

Tantangan dan Peluang pada *Question Generation*

Wiwin Suwarningsih^{1,2}, Iping Supriana³, Ayu Purwarianti⁴
^{1,3,4} Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung
² Pusat Penelitian Informatika-LIPI
¹ e-mail: wiwin.suwarningsih@lipi.go.id;
³ e-mail: iping@informatika.org;
⁴ e-mail: ayu@informatika.org.

Abstrak

Pada makalah ini, kami melakukan survey beberapa penelitian yang membahas mengenai *question generation (QG)*. *QG* adalah sebuah teknik untuk membangkitkan pertanyaan yang berasal dari sebuah kalimat atau teks dalam bentuk bahasa alami. Kami mencoba menelaah garis besar konseptual *question generation* yang terdiri dari tiga kategori yaitu : berbasis sintaks, berbasis semantik, dan berbasis template. Sistem *question generation* dalam kategori sintaksis sering menggunakan unsur semantik dan sebaliknya. Sedangkan sistem yang berbasis template menggunakan beberapa tingkat sintaksis dan/atau informasi semantik. Hasil akhir dari survey ini adalah sebuah review berupa tantangan dan peluang dalam pengembangan penelitian di masa mendatang, yaitu berupa : (a) Tantangan pada isu semantik leksikal dan sintaktik, (b) penggunaan alternatif segitiga Vauquois, shallow parser dan (c) representasi sintaksis dengan struktur pohon frasa.

Kata kunci : *question generation*, leksikal, sintaksis, transformasi kalimat, segitiga Vauquois.

Abstract

In this paper, we reviewed the current state of the art in the *question generation (QG)*. *Question Generation (QG)* is the task of generating reasonable questions from a text or sentence of natural language. We attempted to examine the question of conceptual outline generation consisting of three categories: Syntax based, semantic-based and template-based. *Question generation* system in the syntactic category often uses semantic elements and vice versa. While the template-based system using multiple levels of syntactic and / or semantic information. The final results of this survey is a review in the form of challenges and opportunities in the development of future research, which are: (a) challenge on the issue of lexical semantic and syntactic, (b) the use of alternative Vauquois triangular, shallow parser, and (c) the syntactic representation phrase structure tree.

Key word : *question generation*, leksikal, sintaksis, transformasi kalimat, segitiga Vauquois

1. PENDAHULUAN

Salah satu masalah yang paling sulit dalam mengembangkan Sistem Tanya Jawab (STJ) adalah bagaimana membangkitkan pertanyaan (*question generation=QG*) yang berasal dari kalimat atau teks bahasa alami dan menemukan jawaban yang sesuai dengan query pertanyaan[1]. Untuk mengurangi kesulitan dalam pengembangan sistem tanya

jawab ini, banyak penelitian yang mengusulkan algoritma dan pendekatan baru yang secara aktif menggunakan kalimat dalam dokumen sebagai sumber pertanyaan dan sekaligus memberikan sebuah jawaban. Salah satu pendekatan yang dilakukan peneliti terdahulu adalah bagaimana pengolahan bahasa alami dengan menggunakan teknik yang secara otomatis untuk menghasilkan pertanyaan-pertanyaan (Kalady dkk [5]; Varga dan Ha [4]; Ali dkk [6]; Mannem dkk[7]).

Menurut Yao dkk[21], secara umum, sistem QG dapat membantu dalam bidang-bidang berikut : (1) *Intelligent tutoring systems*, QG dapat mengajukan pertanyaan berdasarkan materi pembelajaran untuk memeriksa prestasi peserta didik atau membantu mereka fokus dalam penelitian. QG juga dapat membantu tutor untuk mempersiapkan pertanyaan yang ditujukan untuk peserta didik atau mempersiapkan diri untuk pertanyaan potensial dari peserta didik; (2) *Closed-domain Question Answering (QA) systems*, beberapa sistem QA domain tertutup menggunakan standar pasangan pertanyaan-jawaban jawab untuk menyediakan layanan QA. Dengan menggunakan pendekatan QG sistem tersebut memungkinkan untuk eksplorasi ke domain lain; (3) *Natural language summarization/generation systems* : QG dapat membantu untuk menghasilkan pertanyaan, misalnya, Frequently Asked Questions (FAQ) dari sumber informasi yang diberikan dalam rangka memberikan daftar calon FAQ.

