

**RANCANGAN
PENERJEMAH BAHASA INGGRIS - INDONESIA
MENGUNAKAN TEKNIK KECERDASAN BUATAN**

Yohanes Bowo Widodo

yb_widodo@hotmail.com

ABSTRAK

Natural Language processing is a topic intensely discussed in computer science nowadays. The term natural language is used to represent common human language because in computer science there are other classifications of language, such as programming language, formal language, etc. Researches in natural language area have given the computer ability to understand human language (in a certain level) and to translate between one natural language to another natural language. There are many unstructured aspects should be considered so that good rules could be applied by the computer in order to make good translation or understanding. These unstructured aspects could not be solved by mere conventional computer science concepts.

There are new concepts in computer science could be used to gain good unstructured problem solving. These concepts are artificial intelligence, fuzzy logic, object oriented, hypertext, and neural networks. The growth of hardware capability also giving good support to the unstructured problems found in natural language. This research is intended to analyze and to apply recent concepts discussed above, to solve natural language processing problems, especially in natural language translation, in this case, English to Indonesian

Kata Kunci : *Language, Strutural Language, Unstruktural language*

PENDAHULUAN

Dalam perkembangan era teknologi ternyata menguasai bahasa untuk kebutuhan adalah hal yang penting, sehingga penggunaan bahasa untuk mendukung penulisan format tulisan.

Permasalahan pada Pemrosesan Bahasa Alami

Sebelum melakukan perancangan pemroses bahasa alami, berikut ini akan dibahas beberapa permasalahan bahasa alami yang sering muncul.

Ambiguitas

Bahasa alami dalam bentuk tertulis tersusun dari bagian bagian yang cukup teratur, dengan kombinasi tertentu, sehingga dapat dicari aturan sintaksnya.

Terkadang ada leblh dari satu struktur sintaks

yang cocok dengan suatu kalimat, sehingga terjadi apa yang disebut ***ambiguitas***. *Ambiguitas* merupakan suatu masalah kompleks yang harus ditangani secara benar dalam pengelolaan bahasa alami. Salah satu cara untuk menuntun komputer agar dapat melakukan analisa sintaks secara benar adalah dengan menambahkan aturan yang lebih rinci, yang disebut ***restrictions***. *Restriction* digunakan bukan hanya untuk memilih hasil analisa yang benar tetapi juga untuk membatasi alternatif analisa yang harus dibuat. Sebagai contoh, kalimat *This morning mother read a book*. Kalimat tersebut dapat menghasilkan pengertian 'morning mother' yang tidak diinginkan. Ambiguitas semacam itu dapat menyebabkan pemborosan waktu yang sangat signifikan dalam analisa sintaks jika tidak ditangani dengan menambahkan *restriction*.

Makna Tunggal

Selain kenyataan bahwa suatu kalimat dapat bermakna ganda, ternyata kasus sebaliknya terjadi juga, yaitu untuk mengungkapkan suatu kenyataan yang sama dapat digunakan beberapa kalimat yang berbeda.

Contoh:

Saya lahir pada tanggal 11 November
Hari ulang tahun saya adalah 11
November

Bentuk 'Zeroed'

Selain dan penanganan kata-kata yang benar-benar muncul dalam kalimat, pemroses bahasa alami juga harus berurusan dengan fenomena dimana suatu kata berada dalam bentuk "zeroed", yaitu secara fisik tidak ada tetapi dapat direkonstruksi dari konteks kalimat. Misalnya *They agree and we do too* dalam contoh tersebut terdapat pengertian *agree* setelah kata *do*, dimana kata tersebut tidak benar-benar tertulis tetapi dapat diturunkan dari konteks kalimat. Untuk melengkapi analisa, dan untuk mendapatkan makna sesungguhnya dari suatu kalimat, maka kata yang "zeroed" harus dapat dicari.

Perbedaan Context

Suatu ekspresi yang sama dapat memiliki arti yang berbeda dalam konteks yang berbeda. Contoh:

Where's the water ? (dalam laboratorium kimia, air yang dimaksudkan adalah air murni)

Where's the water ?

(jika diucapkan oleh seorang yang sedang haus, air yang dimaksudkan adalah air minum, bisa teh, air mineral, sirup, dll)

Perkembangan Bahasa

Tidak ada program pengelola bahasa alami yang dapat menyelesaikan semua permasalahan dengan tepat, karena kata baru, ekspresi baru, dan artinya dapat berkembang sangat bebas.

Contoh :

I'll Fax it to you

PEMBASAHAN

Bagian ini akan membahas keterbatasan-keterbatasan dan peluang-peluang pada pemrosesan bahasa alami baik yang berkaitan dengan karakteristik bahasa alami maupun berkaitan dengan teknologi komputer yang ada pada saat ini.

