersebut akan diperlakukan sebagai dokumen pula, dan seterusnya. Pemilihan dan penapisan bagian-bagian yang dipertimbangkan relevan dengan pertanyaan masukan selanjutnya dikerjakan dalam rangka memfokuskan space pencarian jawaban. Guna keperluan ini metode sorting digunakan untuk mengurutkan bagian-bagian sehingga akan menghasilkan urutan dokumen yang paling cocok dimana jawaban pasti diasumsikan tersedia di dalamnya. Hal yang lain lagi, indeksasi basis data dapat juga dijadikan dasar bagi proses retrieval informasi yang mana dilaksanakan dengan cara semua permintaan pencarian dijawab dengan proses indeksasi yang didasarkan pada indeksasi kata kunci dalam pengetahuannya. ID dokumen merekam pertanyaan dan jawaban dengan suatu nomor ID yang bersesuaian [26]. Sebaliknya, sebuah teknik tak gayut basis data dihantarkan oleh [15] yang mengimplementasikan pendekatan model semantik distribusional (*DSM*) yang mana mengkombinasikan beberapa teknik terkemas dalam question cube framework untuk mendapatkan kembali bagian yang mengandung jawaban pasti bagi pertanyaan berbahasa alami.

2.3 Document Analyzer

Penganalisa dokumen menganalisa dokumen-dokumen yang dipilih oleh document retriever guna diidentifikasi frase yang berjenis sesuai atau mendekati. Biasanya hal ini dilaksanakan dengan menggunakan suatu pengidentifikasi entitas ternamai (*named entity identifier*) sehingga string-string multiword sebagai nama orang, tanggal, lokasi, dan lain-lain dapat dikenali dan diklasifikasi.

2.4 Answer Selector

Sebagai sebuah komponen akhir dalam arsitektur *QAS*, module answer selection bertanggung-jawab pada identifikasi, ekstraksi, dan validasi jawaban yang terkumpul dari bagian-bagian yang telah terurut padanya. Pada umumnya validasi yang dikerjakan dengan berdasarkan pada pendekatan *ML* bertanggung jawab dalam menentukan kebenaran sebuah jawaban, yaitu apakah jawaban dari sebuah *QAS* adalah benar atau salah, sebagaimana dikerjakan oleh [7] dan [8]. Sesuai dengan [18] sistem pemvalidasi jawaban harus memberikan sebuah keputusan bagi jawaban yang dipilih, divalidasi atau ditolak, dari sekumpulan dari triplets yang mengandung pertanyaan, jawaban, dan teks pendukung yang diterima. Lebih lanjut, [7] membandingkan unjuk kerja pemvalidasian antara validasi jawaban yang terpilih pada sistem berbasis teks dan berbasis web.

3. ARAH PENGEMBANGAN *QAS* BERBAHASA INDONESIA

QAS adalah suatu sistem untuk memperoleh kembali informasi berdasarkan pada masukan pertanyaan berbahasa alami guna mendapatkan jawaban pendek dan pasti. Sejumlah model *QAS* telah diciptakan oleh banyak peneliti yang mana *QAS* tersebut telah dikembangkan dalam multi-lingual sebagaimana dalam [5] atau pun dalam bahasa tertentu, seperti berbasis *Tibetan* dalam [23], berbasis *Turkish* dalam [2], berbasis *Chineses* dalam [13] dan [25], serta berbasis *French* [7]. Lebih lanjut, pengembangan berbasis *Indonesian* juga telah dilakukan oleh [3][6][9][12][14][28][29].