Garis besar konseptual QG terdiri dari tiga kategori yaitu : berbasis sintaks, berbasis semantik, dan berbasis template. Seperti penelitian yang dilakukan oleh Kalady dkk [5]; Varga dan Ha [4] dan Ali dkk [6], mengusung metode berbasis sintaks untuk melakukan QG. Meskipun masing-masing upaya ini berbeda pada beberapa rincian, mereka telah mengikuti strategi dasar yang sama, yaitu : mengurai kalimat menggunakan parser sintaksis, menyederhanakan kalimat kompleks, mengidentifikasi frase kunci, dan menerapkan aturan transformasi sintaksis dan penggantian kata pertanyaan. Mannem dkk [7], memperkenalkan sistem QG berbasis semantik yang menggabungkan *semantic role labeling* (SRL) dengan transformasi sintaksis. Pada tahap seleksi konten, satu kalimat pertama diurai dengan peran Labeler semantik untuk mengidentifikasi target potensial. Target dipilih menggunakan kriteria seleksi sederhana.

Salah satu kelemahan dari metode berbasis sintaks dan semantik adalah bahwa mereka menghasilkan pertanyaan dengan menata ulang bentuk permukaan kalimat (*surface-form sentences*) [3][4][7]. QG berbasis template menawarkan kemampuan untuk mengajukan pertanyaan dengan kata-kata yang tepat dari sumber teks. Template yang dimaksud adalah teks yang telah ditetapkan dengan variabel diganti dengan isi dari teks sumber. QG berbasis template memungkinkan sistem meningkatkan keahlian manusia di *language generation*.

Sedangkan penelitian QG berbasis template dilakukan oleh Cai dkk [8], mengusulkan sistem QG berbasis template dengan menggunakan *Natural Language Generation Markup Language* (NLGML), bahasa yang dapat digunakan untuk

menghasilkan tidak hanya pertanyaan tapi ekspresi bahasa alami. Lindberg dkk [3], pendekatan yang digunakan adalah metode berbasis template dengan menggunakan informasi yang didominasi semantik, agar dapat digunakan untuk menghasilkan pertanyaan bahasa alami dalam sistem pembelajaran on-line

Pendekatan lain yang digunakan untuk QG adalah *Hidden Markov Model* (HMM) dan *Latent Dirichlet Allocation* (LDA) model, seperti yang dilakukan oleh Ajmera dkk[1], mengusulkan teknik untuk mengeksploitasi informasi yang berasal dari artikel berupa kasus nyata tentang kekhawatiran konsumen terhadap suatu masalah produk yang dikeluarkan oleh suatu perusahaan. Dari artikel ini Ajmera dkk mencoba memetakan pasangan pertanyaan dan jawaban dengan cara membangkitkan arti dari artikel tersebut.

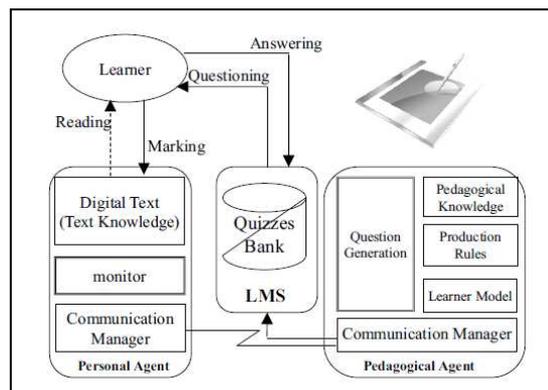
Beberapa penelitian lain yang mengimplementasikan QG pada system pembelajaran siswa di antaranya adalah Bednarik dan Kovács[9], membuat algoritma baru yang dapat melakukan pengelompokan atau klasifikasi sesuai pengaturan parameter dalam kasus dokumen yang berisi ribuan kalimat. Waktu yang dibutuhkan untuk menjawab pertanyaan berdasarkan karakter dokumen, memungkinkan menghasilkan pertanyaan untuk setiap kalimat dalam dokumen. Pada penelitian ini Bednarik dan Kovács[9] membandingkan pula QG otomatis dengan QG manual dengan pengguna terdiri dari siswa yang diajarkan subyek tertentu, siswa yang tidak diajarkan subyek tertentu dan pakar. Dari hasil ini menunjukkan rasio jawaban yang benar dihasilkan dari pendekatan secara otomatis daripada secara manual. Hasil penelitian juga menunjukkan bahwa orang yang diuji berhasil memberikan jawaban yang lebih tepat untuk pertanyaan ilmu linguistik. Yahya[11], mendesain, implementasi, dan evaluasi *Ontologi Question* (OntoQue), sistem QG yang membantu instruktur dalam menghasilkan penilaian otomatis dengan menggunakan domain ontologi. Domain ontologi adalah struktur representasi pengetahuan dengan model tertentu yang menyediakan representasi entitas sesuai dengan domain yang digunakan. Entitas pada domain ini terdiri dari kelas, properti obyek, properti data, dan individu. Kelas merupakan set individu, objek dan data properti mewakili hubungan antara individu dalam domain, dan individu mewakili objek yang sebenarnya dari domain tersebut. Entitas Domain diterjemahkan ke dalam aksioma ontologis yang merupakan sumber daya dari item penilaian yang didapat secara otomatis.

