

Survei: Question Classification untuk Question Answering System

Abdiansah

Laboratorium Sistem Cerdas, UGM
Departemen Ilmu Komputer, Universitas Sriwijaya
Email: abdiansah@unsri.ac.id

Anny K. Sari

Laboratorium Sistem Cerdas, UGM
Departemen Ilmu Komputer dan Elektronika, Universitas
Gadjah Mada
Email: a_kartikasari@ugm.ac.id

Abstract—*Question Classification (QC)* merupakan salah satu dari tiga komponen utama yang ada dalam *Question Answering System (QAS)*. QC berfungsi untuk mereduksi ruang pencarian sehingga dapat meningkatkan kecepatan dan akurasi QAS. Secara umum kajian tentang QC dapat dibagi menjadi dua bidang yaitu memperdalam analisis fitur yang meliputi analisis leksikal, sintaksis dan semantik serta improvisasi algoritma klasifikasi. Artikel ini berisi laporan survei tentang algoritma klasifikasi untuk QC berdasarkan tiga pendekatan yang sering digunakan yaitu: *Knowledge-Based (KB)*, *Machine Learning (ML)* dan *Hybrid* antara KB dan ML, serta dengan metode *Feature Selection (FS)*. Tujuan dari survei adalah untuk melihat perkembangan QC melalui pendekatan metode yang digunakan serta melihat peluang penelitian yang dapat dilakukan untuk meningkatkan performa QC. Dari hasil survei dapat diperoleh gambaran secara umum bahwa kajian di bidang QC masih bisa dieksplorasi lebih dalam lagi terutama pada wilayah *open-domain* dengan ukuran data yang besar serta domain bahasa yang digunakan.

Keywords—*Question Classification; Question Answering System; Knowledge-Based; Machine Learning; Hybrid; Feature Selection.*

I. PENDAHULUAN

Information Retrieval (IR) merupakan teknik yang digunakan untuk mencari informasi dari sekumpulan dokumen, yang hasilnya berupa daftar informasi yang relevan dengan *keyword* pengguna, teknik ini umumnya digunakan oleh mesin pencari seperti Google, Bing dan lainnya. Seiring perkembangan informasi yang pesat, terkadang pengguna ingin memperoleh jawaban yang cepat dan langsung tanpa harus membuka tautan informasi yang diberikan oleh mesin pencari. Oleh sebab itu dikembangkanlah sebuah sistem yang dapat memenuhi kebutuhan tersebut, yang dikenal dengan *Question Answering System (QAS)* yaitu sebuah sistem yang dapat memproses kueri pengguna dalam bahasa alami dan dapat memberikan jawaban secara langsung kepada pengguna.

Awal mula perkembangan QAS dimulai dari konferensi TREC-8 (*eight Text ReErieval Conference*) yang memberikan kompetisi pertama untuk menjawab pertanyaan berbasis fakta seperti: "Who came up with the name, El Nino?". Dalam kompetisi tersebut, para peneliti menggunakan teknik IR dan *Information Extraction (IE)* untuk membuat QAS yang menghasilkan metode dan peluang penelitian baru, dengan cara menggabungkan metode pengolahan teks dan inferensi

teks berbasis pengetahuan. Secara umum terdapat dua poin hasil dari konferensi tersebut, 1) memperoleh semantik dari suatu pertanyaan; dan 2) menjustifikasi pertanyaan yang benar [10]. Arsitektur QAS umumnya terdiri dari tiga komponen utama yaitu: 1) *Question Analysis*; 2) *Passage Retrieval*; dan 3) *Answer Extraction* [8]. Setiap komponen tersebut mempunyai modul-modul yang spesifik terhadap permasalahan tertentu, misalnya pada komponen *Question Analysis* terdapat beberapa modul seperti *Parsing*, *Question Classification*, *Query Reformulation* [6].

Question Classification (QC) merupakan bagian penting dari QAS karena akan berpengaruh langsung terhadap akurasi ekstraksi jawaban serta menentukan kualitas dan performa keseluruhan dari QAS [1]. QC berfungsi untuk memahami arah pertanyaan yang diajukan, misalnya ada pertanyaan: "Who was the first American in space?" dari pertanyaan tersebut dapat dipastikan bahwa yang ditanyakan adalah nama orang. QC berfungsi untuk menentukan *Expected Answer Type (EAT)* sehingga dapat mereduksi ruang pencarian [13]. EAT merupakan label yang diberikan untuk suatu pertanyaan, misalnya pada contoh sebelumnya maka EAT-nya adalah HUMAN [18].

Menentukan EAT dari suatu pertanyaan tidaklah mudah karena disebabkan oleh beberapa hal diantaranya adalah: 1) suatu pertanyaan dapat memiliki EAT lebih dari satu, misalnya dalam pertanyaan: "Who made the first airplane?" terdapat beberapa EAT seperti: *person*, *company* dan *organization*; 2) Penentuan EAT berdasarkan *wh-group (what, where, who, when, why dan how)* dapat menyebabkan ambiguitas, misalnya untuk pertanyaan: "How long would it take to get to Mars?", kata *how* tidak menyatakan arti tetapi menyatakan waktu/jarak [23]; 3) suatu pertanyaan memiliki *focus* [23] yang menentukan semantik dan arah pertanyaan, misalnya dalam pertanyaan: "What is the capitol of Uruguay?", dimana fokus pertanyaan adalah *capitol* bukan *Uruguay*; 4) banyak pertanyaan memiliki satu jawaban [13], misalnya dalam pertanyaan: "What is the age of the Queen of Holland?" dan "How old is the Netherlands' queen?" yang memiliki jawaban: "since January 1938"; 5) satu pertanyaan memiliki lebih dari satu jawaban [13], misalnya pada pertanyaan: "Who invented the gas laser?" yang bisa dijawab dengan "Ali Javan" atau "a scientist at MIT".

Survei ini bertujuan untuk melihat perkembangan QC terutama dari pendekatan metode yang digunakan serta melihat peluang penelitian yang dapat dilakukan untuk meningkatkan performa QC. Selanjutnya pada Bagian 2, berisi penjelasan

teknis tentang metodologi survei yang dilakukan. Bagian 3 menjelaskan teknik seleksi fitur yang dapat dikombinasikan dengan ML. Bagian 4 berisi penjelasan tentang ketiga pendekatan yang digunakan untuk QC yang terdiri dari penjelasan singkat metode dan ringkasan hasil penelitian yang sudah ada. Bagian 5 memberikan diskusi hasil survei terutama tentang peluang penelitian di bidang QC, dan terakhir Bagian 6 yang berisi kesimpulan.