Daya Ingat Komputer

Meskipun hingga saat ini para ahli masih beranggapan bahwa penterjemahan oleh manusia masih lebih luwes dibandingkan dengan penterjemahan oleh komputer, tetapi ada satu karakteristik komputer yang memberi harapan terhadap permasalahan bahasa alami yaitu daya ingat (memori) komputer. Komputer dapat menyimpan berjuta-juta informasi tentang suatu kata dan terjemahannya secara konsisten, dan dapat ditambah setiap saat yang diinginkan. Dilain pihak kemampuan dan daya ingat manusia relatif tidak konsisten sehingga ada kemungkinan kekeliruan atau lupa mengartikan suatu kata.

Metoda Baru dalam Ilmu Komputer

Belakangan ini Ilmu Komputer telah maju sangat pesat. Banyak metoda-metoda baru diteliti dan dikembangkan untuk mendukung permasalahan yang semakin rumit. Metoda-metoda tersebut antara lain adalah *Object Oriented Programming Paralel Processing, Jaringan Syaraf Tiruan, Fuzzy Logic, Hypertext*, dan lain-lain. Metoda-metoda tersebut memberikan kemungkinan baru bagi penyelesaian permasalahan yang rumit.

Keteraturan Tatabahasa

Walaupun banyak sifat bahasa alami yang tidak teratur (*unstructured*), beberapa kalimat yang kita jumpai sehari-hari banyak juga yang memiliki keteraturan, terutama

bahasa resmi (baku), sehingga dapat dicari aturannya dan disimpan dalam basis pengetahuan. Bahkan kenyataan tersebut dapat memberikan peluang baru, yaitu program bantu penyuntingan teks. Program tersebut dapat membantu mengatur suatu teks yang tidak tertulis dalam bentuk standar, yang sulit dimengerti, menjadi teks lain yang standar, lebih mudah dimengerti, tanpa menghilangkan arti semula.

Pemahaman Bahasa Alami

Pemrosesan Bahasa Alami merupakan topik yang pada saat ini menarik minat peneliti di beberapa bidang ilmu. Ahli komputer, pengajar, psikolog, ahli linguistik, *information specialist*, administrator medis, ahli pengolahan file, dan orang yang non spesialis mempertanyakan seberapa jauh komputer dapat menafsirkan bahasa alami atau bahasa manusia. Jelas bahwa banyak hal yang dapat dikerjakan jika komputer dapat di program untuk mengekstrak isi suatu pesan yang tersimpan dalam bentuk bahasa alami.

Sebagai contoh, jika seorang pemrogram dapat memberi instruksi dalam bahasa Indonesia untuk menyatakan apa yang dia inginkan dan komputer dapat memberikan interaksi yang tepat, maka dengan mudah dapat tercipta suatu program yang menangani apa yang diinginkan oleh pemrogram, tanpa perlu menuliskan program tersebut dalam bahasa komputer (Tingkat Tinggi maupun Tingkat Rendah).

Proses penterjemahan suatu teks bahasa alami menjadi teks bahasa alami yang lain tidak terlepas dari masalah pemahaman teks sumber. Sebelum teks sumber diterjemahkan menjadi teks sasaran, teks sumber tersebut harus 'dipahami' dulu dengan benar sehingga hasil terjemahan yang diperoleh menjadi optimal.

Pemahaman Teks.

Dalam komunikasi sehari-hari, manusia menggunakan bahasa dalam bentuk

rangkaian kalimat yang saling berkaitan. Rangkaian kalimat tersebut membentuk satu kesatuan pikiran yang dikomunikasikan. Teks merupakan deretan kalimat yang memiliki kaitan makna tertentu di antara kalimat-kalimatnya.

Di sini terlihat bahwa dalam membicarakan teks kita harus memahami segala sesuatu yang berhubungan dengan kalimat. Definisi kalimat adalah 'bagian terkecil dari ujaran atau teks yang mengungkapkan pikiran yang utuh secara ketatabahasaan'. Teks baru mungkin terbentuk jika ada kalimat yang letaknya

Rancangan Penerjemah Bahasa
(Yohanes Bowo W.)

berurutan dan berdasarkan kaidah tertentu. Dengan demikian pengenalan kalimat dalam teks hendaknya dimulai dari bagian awal teks, atau setidaknya dimulainya dari bagian awal setiap alinea.