Berdasarkan hal tersebut diatas, kami mencoba melakukan *review* kecil atas beberapa pendekatan dan teknik yang telah digunakan untuk mengimplementasikan QG. Hasil dari *review* ini berupa tantangan dan peluang yang sedang muncul, seperti misalnya penggunaan koleksi data, pengkombinasian bermacam teknik, dan juga kemungkinan pengusulan pendekatan atau metoda baru.

Secara keseluruhan, paper ini diorganisasikan kedalam bagian-bagian berikut ini : bagian 2 menyajikan kerangka kerja dari QG. Bagian 3 menjelaskan dan membahas tantangan dari QG. Bagian 4 menjelaskan peluang QG untuk domain bahasa Indonesia. Bagian 5 menjelaskan kesimpulan dari hasil survey dan review.

2. KERANGKA KERJA QUESTION GENERATION

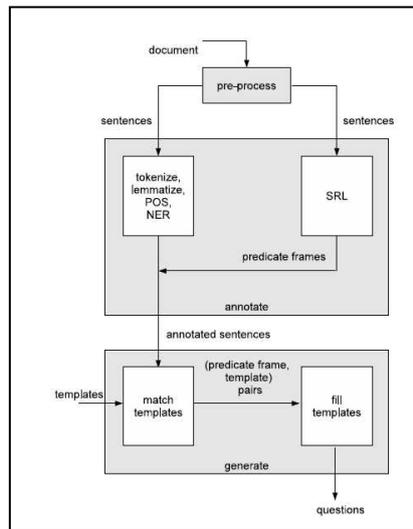
Salah satu tujuan utama pembangunan *QG* adalah untuk menyediakan arsitektur yang berisi tahapan kerja dan komponen yang akan dikembangkan untuk pengelolaan dokumen yang digunakan dalam bahasa tertentu. Kerangka kerja *QG* yang diperkenalkan oleh Iwane dkk [2] terdiri dari template dan grafik pengetahuan (lihat gambar 1). Kerangka kerja ini menjamin untuk menutupi seluruh pengetahuan dalam buku teks dan mempromosikan kepada siswa system pembelajaran terpusat. Dalam kerangka ini grafik pengetahuan dan template, menganggap bahwa buku teks adalah teks ekspositori yang terdiri dari jawaban atas pertanyaan dan penjelasan dari pengetahuan baru kepada peserta didik. Jadi seorang pelajar harus mengidentifikasi pertanyaan dan jawaban untuk menyatakan pertanyaan-pertanyaan mereka sendiri dan jawaban dalam teks.



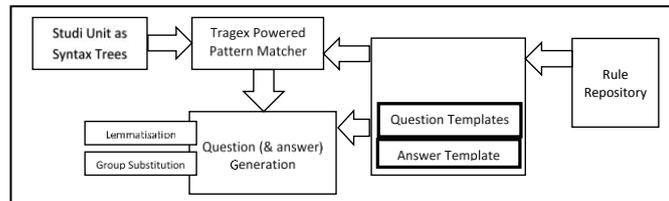
Gambar 1. Kerangka *question generation* Iwane dkk [2]

Berbeda dengan Lindberg dkk[3], salah satu hal yang paling penting untuk mengamati tentang arsitektur *QG* adalah bahwa template adalah input eksternal (lihat gambar 2). Mereka sama sekali tidak digabungkan ke sistem dan dapat dimodifikasi sesuai kebutuhan tanpa adanya modifikasi sistem. Lindberg dkk[3] melakukan sedikit pra-pengolahan. Metode berbasis sintaks khususnya telah termotivasi untuk melakukan penyederhanaan kalimat, karena metode ini lebih mungkin untuk menghasilkan pertanyaan berarti dari kalimat yang pendek dan ringkas.