II. METODOLOGI

Tabel 1 berisi ringkasan beberapa artikel yang telah disurvei. Terdapat empat fokus analisis artikel yang tercermin dalam kolom *Dataset*, *Metode*, *Bahasa* dan *Kelompok*. Kolom *Dataset* secara umum terdiri dari: dataset yang berasal dari literatur, dataset yang bersifat domain khusus, dan dataset modifikasi. Kolom *Metode* berisi metode-metode yang digunakan untuk melakukan klasifikasi. Kolom *Bahasa* berisi

bahasa yang digunakan dalam penelitian, pada survei ini umumnya banyak digunakan bahasa Inggris dan Cina. Kolom terakhir, kolom *Kelompok* berisi kelompok dari masing-masing artikel. Salah satu pekerjaan survei yang dilakukan adalah mengelompokkan metode-metode yang digunakan ke dalam tiga kelompok yaitu: KB (*Knowledge-Based*), ML (*Machine Learning*), *Hybrid* antara KB dan ML serta dengan FS (*Feature Selection*).

Metodologi survei dilakukan sebagai berikut: mengumpulkan artikel terkait dengan QC. Melakukan analisis terhadap metode dan akurasi dari hasil penelitian. Dari sekian banyak artikel terkait dengan topik QC, kami hanya mengambil artikel yang menggunakan metode dengan hasil akurasi yang cukup memuaskan. Berikutnya dilakukan pengelompokan artikel ke dalam tiga kelompok (KB, ML dan *Hybrid*), setelah itu dilakukan analisis untuk mencari peluang penelitian QC terbaru.

TABEL 1. RANGKUMAN ARTIKEL PENELITIAN *QUESTION CLASSIFICATION*

Peneliti	Tahun	Dataset	Metode	Bahasa	Kelompok
Li & Roth [18]	2002	TREC-10	Sparse Network of Winnows (SnoW)	Inggris	ML
Zhang & Lee [40]	2003	USC, UIUC, TREC	Nearest-Neighbors (NN), Naive Bayes (NB), Decision Tree (DT), Snow, Support Vector Machines (SVM)	Inggris	ML
Li & Roth [19]	2004	TREC-10 dan 11	SnoW	Inggris	ML
Skowron & Araki [29]	2004	USC, UIUC, TREC	SVM	Inggris	ML
Day et al. [4]	2005	CLQA	SVM, INFOMAP inference engine (knowledge-based approach)	Cina	Hybrid ML+KB
Purwarianti et al. [26]	2006	dataset taken from Indonesian newspaper sites	feature from shallow parser (SP), WordNet (translated, modified)	Indonesia	ML
Day et al. [5]	2007	UIUC & NTCIR-6 CLQA (English), IASLQ2322C & NTCIR06 CLQA (Cina)	Genetic Algorithm (GA), Conditional Random Field (CRF), SVM	Cina dan Inggris (cross-language)	Hybrid ML+FS
Huang et al. [14]	2007	UIUC, TREC-10 QA Track	SnoW, SVM, Passive-Aggressive Algorithm (PA) (online algorithm)	Inggris	ML
Jia et al. [16]	2007	TREC-2001 & UIUC (translated and modified)	Ensemble Learning: Feature Frequency Classifier (FFC), Weighted FFC (WFFC)	Cina	ML
Xia et al. [37]	2008	dataset cooking (restricted-domain)	Rule-Based (RB), SVM	Cina	Hybrid ML+KB
Fu et al. [9]	2009	dataset computer service & support (restricted-domain)	SVM, Semantic Similarity Model (SSM)	Cina	Hybrid ML+KB
Su et al. [31]	2009	dataset tourism (restricted-domain)	Ensemble Learning: Bagging, AdaBoost	Cina	ML
Ray et al. [28]	2010	UIUC, USC, TREC-8,9,10	Algorithm QC	Inggris	KB
Sun [33]	2010	TREC-8 (translated and modified)	Association Rules (AR)	Cina	KB
Yu et al [39]	2010	dataset tourism (restricted-domain)	Co-Training (CT)	Cina	ML
Loni et al [20]	2011	UIUC	Laten Semantic Analysis (LSA), Backpropagation Neural Network (BPNN), SVM	Inggris	Hybrid ML+FS
Su et al. [32]	2011	dataset tourism	Ensemble Learning: Multiboost	Cina	ML

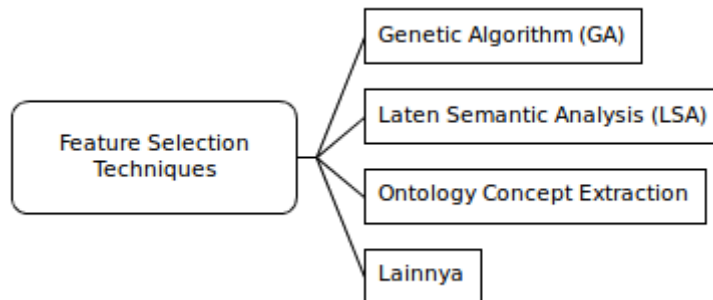
(restricted-domain)

Zheng-wen & Shan- 2012 liang [42]	HIT-IR-Lab	Rule Database Generation	Cina	KB	
Biswas et al [1]	2014	TREC-10	Syntactic & Rule-Based	Inggris	KB