Bahasa Indonesia yang dilafalkan oleh lebih 167 juta orang adalah bahasa resmi di Indonesia. Sepanjang pengetahuan kami, bahasa Indonesia juga sedang dipelajari oleh orang-orang di beberapa negara seantero dunia. Berdasarkan pada fakta ini, maka sangatlah penting untuk melaksanakan pengembangan yang lebih intensive terhadap riset *QAS* berbahasa Indonesia. Berawal dari kenyataan inilah maka riset *QAS* berbahasa Indonesia perlu untuk dialamati. *QAS* berbahasa Indonesia telah dikerjakan oleh [9] dan [12] dengan analisis semantik untuk memperoleh representasi secara semantik dari sebuah kalimat. Kelanjutan analisis semantik dikerjakan oleh [12] dengan penambahan sejumlah aksioma yang dirancang untuk mengkode pengetahuan yang bermanfaat membangun jawaban. Adanya penambahan aksioma ini telah menjadikan *QAS* yang diusulkan oleh [12] lebih tahan dari *QAS* dalam [9] dimana sistem telah mempunyai kemampuan untuk menjawab pertanyaan yang sebelumnya tidak mampu dijawab pada riset dalam [9].

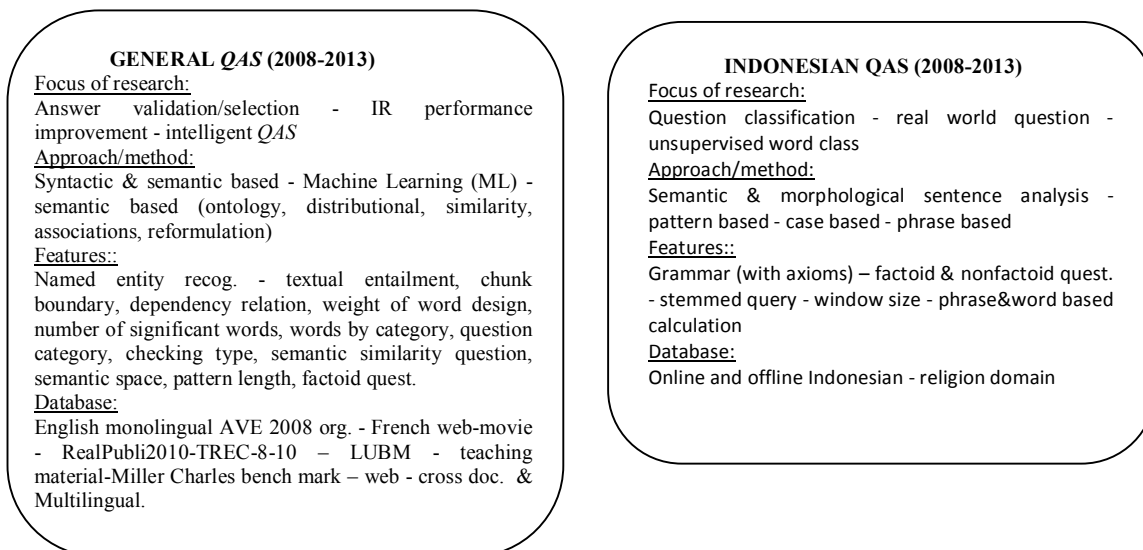
Salah satu tantangan besar dalam pengembangan *QAS* secara umum adalah pada kategorisasi pertanyaan kedalam suatu jenis tertentu. Pendekatan umum untuk kategorisasi pertanyaan adalah dengan pencocokan pola setiap jenis pertanyaan berdasarkan pada posisi dari kata tanya dan kata kunci pertanyaan yang merepresentasikan sebuah jenis pertanyaan. Pertanyaan dapat dikelompokkan kedalam jenis pertanyaan *factoid* (semisal orang, organisasi, lokasi, kuantitas, tanggal-waktu) dan *non-factoid* (semisal definisi, alasan, metode)[28]. Sesuai dengan hal itu, [28] mengkonstruksi *QAS* untuk menangani pertanyaan *factoid* dan *non-factoid* dengan menggunakan pendekatan *monolingual* dan kemudian dilanjutkan oleh [29] dengan melakukan pendekatan yang berdasarkan frase dari jenis pertanyaan *factoid* berbasis *ML*. Sebuah modifikasi pada *QAS* telah dikerjakan pula oleh [6] dimana *QAS* yang didasarkan pada kasus hanyaterdapat bagian analisis pertanyaan dan komponen *retrieval* kasus. Sebagaimana jawaban telah tersedia dalam basis kasus, [6] mengganti komponen retriever document dengan *retririver* kasus. Disamping hal-hal tersebut, umumnya dalam sebuah kalimat bahasa Indonesia bisanya kata kerja (*verb*) sebagai predikat tidaklah muncul dalam bentuk kata asli (kata dasar) tapi biasanya dilengkapi dengan imbuhan. Penyertaan imbuhan (seperti “-kan”) dalam kalimat bahasa Indonesia akan mempengaruhi kalimat secara sintaksis. Variasi daro sintaks kalimat ditangani oleh [14] dengan induksi klas kata yang memperkerjakan pengklasteran *hierarchical agglomerative*.