Aplikasi *automated question generation* (AQG) yang dibangun oleh Bednarik dan Kovács[10], ada beberapa pendekatan di arsitektur yang dibangun untuk diimplementasikan. Salah satu deskripsi pertama dari kerangka dasar sistem AQG terdiri dari tiga bagian utama yaitu : (1) ekstraksi kalimat dengan menggunakan metode pembelajaran preferensi; (2) estimasi posisi kosong dengan bersyarat dengan pendekatan random field; (3) pembangkitan kandidat kata untuk posisi kosong dengan metode pencocokan pola statistik; (4) pengurangan dari himpunan kandidat kata.

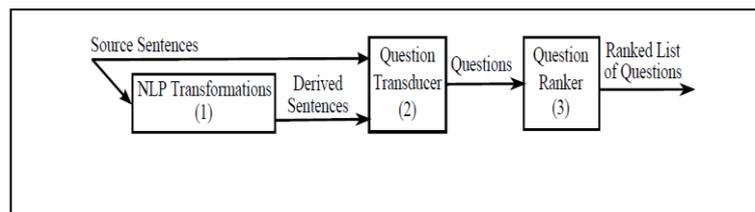


Gambar. 2. Kerangka kerja *question generation* Lindberg dkk[3]



Gambar 3. Kerangka kerja Automated question generation (AQG) Bednarik dan Kovács[10].

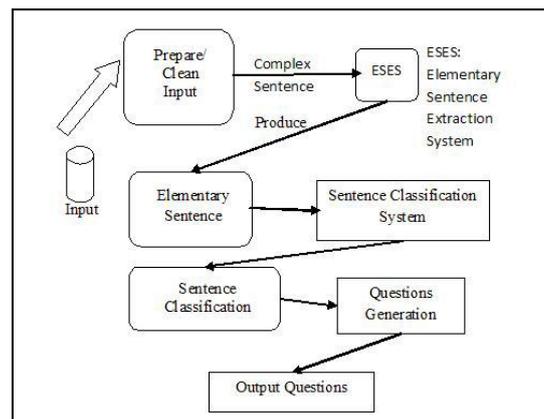
Sedangkan AOG yang dibangun oleh Heilman dan Smith[20] adalah mendefinisikan kerangka kerja untuk menghasilkan satu himpunan peringkat pertanyaan berdasarkan fakta tentang teks yang berasal dari artikel. Dari himpunan ini, pertanyaan-pertanyaan peringkat utama diberikan kepada pengguna yaitu guru untuk menyaring dan merevisi, atau langsung ke siswa untuk digunakan sebagai latihan. Kerangka kerja AQG dapat digunakan untuk dekomposisi menjadi tiga tahap modular yang digambarkan pada Gambar 4. Banyak pertanyaan yang berguna dapat dilihat sebagai transformasi leksikal, sintaksis, atau semantik dari kalimat deklaratif dalam teks. Pada tahap 1, yang dipilih kalimat atau himpunan kalimat dari teks berubah menjadi satu kalimat deklaratif secara opsional untuk mengubah item leksikal, struktur sintaksis, dan semantik. Transformasi pemrosesan bahasa alami dimanfaatkan untuk peringkasan ekstraktif, kompresi kalimat, membelah kalimat, dan substitusi kata. Pada tahap 2, kalimat deklaratif berubah menjadi pertanyaan dengan mengeksekusi satu himpunan sintaksis yang terdefinisi dengan proses transformasi.



Gambar 4. Kerangka Kerja tiga tahap Heilman dan Smith[20]

Kontribusi utama dari kami Heilman dan Smith [20] adalah deskripsi dan evaluasi dari tahap implementasi kerangka kerja yang meliputi : (1) dirancang untuk menjadi domain-umum, (2) menggunakan komposisi aturan extensible untuk mengubah kalimat deklaratif menjadi pertanyaan, (3) secara eksplisit menangani fenomena linguistik yang relevan, (4) modular, memanfaatkan *tool* dari pemrosesan bahasa alami dan (5) mencakup komponen belajar untuk peringkat pertanyaan yang dihasilkan dengan berbagai kriteria