III. SELEKSI FITUR

Seleksi fitur merupakan teknik yang digunakan untuk melakukan optimasi terhadap sekumpulan fitur dengan cara mereduksi fitur dan memilih fitur terbaik sehingga dapat meningkatkan performa suatu sistem. Seleksi fitur dapat dijadikan pra-pengolahan untuk QC, mengingat jumlah fitur yang besar sehingga perlu adanya optimasi. Dari survei yang telah dilakukan ada beberapa peneliti yang menggunakan seleksi fitur diantaranya oleh [5] yang menggunakan *Genetic Algorithm* (GA) untuk seleksi fitur digabung dengan *Conditional Random Field* (CRF) untuk memprediksi *question inform* yang nanti akan dijadikan fitur untuk SVM. *Question inform* merupakan suatu kata yang ada dalam pertanyaan yang

akan dijadikan petunjuk untuk mencari jawaban, misalnya dalam pertanyaan: "*What is the biggest city in the United States?*" maka *question inform* dari pertanyaan tersebut *city*, sehingga *city* akan dijadikan petunjuk untuk QC. Selanjutnya [9] menerapkan seleksi fitur berdasarkan analisis leksikal dan semantik yang mereka sebut dengan *Ontology Concept Extraction*, pendekatan ini bersifat domain spesifik. Terakhir [20] menggunakan *Latent Semantic Analysis* (LSA) untuk mereduksi fitur sehingga ukurannya menjadi lebih kecil dan efisien. Mereka menggunakan *Backpropagation Neural Network* (BPNN) dan SVM untuk mengevaluasi dan didapatkan akurasi BPNN lebih unggul dibanding SVM untuk ukuran fitur yang sedikit. Pada Gambar 1 dapat dilihat teknik-teknik seleksi fitur yang dapat digunakan untuk QC.



Gambar 1. Teknik seleksi fitur yang digunakan dalam *Question Classification*

IV. PENDEKATAN *QUESTION CLASSIFICATION*

Dari hasil konferensi TREC-8 ada beberapa peneliti yang mengusulkan arsitektur QAS yang membahas QC secara umum, diantaranya dilakukan oleh [10], [23] dan [12][13]. Pada penelitian [10], proses pengenalan EAT berdasarkan perbandingan antara representasi semantik pertanyaan dengan representasi suatu simpul taksonomi pertanyaan. Tiap simpul merepresentasikan empat tupel yaitu: 1) representasi semantik pertanyaan; 2) tipe pertanyaan; 3) tipe jawaban; dan 4) kata kunci pertanyaan. Selanjutnya penelitian yang dilakukan oleh [12][13] mencoba menggunakan CONTEX untuk melakukan *syntactic-semantic parser* terhadap pertanyaan. CONTEX merupakan *machine learning deterministic* berbasis *grammar learner/parser* dengan *precision* sebesar 87.6% dan *recall* 88.4% untuk *corpus* berbahasa Inggris. Dari beberapa penelitian terkait dengan QAS diperlihatkan bahwa QC memberikan kontribusi langsung terhadap keberhasilan sistem, sehingga butuh kajian khusus dan mendalam untuk meningkatkan akurasi klasifikasi pertanyaan.

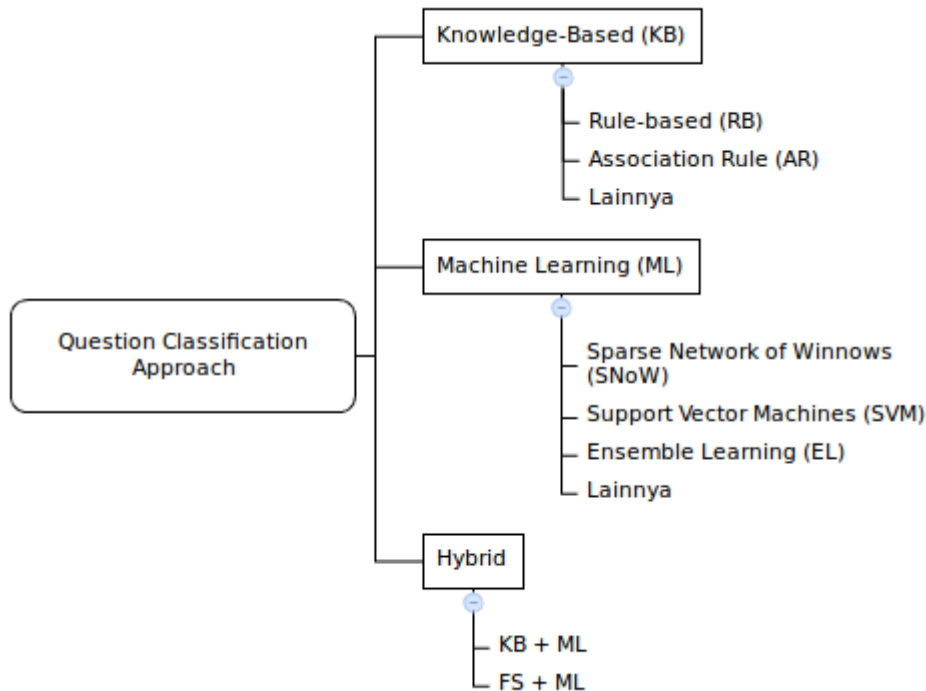
Berdasarkan rekam jejak penelitian QC untuk QAS maka metode QC dapat dibagi menjadi tiga pendekatan yang dapat dilihat pada Gambar 2, yaitu: 1) Pendekatan *Knowledge-Base*

(KB), di mana klasifikasi ditentukan berdasarkan aturan-aturan yang dibuat secara manual, akurasi yang diperoleh cukup tinggi, domain bersifat tertutup dan memerlukan waktu yang lama; 2) Pendekatan *Machine Learning* (ML), menggunakan beberapa dataset yang akan dilatih, kemudian hasil pelatihan tersebut akan digunakan untuk melakukan klasifikasi. Rata-rata akurasi yang diperoleh cukup bagus, bisa digunakan untuk domain terbuka dan tidak memerlukan waktu yang lama; dan 3) Pendekatan *Hybrid*, yaitu menggabungkan kedua pendekatan sebelumnya serta teknik-teknik seleksi fitur.

A. *Knowledge-Based* (KB)

Rule-Based (RB) merupakan cara yang paling sederhana untuk menentukan kelas dari suatu pertanyaan dengan cara membuat seperangkat aturan yang menentukan kelas secara eksplisit. Untuk mempermudah pembuatan aturan, biasanya didahului dengan pra-pengolahan fitur seperti analisis leksikal, sintaksis dan semantik. Pendekatan ini cocok untuk domain yang kecil dan bersifat statis.