4. TANTANGAN DALAM PENGEMBANGAN *QAS*

4.1 *QAS* dalam Cakupan Global

Arah pengembangan *QAS* dalam paper ini diidentifikasi dengan empat komponen, yaitu topik yang diisukan, pendekatan yang digunakan, fitur yang dipekerjakan, dan basis data yang dimanfaatkan. Pada umumnya tidak ada topik yang menjadi focus utama dalam pengembangan *QAS*. Namun pengembangan *QAS* pintar (*intelligent QAS*) dan penggunaan *cross*-dokumen pada

basis data multilingual adalah topic yang sedang banyak diteliti dalam lima tahun terakhir. Pengembangan *QAS* dengan mengimplementasikan pendekatan kearah ontology-semantik dengan menggunakan beragam fitur dan dokumen web juga sedang menjadi *trend*. Sebuah tantangan baru yang sedang muncul dalam analisis pertanyaan adalah *query* berjangkauan (*range query*) yang mana pada pertanyaan jenis ini membutuhkan jawaban dua buah nilai antara. Tentunya *query* berjangkauan ini berbeda dengan *IR* pada umumnya.



Gambar 3. Ilustrasi pengembangan *QAS* dalam lingkup global dan berbahasa Indonesia

4.2 *QAS* berbahasa Indonesia

QAS berbahasa Indonesia telah dikembangkan sampai kearah peningkatan unjuk kerja kategorisasi pertanyaan dalam penentuan jenis pertanyaan. Bahasa Indonesia masih merupakan bahasa “*under-resourced*” ketika dalam pengolahan bahasa alami. Disamping itu bahasa Indonesia juga diucapkan oleh lebih dari 200 suku bangsa (seperti Jawa, Sunda, Bali, Batak, dan sebagainya) yang mempunyai dialek berbeda ketika berbahasa Indonesia. Hal ini tentunya akan mempengaruhi gramatika kalimat. Ditambah lagi adanya beberapa penggunaan kata-kata dalam bahasa Indonesia yang terdiri dua kata namun hanya memiliki satu makna, sebagai contoh “Siapakah pemenang nomor lari 100 meter putrid?”. Disini, “nomor lari” dalam bahasa Indonesia dianggap satu kata tapi dalam bahasa Inggris mungkin akan dianggap dua kata sehingga konsekuensinya akan ditempatkan dalam frase yang berbeda. Selanjutnya adanya penyertaan imbuhan pada kata dasar yang berkedudukan sebagai predikat kalimat, dimungkinkan akan memberikan efek pada saat proses penguraian kalimat. Keberadaan kata majemuk yang membuat kata menjadi lebih kompleks dan beberapa indikator lain sebagaimana tersebut diatas akan memunculkan tantangan tersendiri dalam mengembangkan *QAS* berbahasa Indonesia.