Kerangka kerja AQG yang dikembangkan oleh Ali dkk[6], dibagi menjadi tiga tahapan (lihat gambar 5) yaitu : (1) tahap *preprocessing*, tahap ini merupakan proses pembersihan dan pengolahan data mentah dengan cara menghapus tag berlebihan dari semua dokumen dan pertanyaan-pertanyaan yang diberikan dalam dataset TREC; (2) Konstruksi kalimat dasar, pada tahap ini dilakukan proses mengekstrak kalimat dasar dari kalimat-kalimat kompleks dengan maksud untuk menghasilkan pertanyaan yang lebih akurat. Sintaksis diuraikan dari setiap kalimat kompleks untuk mencapai hal ini. Metoda yang digunakan adalah Charniak parser untuk membangun sebuah representasi pohon sintaksis dan mengurai representasi kalimat; (3). Klasifikasi kalimat dan pembangkitan pertanyaan, sebagai masukan pada modul ini adalah kalimat dasar yang kemudian berdasarkan hubungan informasi antara *Part of speech* (POS) dan *Name of entity* (NE) *tagged* diambil subyek, obyek, preposisi dan kata kerja dari setiap kalimat dasar untuk digunakan proses klasifikasi kalimat. Dari proses klasifikasi kalimat diperoleh lima *coarse classes* yaitu *Human* (nama orang yang berkedudukan sebagai subyek atau obyek), *Entity* (termasuk binatang, tanaman, gunung dan obyek lainnya), *Location* (kata yang merepresentasikan lokasi seperti kota, negara, sekolah dan lain-lain), *Time* (kata yang merepresentasikan waktu, tanggal atau periode seperti tahun, hari Selasa, jam 9 pagi dan lain-lain) dan *Count* (semua benda yang dapat dihitung dan memiliki ukuran seperti jumlah laki-laki ada 10 orang).



Gambar 5. Kerangka kerja AQG Ali dkk[6]

Berdasarkan hasil analisa terhadap lima kerangka kerja diatas, maka dapat disimpulkan bahwa tahapan dalam QG meliputi (1) *Preprocessing* (ekstraksi kamilat untuk pendefinisian POS dan NE), (2) Formulasi pertanyaan (dapat dilakukan dengan klasifikasi kalimat atau pencocokan pola/model kalimat), (3) Pemilihan kandidat pertanyaan.

3. TANTANGAN DALAM QUESTION GENERATION

Pada bagian ini, kami menyajikan diskusi tentang tantangan dan peluang penelitian di bidang komputasi dan linguistik untuk menghasilkan pertanyaan secara otomatis yang berasal dari kalimat atau teks bahasa alami. Tujuan yang harus dicapai menurut Linberg dkk[3], pada proses *question generation* adalah (1) *content selection* : memilih sumber teks (biasanya kalimat tunggal) yang dapat menghasilkan pertanyaan; (2) *target identification* : menentukan kata dan/atau frase tertentu yang dijadikan sebagai kata tanya ; (3) *question formulation* : menentukan pertanyaan yang tepat yang disesuaikan dengan identifikasi konten; (4) *surface form generation* : memproduksi realisasi bentuk akhir dari *surface-form*. Berdasarkan keempat hal diatas, tantangan yang mendasar dalam mengembangkan QG secara umum akan diuraikan berikut ini.

A. Tantangan Pada Isu Semantik Leksikal

Pada bagian ini akan dibahas isu-isu di QG terkait dengan semantik leksikal (yaitu, arti kata dan frase pendek). Pada teks Bahasa Inggris pertanyaan selalu diawali dengan 4 WH [1][5][6] yaitu *Who* (untuk pertanyaan yang menerangkan orang), *Where* (untuk pertanyaan yang menunjukkan tempat), *When* (untuk pertanyaan yang berhubungan dengan waktu) dan *What* (untuk pertanyaan yang berhubungan dengan benda).

Menurut Heilman[19], isu pertama adalah untuk dapat membuat peta yang cukup handal dan benar dari jenis semantik *WH-word* dan untuk mengidentifikasi semantik tingkat tinggi maka diperlukan pengetahuan leksikal substansial. Isu kedua adalah benar mengidentifikasi jenis semantik tidak selalu cukup. Pertanyaan sederhana yang

menggunakan *What* (apa) atau *Who* (siapa) terkadang tidak jelas ketika pertanyaan diambil dari konteksnya. Tantangan lain dalam QG adalah bahwa, idealnya pertanyaan harus menunjukkan beberapa variasi dari teks input. Juga, sebagai alternatif untuk menghasilkan berbagai susunan kata untuk pertanyaan tertentu. Linberg[3], menyatakan bahwa dua kalimat dapat mengekspresikan hubungan semantik yang sama antara dua konsep sebab akibat (contoh fakta bahwa panas yang terperangkap menyebabkan suhu bumi meningkat). Dalam satu kasus, penyebab ini dinyatakan dalam sebuah frase kata sifat, sementara yang lain menggunakan frase preposisional kalimat-awal.