Pada penelitian [1] menggunakan RB dan analisis sintaksis dengan fokus pada pertanyaan *5WH* (*why, when, where, which, who* dan *how*). Hasil akurasi rata-rata mencapai



Gambar 2. Pendekatan dalam *Question Classification*

97.5%. Penelitian [3] menggunakan *regex* dan *sememes model*, dengan fitur *interrogative words*, *question focus word*, dan *first sememes*. Artikel menyebutkan bahwa hasil percobaan menunjukkan lebih baik dari *pattern matching* dan *machine learning approach* tetapi tidak menyebutkan metode yg dibandingkan. Tabel 2 memperlihatkan algoritma yang digunakan untuk klasifikasi pertanyaan dengan tipe *wh-words* yang digunakan oleh [1].

TABEL 2. ALGORITMA KLASIFIKASI PERTANYAAN

Case :	Question Word = "Why" Return(" Descriptive Type") ;
Case :	Question Word = "How" If Question, Pattern = QW + [Adj / Adverb]; Return("Factoid Type") ; Else if Question Pattern = QW + [Verb / Model]; Return(" Descriptive Type") ;
Case :	Question Word = "When" Return("Factoid Type") ;
Case :	Question Word = "Where" Return("Factoid Type") ;
Case :	Question Word = "Who" If Question = QW to + aux + NOUN; Return("Descriptive Type") Else Return("Factoid Type") ;
Case :	Question Word = " Which" Return(" Factoid Type") ;
Case :	Question Word = "What" If Question Pattern = (QW + aux + [definition of] + NOUN /JJ+NOUN) Return(" Definition Type") ; Else Return("Descriptive Type Or FactoidType") ;
Case :	Default Return(" Functional Question") ;

Penelitian [33] menggunakan *Association Rule* (AR) dengan fitur sederhana: *words* dan *bi-gram*. Hasil penelitian menunjukkan bahwa AR dapat digunakan untuk QC dan memberikan performa yang bagus sekitar 86.4%. Penelitian [28] dan [30] melakukan penelitian menggunakan *Semantic Approach* (SA) dan *Semantic Pattern* (SP). SA menggunakan WordNet dan term yang ada di Wikipedia. Kombinasi antara WordNet dan Wikipedia dapat menghasilkan akurasi QC sebesar 89.5%. SP menggunakan tiga fitur sederhana yaitu *stopword*, *stemming* dan *information gain*. Hasil penelitian menunjukkan akurasi metode yang diusulkan mendekati metode SVM dengan dataset yang sama.

B. Machine Learning (ML)

Penelitian klasifikasi pertanyaan yang menggunakan *Machine Learning* biasanya berbasis *Supervised Learning*. Dalam rekam jejak penelitian QC, ada beberapa metode yang sering digunakan yaitu: *Sparse Network of Winnows* (SNoW), SVM, *Ensemble Learning* (EL) dan lainnya. Pada bagian ini akan dijelaskan metode-metode tersebut serta penerapannya pada QC.

Sparse Network of Winnows (SNoW)

Sparse Network of Winnows (SNoW) merupakan sebuah arsitektur pembelajaran untuk multi-kelas yang secara khusus dapat digunakan untuk pembelajaran di ruang dimensi tinggi. Arsitektur ini menggunakan fungsi linear sebagai pemisah untuk masing-masing kelas. Fungsi linear menggunakan *update-rule*. Beberapa *update-rule* seperti *Naif Bayes* (NB),

Perceptron dan *Winnows* dapat digunakan untuk fungsi linear tersebut.

Penelitian [18][19] menggunakan model SNoW dan mengenalkan konsep hirarki klasifikasi. Mereka mengusulkan dua kelas klasifikasi yaitu: 6 *coarse-class* dan 50 *fine-class* yang diuraikan pada Tabel 2. Hirarki pertanyaan dengan dua layer yang dibuat oleh [18] pada Tabel 3 dijadikan sebagai standar untuk penelitian QC. Dataset yang digunakan diambil dari UIUC sebanyak 5.500 data untuk *training* dan dari TREC sebanyak 500 data untuk *testing*. Dari hasil pengujian didapat akurasi sebesar 98.0% untuk *coarse-class* dan 95.0% untuk *fine-class*.

Support Vector Machines (SVM)

Support Vector Machines (SVM) adalah model pembelajaran non-probabilistik untuk mengklasifikasikan data. SVM telah berhasil digunakan untuk data dimensi tinggi. SVM adalah model linear diskriminan yang mencoba untuk menemukan *hyperplane* dengan *margin* yang maksimal untuk memisahkan kelas. [15] dan [44] menggunakan SVM sebagai metode klasifikasi dengan kernel linear dimana akurasi yang diperoleh oleh [15] sebesar 89.2% untuk *fine-class* dan 93.4% untuk *coarse-class* menggunakan dataset TREC. Sedangkan [44] memperoleh akurasi sebesar 90.8% untuk *fine-class* dan

95.0% untuk *coarse-class* dengan dataset yang sama. Perbedaan kedua penelitian tersebut terletak pada improvisasi fitur. [42] juga menggunakan SVM tetapi dengan kernel *Radial Basis Function* (RBF). Mereka memperoleh hasil yang berbeda untuk fitur yang berbeda. Akurasi yang diperoleh sebesar 83.6% untuk *fine-class* dan 90.2% untuk *coarse-class* dengan dataset TREC. Klasifikasi pertanyaan menggunakan SVM dengan kernel linear memberikan hasil yang lebih baik dibanding dengan kernel lainnya. Hal ini telah diuji oleh [19] dengan membandingkan akurasi SVM menggunakan empat kernel yaitu: *Linear*, *Polynomial*, *RBF* dan *Sigmoid*. Pada Tabel 4 dapat dilihat hasil perbandingan tiap kernel.

Penelitian [40] membuktikan bahwa metode SVM dapat digunakan untuk QC serta lebih baik dibandingkan dengan metode *Neural Network*, *Naive Bayes*, *Decision Tree* dan juga *SnoW*. Bahkan untuk representasi dan ekstraksi fitur yang sederhana (*Bag of Word* dan *n-grams*) SVM memperoleh akurasi di atas 80%. Zhang dan Lee juga melakukan improvisasi dengan membuat *Kernel Tree* yang dibuat dari struktur sintaksis pertanyaan, akurasi klasifikasi meningkat dan di atas 90%. Selanjutnya [43] juga menggunakan *Kernel Tree* berdasarkan struktur semantik yaitu dengan cara mengukur *semantic similarity*. Akurasi yang diperoleh sebesar 94.0% untuk *coarse-class* menggunakan dataset yang sama dengan [40].