5. KESIMPULAN

Sebuah survey kecil pada *QAS* telah disajikan pada paper ini. Baik bagi *QAS* secara global maupun *QAS* berbahasa Indonesia keduanya mempunyai arah kecenderungan yang hampir sama dalam sisi pengembangannya. Sejumlah pendekatan telah digunakan dalam riset *QAS*. Bagaimanapun *QAS* berbahasa Indonesia akan mempunyai tantangan yang lebih besar sehingga membutuhkan perhatian khusus terutama bagi para calon peneliti yang berkecimpung di area sistem tanya jawab. Hal ini tidak hanya sekedar persoalan dalam parsing kalimat tetapi juga dalam pemahaman kalimat secara semantic dan sintaksis berkaitan dengan banyaknya ragam kata dan dialek dalam bahasa Indonesia.

DAFTAR PUSTAKA

- [1] Cali, A., Gottlob, G., and Lukasiewicz, T., “A general datalog-based framework for tractable query answering over ontologies”, Journal of Web semantics: services and agents on the world wide web, Elsevier, Vol. 14, 2012, pp. 57-83, doi: 10.1016/j.websem.2012.03.001.

- [2] Celebi, E., Gunel, B., and Sen, B., "Automatic question answering for Turkish with pattern matching", *IEEE Trans.*, 2011, pp. 389-393
- [3] Ferdian, F., and Purwarianti, A., "Implementation of semantic analysis in Indonesian text-understanding evaluation system", In *Proceedings of IEEE International Conference on Computational Intelligent and Cybernetics*, Bali, 2010.
- [4] Ferrández, O., Izquierdo, R., Ferrández, S., and Vicedo, J.L., "Addressing ontology-based question answering with collections of user queries", *Journal of Information Processing and Management*, vol. 45, 2009, pp. 175–188, doi: 10.1016/j.websem.2011.01.002.
- [5] Ferrandez, O., Spurk, C., Kouylekov, M., Dornescu, I., Ferrandez, S., Negri, M., Izquierdo, R., Tomas, D., Orasan, C., Neumann, G., Magnini, B., and Vicedo, J.L., "The QALL-ME framework: a specific-domain multilingual question answering architecture", *Journal of web semantics: science, services and agents on the world wide web*, Elsevier. Vol. 9, 2011, pp. 137-145.
- [6] Fikri, A., and Purwarianti, A., "Case based Indonesian closed domain question answering system with real word questions" In *Proceedings of IEEE 7th the International Conference on Telecommunication Systems, Services, and Applications (TSSA 12)*, 2012, pp. 181-186.
- [7] Grappy, A., Grau, B., Falco, M.H., Ligozat, A.L., Robba, I., and Vilnat, A., "Selecting answers to question from web documents by a robust validation process", In *Proceedings of IEEE International Conference on Web Intelligence and Intelligent Agent Technology*, IEEE Press, 20011, pp. 55-62, doi: 10.1109/WI-IAT.2011.210.
- [8] Huang, Z., Thint, M., and Qin, Z., "Question classification using head words and their hypernyms", In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 08)*, IEEE Press, 2008, pp. 927-936.
- [9] Larasati, S.D., and Manurung, R., "Towards a semantic analysis of bahasa Indonesia for question answering", In *Proceeding of the 10th Conference of the Pacific Association for Computational Linguistics*, IEEE Press, 2008, pp. 273-280.
- [10] Loni, B., Khoshnevis, S.H., and Wiggers, P., "Latent semantic analysis for question classification with neural networks", In *Proceedings of IEEE International Conference on ASRU*, 2011, pp. 437-442.
- [11] Loni, B., Tulder, G.V., Wigger, P., Tax, D.M.J., and Loog, M., "Question classification by weighed combination of lexical , syntactical and semantic feature", 2011.
- [12] Mahendra, R., Larasati, S. D., and Manurung, R., "Extending an Indonesian semantic analysis-based question answering system with linguistic and world knowledge axioms", In *Proceedings of IEEE The 22nd Pacific Asia Conference on Language, Information, and Computation*, Cebu, Philippines, 2008, pp. 262-271.
- [13] Mei, L.H., "Intelligent question answer system of research based ontology on excellent course", In *Proceedings of IEEE 4th International Congference on Computational and Information Science*, IEEE Press, 2012, pp. 784-787, doi:10.1109/ICCIS.2012.177