B. Tantangan pada Isu Pembentukan Sintaktik

Bagian ini membahas beberapa tantangan utama dalam QG yang terkait dengan struktur sintaksis bahasa alam. Untuk menghasilkan pertanyaan, sistem QG perlu model beberapa kompleksitas sintaksis bahasa alami : (1) mengidentifikasi batas-batas frase [4][7][8] (misalnya, terdapat sebagian kata yang merupakan bagian dari sebuah jawaban), (2) mengidentifikasi predikat dan argumen [7][9][11], dan (3) ekstrak informasi dari struktur bersarang seperti klausa relative[11][13][14]. Bahkan jika parser mampu memberikan representasi akurat dari struktur sintaksis kalimat, penggalan informasi dapat mengubah struktur yang trivial. Bahkan dengan kalimat kompleks sekalipun kita dapat membangkitkan pertanyaan dengan menggunakan berbagai konstruksi sintaksis untuk menyajikan beberapa fakta individu[14].

Suatu hal yang penting di QG adalah bagaimana untuk menghasilkan pertanyaan singkat dari kalimat kompleks. Salah satu solusi yang mungkin adalah dengan menggunakan teknik kompresi dari kalimat [20][10]. Tujuan dari kompresi adalah untuk mengambil kalimat panjang dan rumit sebagai masukan, dan menghasilkan keluaran berupa versi singkat dan tunggal dimana hasil ini dapat menyampaikan sepotong informasi utama. Namun, kalimat sering menyampaikan beberapa potongan informasi. Di QG, dimungkinkan untuk menghasilkan pertanyaan bukan hanya tentang informasi dalam klausa utama, tetapi juga tentang informasi tertanam dalam berbagai konstruksi bersarang[18][20].

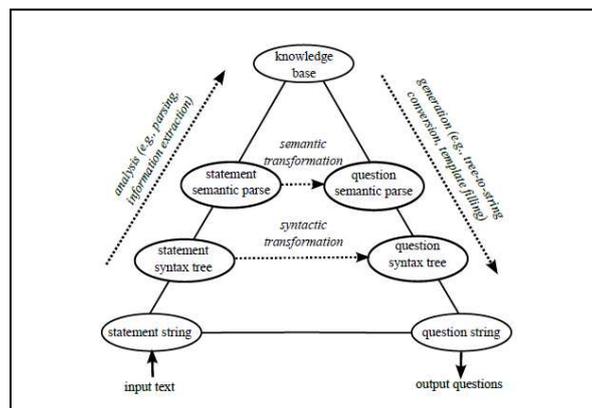
4. PELUANG *QUESTION GENERATION* UNTUK DOMAIN BAHASA INDONESIA

Representasi sistem QG dalam domain bahasa Indonesia masih sangat luas. Salah satu alternatif dapat mengandalkan struktur frase pohon sintaks sebagai representasi utama dari teks input. Pendekatan sintaksis ini tentu saja bukan satu-satunya pendekatan yang mungkin untuk QG. Pada bagian ini, kami akan menjelaskan berbagai representasi serta potensi metoda yang lain, dan menggambarkan bagaimana dengan domain Bahasa Indonesia bisa menggunakan pendekatan QG dengan representasi alternatif.

C. Analogi Segitiga Vauquois

Sebuah analogi yang berguna dapat dihasilkan dari system kerja mesin penterjemah[12][13], salah satunya adalah bagaimana membangkitkan dari teks ke teks. Penelitian yang telah dilakukan dan diekplorasi diantaranya adalah model-model penerjemahan berdasarkan berbagai jenis representasi linguistik, termasuk yang menggunakan pendekatan berbasis shallow-word dan frase yang memanfaatkan dataset besar untuk mempelajari korespondensi antara urutan tanda [12], pendekatan sintaksis didasarkan pada struktur kalimat atau ketergantungan pohon [13][14][15].