TABEL 3. HIRARKI PERTANYAAN DENGAN DUA KELAS [18]

Coarse-class	Fine-class
ABBREVIATION	abbreviation, expression
DESCRIPTION	Defenition, description, manner, reason
ENTITY	Animal, body, color, creative, currency, disease/medicine, event, food, instrument, lang, letter, <i>other</i> , plant, product, religion, sport, substance, symbol, technique, term, vehicle, word
HUMAN	Group, individual, title, description
LOCATION	City, country, mountain, <i>other</i> , state
NUMERIC	Code, count, date, distance, money, order, <i>other</i> , period, percent, speed, temp, volume/size, weight

Penelitian [41] melakukan penelitian untuk improvisasi SVM. Beberapa penelitian berfokus pada improvisasi fitur untuk meningkatkan akurasi SVM, diantaranya dilakukan oleh [29] yang membuat representasi fitur baru untuk mengatasi kelemahan *Bag of Words* and *ngram* model dengan menggunakan *Subordinat Word Category*, *Question Focus* dan *Syntactic-Semantic Structure*. Penelitian [2] fokus pada *multiple-sentence question*. Penelitian [7] menunjukkan bahwa penggunaan *multiple feature* lebih baik dibandingkan dengan *single feature*.

TABEL 4. PERBANDINGAN AKURASI TIAP KERNEL SVM

Kernel	Linear	Polynomial	RBF	Sigmoid
Akurasi	89.2%	85.2%	85.0%	85.2%

Pada penelitian [27] berhasil melakukan kombinasi antara BoW dan *Subset Tree Kernel* untuk memaksimalkan hasil ekstraksi fitur. Penelitian [17] memperoleh hasil akurasi QC

sebesar 94.98% menggunakan fitur *Bag of Words*, *POS Tag*, *Syntactic Tree*, *Chunk*, *Semantic Roles*. Penelitian [35] dengan metode *semantic gram extraction* yang diusulkan memberikan hasil yang cukup baik. [11] fokus pada imrovisasi fitur menggunakan *syntactic feature*, *semantic feature*, *n-grams*, *Term Weighting*. Penelitian [36] melakukan penelitian yang fokus pada *semantik analisis*. [22] menyimpulkan bahwa dari perbandingan dengan penelitian-penelitian sebelumnya dapat dilihat peningkatan akurasi dengan memperbanyak ekstraksi fitur seperti: *lexical feature* (*word level n-grams*, *stemming & stopword removal*, *word shapes*, *syntactic feature*, *question headword*, *question pattern*, *POS Tag*), *sematic feature* (*hypernyms*, *question category*). Penelitian [26] menggunakan SVM untuk domain bahasa Indonesia yang menghasilkan akurasi yang lebih baik dari *baseline* (92.9%) menggunakan *Shallow Parser* (SP) dan *WordNet* yang diterjemahkan dan dimodifikasi, akurasi didapatkan sebesar 95.1%. *Ensemble Learning* (EL)

Dalam statistik dan ML, metode *Ensemble Learning* (EL) menggunakan beberapa algoritma pembelajaran untuk mendapatkan kinerja prediksi yang lebih baik yang dapat diperoleh dari salah satu algoritma pembelajaran konstituen. Beberapa penelitian sebelumnya telah menunjukkan bahwa penggunaan penggabungan beberapa algoritma menghasilkan akurasi yang lebih baik dibandingkan dengan algoritma tunggal [21]. EL termasuk ke dalam pembelajaran terawasi, karena dapat dilatih dan digunakan untuk melakukan prediksi. Beberapa model EL yang biasa digunakan seperti: *Bayes optimal classifier*, *Bootstrap aggregating* (Bagging), *Boosting*, *Bayesian model averaging*, *Bayesian model combination*, *Bucket of models* dan *stacking*. Dari beberapa penelitian menunjukkan bahwa model *Bagging* dan *Boosting* memberikan hasil yang cukup baik.

Penerapan EL untuk QC telah dilakukan oleh beberapa peneliti diantaranya oleh [16] yang menggunakan metode *Feature Frequency Classifier* (CFF) dan WFFC (*Weighted FFC*) dengan fitur *n-gram* dan *POS Tag*. Hasil akurasi klasifikasi mencapai 87.6% untuk pertanyaan dengan tipe *fine-class*. Penelitian yang dilakukan oleh [31] yang menggunakan algoritma *Bagging* dan *AdaBoost* dengan akurasi *Bagging*: 91.4% (*coarse-class*), 83% (*fine-class*) dan *AdaBoost*: 91.3% (*coarse-class*), 82.7 (*fine-class*). Perbedaan antara kedua algoritma *AdaBoost* dan *Bagging* terletak pada *training-set*. *AdaBoost* tidak independen dan ditentukan dari iterasi latihan terakhir, sehingga klasifikasi dari *AdaBoost* harus dilatih secara berurutan. Sedangkan pada *Bagging* dapat dihasilkan secara paralel sehingga biaya waktu pelatihan dapat disimpan. Selanjutnya pada [32] melanjutkan penggunaan EL untuk QC menggunakan metode *MultiBoost*.

Pada Tabel 6 dapat dilihat algoritma *MultiBoost* yang digunakan oleh [32]. Hasil penelitian mereka menggunakan *MultiBoost* memberikan akurasi mencapai 91.8% untuk *coarse-class* dan 85.6% untuk *fine-class*. Dari beberapa percobaan menunjukkan bahwa pendekatan EL sangat efektif karena memanfaatkan beberapa algoritma untuk meningkatkan tingkat akurasi klasifikasi pertanyaan. Pada Tabel 5 dapat dilihat hasil perbandingan akurasi penggunaan metode EL yang telah dilakukan.

TABEL 5. PERBANDINGAN AKURASI BAGGING & ADABOOST

Peneliti	Coarse-class	Fine-class
Jia et al. (2007) [16]	-	87.6%
Su et al. (2009) [31]	Bagging: 91.4% AdaBoost: 91.3%	Bagging: 83.0% AdaBoost: 82.7%
Su et al. (2011) [32]	91.8%	85.6%

TABEL 6. ALGORITMA MULITBOOST [32]

Input:	training data set D , base learning algorithm L , number of learning rounds T , Vector I_i specifying the number of iterations for each sub-committee.
Output:	ensemble H^* .
Process:	

$D_i(i)=1/m$.	% initialize the weight distribution for all i -th instance.
Set $k=1$.	
for $t=1$ to T	
If $I_i=t$ then	
RandomWeight(D_i)	% drawn from continuous Poisson
distribution.	
	Standardize (D_i).