- [14] Mistica, M., Lau, J.H., and Baldwin, T., "Unsupervised word class induction for under-resourced language: a case study on Indonesian", In *Proceedings of The 6th International Joint Conference on Natural Language Processing (IJCNLP 13)*, Nagoya, Japan, 2013.
- [15] Molino, P., Basile, P., Caputo, A., Lops, P., and Semeraro, G., "Exploiting distributional semantic model in question answering", In *Proceedings of IEEE 6th International Conference on Semantic Computing*, IEEE Press, 2012, pp. 146-153.
- [16] Muthukrishnan Ramprasath and Hariharan, S. 2012. Using ontology for measuring semantic similarity for question answering system. In *Proceedings of IEEE International Conference on Advanced Communication Control and Computing Technology (ICACCCT 12)*, pp. 218-223.
- [17] Najmi, E., Hashmi, K., Khazalah, F., and Malik, Z., "Intelligent semantic question answering system", In *Proceeding of IEEE International Conference on Cybernetics*, 2013, pp. 255-260.
- [18] Pakray, P., Pal, P., Bandhyopadhyay, S., and Gelbukh, A., "Automatic answer validation system on English Language", In *Proceeding of IEEE 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE 10)*, IEEE Press, 2010, pp. 329-333.
- [19] Ramprasath, M., and Hariharan, S., "Using ontology for measuring semantic similarity for question answering system", In *Proceedings of IEEE International Conference on Advanved Communication Control and Computing (ICACCCT 12)*, IEEE Press, 2012, pp. 218-223.
- [20] Ramprasath, M., and Hariharan, S., "Improving QA performance through semantic reformulation", In *Proceedings of IEEE International Conference on Engineering*, Nirma University, Dec. 2013, pp. 1-4.
- [21] Silva, J., Coheur, L., Mendes, A., and Wichert, A., "From symbolic to sub-symbolic information in question classification", *Artificial Intelligence Review*, vol. 35, no.2, 2011, pp. 137-154.
- [22] Tartir, S., Arpinar, I.B., and McKnight, B., "Semantic QA: Exploiting semantic associations for cross-document question answering", In *Proceedings of IEEE 4th International Symposium on Innovation in Information and Communication Technology*, 2011, pp. 1-6.
- [23] Te, R., "Research on question classification method of Tibetan online automatic question answering system", In *Proceedings of IEEE 4th International Conference on Intelligent Networks and Intelligent System*, IEEE Press, pp. 211-213, doi: 10.1109/ICINIS.2011.42.
- [24] Walke, P.P., and Karale, S., "Implementation approach for various categories of question answering system", In *Proceeding of IEEE Conference on Information and Communication Technology (ICT 2013)*, pp. 402-407.
- [25] We, Z., Xuan, Z., Wei, Z., and Junjie, C., "Design and implementation of influenza question answering system on multi-strategies", In *Proceedings of IEEE International Conferences*, IEEE Press, 2012, pp. 720-722.
- [26] Yunjuan, L., Lijuan, M., Lijun, Z., and Qinlin, M., "Research and application of information retrieval techniques in intelligent question answering system", In *Proceedings of IEEE International Conference*, 2011, pp. 188-190.
- [27] Zhang, G., Zhang, W., Bai, Y., Kang, S., and Wang, P., "An open domain question answering system for NTCIR-8 C-C task", In *Proceedings of IEEE NTCIR-8 Workshop Meeting*, Tokyo, Japan, 2010.
- [28] Zulen, A. A., and Purwarianti, A., "Study and implementation of monolingual approach on Indonesian question answering for factoid and non-factoid question", In *Proceedings of IEEE 25th Pacific Asia Conference on Language, Information , and Computational (PACLIC 2011)*, 2011, Singapore.
- [29] Zulen, A.A, and Purwarianti, A., "Using phrase-based approach in machine learning based factoid Indonesian question answering", In *Proceedings of CISAK*, 2013.