Sebuah visualisasi umum dari pendekatan berbasis mesin pembelajaran adalah Segitiga Vauquois [16], ditunjukkan pada Gambar 4. Dalam kasus ini input dan output yang diwakili di sudut bawah segitiga adalah kalimat dalam satu bahasa. Untuk QG, input teks, dan output adalah seperangkat pertanyaan. Dimensi vertikal segitiga merupakan kedalaman analisis linguistik yang diperlukan untuk analisis teks dan generasi, dan dimensi horizontal menunjukkan jumlah pengolahan yang diperlukan untuk transfer dari input di bagian kiri bawah ke output di bagian kanan bawah. Tingkat vertikal yang berbeda dalam segitiga tersebut merupakan representasi bahasa, dan tingkat yang lebih tinggi menunjukkan tingkat yang lebih besar dari abstraksi. Jarak horizontal dan vertikal kira-kira merupakan tahapan untuk melakukan operasi seperti parsing, transformasi dan pembangkitan (*generation*). Sebagai contoh, representasi semantik akan membuat parsing input ke representasi (yaitu, bergerak jauh di atas segitiga) relatif sulit, dan mungkin akan menghasilkan output (yaitu, bergerak kembali ke segitiga) juga relatif sulit, tetapi juga akan mungkin membuat transformasi antara input dan output (yaitu, bergerak di bagian atas segitiga) relatif mudah. Sebaliknya, representasi linguistik dangkal umumnya akan memiliki biaya rendah untuk parsing dan pembangkitan (*generation*), tetapi biaya tinggi untuk transfer (misalnya, karena kasus yang lebih khusus harus ditangani).



Gambar 4. Segitiga Vauquois[16]

D. Shallow Parser

Pendekatan lain untuk QG akan yaitu menggunakan *Shallow parser* [17] atau alat serupa untuk mengekstrak representasi *shallower syntactic*[18]. Pendekatan seperti ini akan memiliki langkah parsing sederhana yaitu dari proses parsing sintaksis dan melakukan proses similaritas dari proses pembangkitan teks. Namun, langkah pengalihan (yaitu, mengubah kalimat deklaratif kompleks menjadi pertanyaan) akan lebih sulit pada umumnya. Menggunakan *Shallow parser* untuk mengekstrak pertanyaan dari berbagai konstruksi linguistik kemungkinan akan memerlukan penanganan sejumlah besar kasus-kasus khusus.

Dengan representasi semantik, analisis kalimat masukan akan lebih menantang karena parser semantik biasanya lebih rentan terhadap kesalahan daripada parser sintaksis[19]. Transformasi representasi semantik ke representasi pertanyaan akan menjadi sederhana jika representasi ini cocok dengan kebutuhan QG, tetapi representasi semantik mungkin terlalu abstrak karena memerlukan berbagai template untuk menghasilkan pertanyaan[20].

E. Representasi Sintaksis dengan Struktur Pohon Frasa

Salah satu bidang yang menarik untuk pekerjaan di masa depan adalah eksplorasi representasi alternatif untuk QG. Pendekatan potensi untuk faktual QG yang dapat diandalkan adalah representasi sintaksis dengan menggunakan struktur pohon frasa[19]. Heilman[19], mengidentifikasi beberapa tantangan utama yang tersisa mungkin akan relevan terlepas dari representasi yang digunakan oleh sistem QG, dan membahas bagaimana tantangan-tantangan ini bisa diatasi.

Penelitian lain yang dilakukan oleh Heilman dan Smith [20], menerapkan aturan untuk mengubah kalimat sumber ke pertanyaan, dengan menggunakan *Tregex* (bahasa query pohon) dan *Tsurgeon* (bahasa manipulasi pohon). Bahasa *Tregex* mencakup berbagai operator relasional berdasarkan hubungan primitif dominasi langsung (dilambangkan "<") dan diutamakan langsung (dilambangkan ":"). Selain mendukung query yang melibatkan hubungan pohon standar, *Tregex* juga mencakup beberapa kendala termotivasi bahasa seperti kepemimpinan, dominasi dibatasi, dan didahulukan (misalnya, dominasi melalui rantai frase kata kerja). *Tsurgeon* menambahkan kemampuan untuk memodifikasi pohon dengan relabeling, menghapus, memindahkan, dan memasukkan node. Dengan menggunakan *Tregex*, pertanyaan dapat ditemukan dengan node yang sesuai dengan pola; node ini kemudian dinamai dan dimanipulasi menggunakan operasi *Tsurgeon*.

5. KESIMPULAN

Makalah survey ini dibuat untuk menyediakan layanan bagi komunitas ilmiah. Pada proses survey kami mencoba meringkas dan melakukan review kecil dari setiap hasil penelitian terbaru dengan cara mengintegrasikan dan menambahkan pemahaman untuk bekerja di bidang *question generation* (QG). Review ini mencoba menekankan klasifikasi literatur yang ada, mengembangkan perspektif di ranah tertentu dan mengevaluasi tren yang sedang berkembang. Namun, karena tidak mungkin untuk memasukkan semua atau bahkan sebagian besar penelitian sebelumnya, survei ini hanya mencoba memunculkan kegiatan progresif dan menunjukkan keterbatasan yang ditemukan dan harus dihadapi oleh peneliti lain.