Metode Lainnya

Penelitian [15] menggunakan metode *Passive Aggressive Algorithm* (PSA) dengan fitur yang sederhana menggunakan *2-gram model*. PA dibandingkan dengan SVM dan SnoW, PA memiliki akurasi yang hampir sama dengan SVM untuk *coarse-class* dengan selisih 0.2%. Sedangkan untuk *fine-class* PA lebih baik dari SVM dengan selisih akurasi 1.6%.

Penelitian [38] menggunakan *Sequential Minimal Optimization* (*Self-SMO*) dengan fitur *tf-idf*. *Self-Revised SMO* memiliki tingkat akurasi yang tinggi dibandingkan metode lain (*WH-Chunk*, *Bayes*, *SMO*) untuk QC dengan domain bahasa Cina. Penelitian [41] dengan *Multilevel Random Walk* yang bersifat semi-supervisi dan fitur *words*, *POS Tag*, *named entity*, *semantic*, mampu mendeteksi pertanyaan yang belum ada label. Penelitian [39] menggunakan metode *Co-Training* (semi-supervisi) dengan fitur *tf-idf* dan *semantic similarity*, hasil penelitian menunjukkan bahwa penggoaan *Co-Training* dapat digunakan untuk QC. Penelitian [24] menggunakan *Conditional Random Field* (CRF) dan fitur: *n-gram* dan *POS-Tag*, memberikan hasil akurasi cukup bagus untuk bahasa Persia.

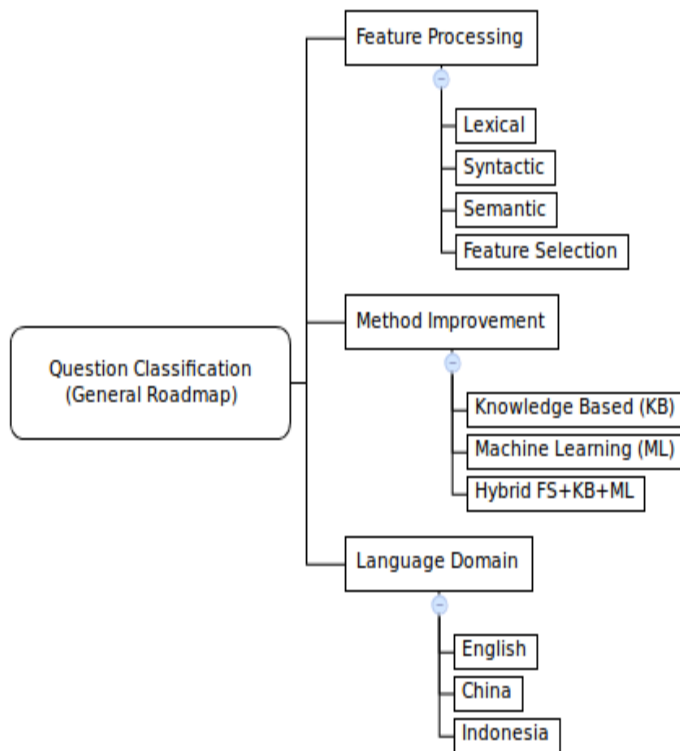
C. Hybrid

Untuk memperoleh hasil akurasi yang maksimal, ML dapat di-*hybrid* dengan metode lain. Ada beberapa penelitian yang telah melakukannya diantaranya dilakukan oleh [4] yang menggunakan metode INFOMAP. *Hybrid Knowledge-based Approach* (INFOMAP) dengan SVM dapat meningkatkan akurasi klasifikasi sampai 92.0%. Penelitian [25] menggunakan algoritma semi-supervisi *Tri-Training* dan SVM. Penelitian [5] menggunakan *syntactic Feature* dan *semantic Feature* dengan metode *hybrid*: GA (*Genetic Algorithm*) untuk *feature selection*, CRF (*Conditional Random Field*) untuk *question informer prediction* dan SVM untuk *question classification*. *Hybrid* dari arsitektur (GA-CRF-SVM) dapat meningkatkan QA dengan memprediksi *question informer*.

Penelitian [14] melakukan improvisasi fitur dengan memberikan pembobotan menggunakan metode *Weighting Model* dengan fitur *n-gram*. Hasil akurasi untuk *fine-class* dan *coarse-class* diatas 85.0%. Penelitian [34] menggunakan *Template Alignment* (TA) dan SVM dengan fitur *POS-Tag* dan *Named Entity Recognition* (NER) yang menghasilkan akurasi lebih baik dibanding penelitian [4]. Penelitian [9] menggunakan *Question Semantic Similarity* dan SVM dengan fitur *lexical analysis*, *semantic analysis* dan *ontology concept extraction* yang menghasilkan akurasi mencapai 91.5%. Terakhir penelitian dilakukan oleh [41] menggunakan *Semantic Similarity Analysis* (SSA). Gabungan SVM dan SSA menghasilkan akurasi di atas 90% dengan domain yang spesifik.

V. DISKUSI

Berdasarkan hasil pengamatan dari survei yang telah dilakukan maka dapat ditarik beberapa hal untuk dijadikan diskusi pemetaan penelitian dibidang QC. Secara umum ada tiga hal yang dapat dilakukan untuk meningkatkan akurasi klasifikasi: 1) dengan memperdalam teknik pengolahan fitur yang meliputi leksikal, sintaksis dan semantik serta penggunaan seleksi fitur; 2) improvisasi metode klasifikasi (KB, ML, Hybrid) dan; 3) menggabungkan kedua hal tadi. Setiap metode yang digunakan pasti memiliki kelebihan dan kekurangan, seperti contoh metode SVM yang telah direkomendasikan oleh komunitas QAS untuk QC ternyata memiliki kelemahan pada fitur dimensi tinggi. Oleh karena itu diperlukan proses pengolahan fitur seperti teknik seleksi fitur. Penelitian ML menggunakan Ensemble Learning (EL) menunjukkan hasil akurasi yang cukup tinggi sehingga memberikan peluang untuk dilakukan improvisasi lagi terutama pemilihan metode EL. Beberapa literatur menyebutkan bahwa algoritma *Bagging* dan *AdaBoost* sering digunakan untuk EL tetapi tidak menutup kemungkinan untuk digunakan metode EL yang lain.



Gambar 3. Pemetaan umum penelitian *Question Classification*

Dari beberapa hasil penelitian menunjukkan bahwa *hybrid* antara KB+ML serta FS+ML dapat meningkatkan akurasi klasifikasi, namun belum ditemukan penelitian yang menggabungkan ketiganya FS+KB+ML. Oleh karena itu salah satu hasil dari survei ini adalah mengusulkan untuk mencoba teknik gabungan tersebut. Selain dari sisi pengolahan fitur dan improvisasi metode klasifikasi, domain bahasa memiliki peran yang cukup signifikan untuk QAS, karena perbedaan bahasa menyebabkan QAS spesifik terhadap bahasa tertentu. Dari survei yang dilakukan umumnya banyak menggunakan domain

berbahasa Inggris dan Cina, padahal perbedaan struktural setiap bahasa berbeda sehingga diperlukan metode yang spesifik untuk mengolah fitur dari bahasa tertentu. Dari hasil analisis ini maka dapat dibuat pemetaan umum untuk penelitian QC yang dapat dilihat pada Gambar 3.

VI. KESIMPULAN

Question Classification (QC) merupakan salah satu tahapan yang sangat penting dalam *Question Answering System* (QAS) karena dapat berpengaruh langsung terhadap QAS. QC berfungsi untuk mereduksi ruang pencarian sehingga dapat meningkatkan kecepatan dan akurasi QAS. Secara umum kajian tentang QC dapat dibagi menjadi dua bidang yaitu memperdalam analisis fitur yang meliputi analisis leksikal, sintaksis dan semantik serta improvisasi algoritma klasifikasi. Artikel ini berisi laporan survei tentang algoritma klasifikasi untuk QC berdasarkan tiga pendekatan yang sering digunakan yaitu: *Knowledge-Based* (KB), *Machine Learning* (ML) dan *Hybrid* antara KB dan ML, serta dengan metode *Feature Selection* (FS).

Survei ini bertujuan untuk melihat perkembangan QC terutama dari pendekatan metode yang digunakan serta melihat peluang penelitian yang dapat dilakukan untuk meningkatkan performa QC. Umumnya pendekatan *hybrid* menggunakan gabungan antara KB+ML dan FS+ML, tetapi belum ada yang menggabungkan ketiganya FS+KB+ML. Selain dari sisi pengolahan fitur dan improvisasi metode klasifikasi, domain bahasa memiliki peran yang cukup signifikan untuk QAS, karena perbedaan bahasa menyebabkan QAS spesifik terhadap bahasa tertentu. Dari hasil survei yang dilakukan dapat dibuat pemetaan umum untuk penelitian QC. Dari hasil survei dapat diperoleh gambaran secara umum bahwa kajian di bidang QC masih bisa dieksplorasi lebih dalam lagi terutama untuk wilayah *open-domain* dengan ukuran data yang besar dan domain bahasa yang digunakan.

REFERENSI

- [1] Biswas, P., Sharan, A., & Kumar, R. (2014). Question Classification using syntactic and rule based pproach. In *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on* (pp. 1033-1038). IEEE.
- [2] Blooma, M. J., Goh, D. H. L., Chua, A. Y. K., & Ling, Z. (2008). Applying question classification to Yahoo Answers. In *Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the* (pp. 229-234). IEEE.
- [3] Cai, D., Bai, Y., Dong, Y., & Liu, L. (2007). Chinese Question Classification Using Combination Approach. In *Semantics, Knowledge and Grid, Third International Conference on* (pp. 334-337). IEEE.
- [4] Day, M. Y., Lee, C. W., Wu, S. H., Ong, C. S., & Hsu, W. L. (2005). An integrated knowledge-based and machine learning approach for Chinese question classification. In *Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE'05. Proceedings of 2005 IEEE International Conference on* (pp. 620-625). IEEE.
- [5] Day, M. Y., Ong, C. S., & Hsu, W. L. (2007). Question classification in english-chinese cross-language question answering: an integrated genetic algorithm and machine learning approach. In *Information Reuse and Integration, 2007. IRI 2007. IEEE International Conference on* (pp. 203-208). IEEE.