Ucapan Terima Kasih

Kami berterima kasih kepada para *reviewer* atas komentar yang sangat membantu.

Daftar Pustaka

- [1] Jitendra Ajmera, Sachindra Joshi, Ashish verma, Amol Mittal, “Automatic Generation of Question Answer Pairs From Noisy Case Logs”, 2014 IEEE 30th International Conference on Data Engineering (ICDE), pp : 436-447, 2014.
- [2] Noriyuki Iwane, Chunming Gao, Makoto Yoshida, “Question Generation for Learner Centered Learning”, 2013 IEEE 13th International Conference on Advanced Learning Technologies, pp : 330-332, 2013.
- [3] David Lindberg, Fred Popowich, John Nesbit Phil Winne, “Generating Natural Language Questions to Support Learning On-Line”, 14th European Workshop on Natural Language Generation, pages 105–114, Bulgaria, August 2013, Association for Computational Linguistics 2013
- [4] Andrea Varga and Le An Ha. “Wlv: A question generation system for the qgstec 2010 task b”. In Proceedings of QG2010: The Third Workshop on Question Generation, pages 80–83. 2010
- [5] Saidalavi Kalady, Ajeesh Elikkottil, and Rajarshi Das. “Natural language question generation using syntax and keywords”. In Proceedings of QG2010 : The Third Workshop on Question Generation, pages 1–10. 2010
- [6] Husam Ali, Yllias Chali, and Sadid A Hasan. “Automation of question generation from sentences”. In Proceedings of QG2010: The Third Workshop on Question Generation, pages 58–67. 2010.
- [7] Prashanth Mannem, Rashmi Prasad, and Aravind Joshi. “Question generation from paragraphs at upenn: Qgstec system description”. In Proceedings of QG2010: The Third Workshop on Question Generation, pages 84–91. 2010.
- [8] Zhiqiang Cai, Vasile Rus, Hyun-Jeong Joyce Kim, Suresh C. Susarla, Pavan Karnam, and Arthur C. Graesser. Nlqml: A markup language for question generation. In Thomas Reeves and Shirley Yamashita, editors, Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2006, pages 2747–2752, Honolulu, Hawaii, USA, October. 2006.

- [9] László Bednarik and László Kovács, "Implementation and assessment of the automatic question generation module", 3rd IEEE International Conference on Cognitive Infocommunications, Slovakia, pages 687-690, 2012.
- [10] Laszlo Bednarik and Laszlo Kovacs, "Automated EA-type Question Generation from Annotated Texts", 7th IEEE International Symposium on Applied (omputational Intelligence and Informatics 2012, Romania, pages 191-195, 2012.
- [11] Maha Al-Yahya, "OntoQue: A Question Generation Engine for Educational Assesment Based on Domain Ontologies", 2011 11th IEEE International Conference on Advanced Learning Technologies, pages : 393-395, 2011.
- [12] Koehn, P., Och, F. J., and Marcu, D. "Statistical phrase-based translation". In Proc. of HLTNAACL, pages 127–133, 2003.
- [13] Quirk, C., Menezes, A., and Cherry, C. "Dependency treelet translation: Syntactically informed phrasal SMT". In Proc. of ACL, 2005
- [14] Graehl, J., Knight, K., and May, J. "Training tree transducers". Journal of Computational Linguistics, 34.2008.
- [15] Gimpel, K. and Smith, N. A. "Feature-rich translation by quasi-synchronous lattice parsing". In Proc. of EMNLP. 2009.
- [16] Jurafsky, D. and Martin, J. H. "Speech and Language Processing". Prentice Hall, chapter 25, 2nd edition. 2008.
- [17] Sha, F. and Pereira, F. "Shallow parsing with conditional random fields". In Proc. of HLTNAACL. 2003.
- [18] Mitkov, R. and Ha, L. A. "Computer-aided generation of multiple-choice tests". In Proc. of the HLT-NAACL workshop on Building educational applications using natural language processing. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73. 2003
- [19] Michael Heilman, "Automatic Factual Question Generation from text", thessis for Doctor of Philosophy at Carnegie Mellon University. 2011
- [20] Michael Heilman and Noah A. Smith, "Question Generation via Overgenerating Transformations and Ranking", www.lti.cs.cmu.edu. 2009