- [6] Dwivedi, S. K., & Singh, V. (2013). Research and Reviews in Question Answering System. *Procedia Technology*, 10, 417-424.
- [7] Duan, L., Niu, Y., & Chen, J. (2011). Study on question classification approach mixing multiple semantic characteristics together. In *Computer Research and Development (ICCRD), 2011 3rd International Conference on* (Vol. 1, pp. 354-357). IEEE.
- [8] Ezzeldin, A., & Shaheen, M. (2012). A survei of Arabic question answering: Challenges, tasks, approaches, tools, and future trends. In *Proceedings of The 13th International Arab Conference on Information Technology (ACIT 2012)* (pp. 1-8).
- [9] Fu, J., Qu, Y., & Wang, Z. (2009). Two level question classification based on SVM and question semantic similarity. In *Electronic Computer Technology, 2009 International Conference on* (pp. 366-370). IEEE.
- [10] Harabagiu, S. M., Paşca, M. A., & Maiorano, S. J. (2000). Experiments with open-domain textual question answering. In *Proceedings of the 18th conference on Computational linguistics-Volume 1* (pp. 292-298). Association for Computational Linguistics.
- [11] Harb, A., Beigbeder, M., & Girardot, J. J. (2009). Evaluation of question classification systems using differing features. In *Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for* (pp. 1-6). IEEE.
- [12] Hovy, E. H., Gerber, L., Hermjakob, U., Junk, M., & Lin, C. Y. (2000). Question Answering in Webclopedia. In *TREC* (Vol. 52, pp. 53-56).
- [13] Hovy, E., Gerber, L., Hermjakob, U., Lin, C. Y., & Ravichandran, D. (2001). Toward semantics-based answer pinpointing. In *Proceedings of the first international conference on Human language technology research* (pp. 1-7). Association for Computational Linguistics.
- [14] Huang, P., Bu, J., Chen, C., Qiu, G., & Zhang, L. (2007). Learning a Flexible Question Classifier. In *Convergence Information Technology, 2007. International Conference on* (pp. 1608-1613). IEEE.
- [15] Huang, P., Bu, J., Chen, C., & Qiu, G. (2007). An effective feature-weighting model for question classification. In *Computational Intelligence and Security, 2007 International Conference on* (pp. 32-36). IEEE.
- [16] Jia, K., Chen, K., Fan, X., & Zhang, Y. (2007). Chinese question classification based on Ensemble Learning. In *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on* (Vol. 3, pp. 342-347). IEEE.
- [17] Kang, X., Wang, X., & Ren, F. (2008). Exploiting syntactic and semantic information in coarse chinese question classification. In *Natural Language Processing and Knowledge Engineering, 2008. NLP-KE'08. International Conference on* (pp. 1-7). IEEE.
- [18] Li, X., & Roth, D. (2002). Learning question classifiers. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1* (pp. 1-7). Association for Computational Linguistics.
- [19] Li, X., & Roth, D. (2006). Learning question classifiers: the role of semantic information. *Natural Language Engineering*, 12(03), 229-249.
- [20] Loni, B., Khoshnevis, S. H., & Wiggers, P. (2011). Latent semantic analysis for question classification with neural networks. In *Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on* (pp. 437-442). IEEE.
- [21] Maclin, R., & Opitz, D. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*.
- [22] Mishra, M., Mishra, V. K., & Sharma, H. R. (2013). Question Classification using Semantic, Syntactic and Lexical features. *International Journal*.
- [23] Moldovan, D., Harabagiu, S., Pasca, M., Mihalcea, R., Girju, R., Goodrum, R., & Rus, V. (2000). The structure and performance of an open-domain question answering system. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics* (pp. 563-570). Association for Computational Linguistics.
- [24] Mollaei, A., Rahati-Quchani, S., & Estaji, A. (2012). Question classification in Persian language based on conditional random fields. In *Computer and Knowledge Engineering (ICCKE), 2012 2nd International eConference on* (pp. 295-300). IEEE.
- [25] Nguyen, T. T., Nguyen, L. M., & Shimazu, A. (2007). Improving the Accuracy of Question Classification with Machine Learning. In *Research, Innovation and Vision for the Future, 2007 IEEE International Conference on* (pp. 234-241). IEEE.
- [26] Purwarianti, et. al. (2006). Estimation of Question Types for Indonesian Question Sentence. Department of Information and Computer Sciences, Toyohashi University of Technology.
- [27] Rahman, M. A., & Scurtu, V. (2008). Performance maximization for question classification by subSet tree kernel using support vector machines. In *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on* (pp. 230-235). IEEE.
- [28] Ray, S. K., Singh, S., & Joshi, B. P. (2010). A semantic approach for question classification using WordNet and Wikipedia. *Pattern Recognition Letters*, 31(13), 1935-1943.
- [29] Skowron, M., & Araki, K. (2004). Evaluation of the new feature types for question classification with support vector machines. In *Communications and Information Technology, 2004. ISIT 2004. IEEE International Symposium on* (Vol. 2, pp. 1017-1022). IEEE.
- [30] Song, W., Wenyin, L., Gu, N., Quan, X., & Hao, T. (2011). Automatic categorization of questions for user-interactive question answering. *Information Processing & Management*, 47(2), 147-156.
- [31] Su, L., Liao, H., Yu, Z., & Zhao, Q. (2009). Ensemble learning for question classification. In *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on* (Vol. 3, pp. 501-505). IEEE.
- [32] Su, L., Yu, Z., Guo, J., Mao, C., & Liao, Y. (2011). Question classification using MultiBoost. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on* (Vol. 3, pp. 1831-1835). IEEE.
- [33] Sun, Y. L. (2010). Chinese question classification based on mining association rules. In *ICMLC* (pp. 413-416).
- [34] Sung, C. L., Day, M. Y., Yen, H. C., & Hsu, W. L. (2008). A template alignment algorithm for question classification. In *Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on* (pp. 197-199). IEEE.
- [35] Wang, L., Zhang, H., Wang, D., & Huang, J. (2009). Chinese Question Classification Based on Semantic Gram and SVM. In *Computer Science-Technology and Applications, 2009. IFCSTA'09. International Forum on* (Vol. 1, pp. 432-435). IEEE.
- [36] Wei, Z., Junjie, C., & Yanqing, N. (2011). Research on Chinese question classification based on hownet and dependency parsing. In *2011 3rd International Workshop on Intelligent Systems and Applications* (pp. 1-4).
- [37] Xia, L., Teng, Z., & Ren, F. (2008). Question classification in chinese restricted-domain based on SVM and domain dictionary. In *Natural Language Processing and Knowledge Engineering, 2008. NLP-KE'08. International Conference on* (pp. 1-6). IEEE.
- [38] Xu, B., Deng, Y., & Liu, J. (2008). Self-revise hierarchical classifier

based on SMO algorithm for Chinese question classification. In *Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference on* (pp. 1683-1687). IEEE.

- [39] Yu, Z., Su, L., Li, L., Zhao, Q., Mao, C., & Guo, J. (2010). Question classification based on co-training style semi-supervised learning. *Pattern Recognition Letters*, 31(13), 1975-1980.
- [40] Zhang, D., & Lee, W. S. (2003). Question classification using support vector machines. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 26-32). ACM.
- [41] Zhang, K., & Zhao, J. (2010). A Chinese question-answering system with question classification and answer clustering. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on* (Vol. 6, pp. 2692-2696). IEEE.
- Chinese question Classification using Multilevel Random Walk. In *Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on* (Vol. 3, pp. 515-519). IEEE.
- [42] Zheng-wen, X., & Shan-liang, P. (2011). Chinese Question classification based on multi strategy method. In *Transportation, Mechanical, and Electrical Engineering (TMEE), 2011 International Conference on* (pp. 1605-1609). IEEE.
- [42] Metzler, D & Croft, W, B. (2005). Analysis of statistical question classification for fact-based questions. *Inf. Retr.*, 8:481–504, May.
- [43] Pan, Y. Tang, Y. Lin, L. & Luo, Y. (2008). Question classification with semantic tree kernel. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '08*, pages 837–838, New York, NY, USA. ACM.
- [44] Silva, J. Coheur, L. Mendes, A and Wichert, A. (2011). From symbolic to sub-symbolic information in question classification. *Artificial Intelligence Review*, 35(2):137–154, February.