Dan segi sintaksis, kalimat terdiri atas bagian inti dan bukan inti. Bagian kalimat yang tidak dapat dihilangkan tanpa mengubah makna kalimat adalah bagian inti, sedangkan yang dapat dihilangkan adalah bagian bukan inti. Kalimat yang terdiri atas satu kesatuan bagian inti, baik dengan maupun tanpa bagian bukan inti disebut sebagai *kalimat tunggal*. Kalimat yang terdiri atas lebih dari satu bagian inti, baik dengan maupun tanpa bagian bukan inti disebut sebagai *kalimat majemuk*. Kata atau kelompok kata yang membentuk kesatuan yang menduduki jabatan tertentu dalam kalimat disebut sebagai *konstituen*. Dalam kalimat, kita dapat menemukan salah satu konstituen yang mengendalikan kemunculan konstituen yang lain. Konstituen ini dinamakan *pusat* dari kalimat tersebut, dan biasanya merupakan konstituen yang menduduki jabatan predikat. Konstituen lain yang harus muncul disebut sebagai konstituen pendamping. Kalimat terdiri atas satu atau beberapa klausa yang masing-masing berupa kalimat tunggal. Dari segi maknanya, kalimat terbagi menjadi lima

kelompok yaitu :

- Kalimat berita (kalimat deklaratif), yaitu kalimat yang isinya memberitakan sesuatu kepada pembaca atau pendengar.
- Kalimat perintah (kalimat imperatif), yaitu kalimat yang maknanya memberikan perintah untuk melakukan sesuatu.
- Kalimat tanya (kalimat interogatif), yaitu kalimat yang isinya menanyakan sesuatu atau seseorang.
- Kalimat seru (kalimat interjektif), yaitu kalimat yang mengungkapkan perasaan kagum.
- Kalimat emfatik, yaitu kalimat yang memberikan penegasan khusus kepada subjek, misalnya dengan penambahan partikel *lah* pada subjek, atau penambahan kata sambung '*yang*' di belakang subjek.

Dalam pemahaman teks, sering kali harus diperhatikan lingkup pembicaraan agar dapat diperoleh makna yang sesuai. Lingkup teks terdiri atas berbagai unsur seperti situasi, pembicara, pendengar, waktu, tempat, topik, dan peristiwa. *Diexsis* merupakan gejala semantik yang terdapat pada kata atau konstruksi yang hanya dapat ditafsirkan acuannya dengan memperhitungkan situasi pembicaraan.

Agar kumpulan kalimat membentuk teks diperlukan adanya keterpautan bentuk (kohesi) dan keterpautan makna (koherensi). Teks yang baik memiliki kedua unsur tersebut. Kalimat atau kata dalam teks harus saling berkaitan, pengertian yang satu menyambung pengertian yang lain secara berturut-turut.

Langkah-langkah Penanganan Bahasa Alami

Secara kasar proses dalam penanganan bahasa alami dapat disusun sebagai berikut:

- **Analisa Morfologi** / Analisa Leksik.

Pada proses ini setiap kata dianalisa berdasarkan komponen-komponennya,

sehingga dapat ditentukan apakah kata tersebut ada dalam kamus, apakah merupakan kata dasar, apakah merupakan kata bentukan, apa awalnya, apa sisipannya, apa akhirnya dan sebagainya. Dalam proses ini juga dipisahkan antara kata dengan tanda-tanda baca.

- **Analisa Sintaks.**

Pada proses ini setiap kata pada suatu kalimat dianalisa keterkaitannya antara satu dengan yang lain. Apakah susunan kalimatnya sudah benar, manakah subjek predikat dan objeknya, manakah kalimat intinya, manakah anak kalimatnya, mana yang diterangkan dan mana yang menerangkan, dan sebagainya. Dalam analisa ini juga dapat ditentukan apakah suatu kalimat ditolak karena tidak sesuai dengan aturan yang ada misalnya: 'Boy the go the to store'.

- **Analisa Semantik.** Proses ini akan melakukan pemetaan antara struktur yang sudah dihasilkan oleh penganalisa sintaks dengan makna dari struktur tersebut, Struktur yang tidak memiliki mapping dengan objek makna dapat ditolak. Misalkan kalimat '*Colorless green ideas sleep furiously*' akan ditolak karena tidak memiliki arti.

- **Integrasi.**

Arti dari suatu kalimat dapat tergantung dari kalimat lain yang mengawali kalimat tersebut, dan dapat mempengaruhi makna dari kalimat lain yang mengikutinya. Sebagai contoh kata, 'It' dalam kalimat 'John wanted it' tergantung dari kalimat sebelumnya, sedangkan kata 'John' dapat mempengaruhi arti kalimat berikutnya misalkan 'He always had'.

- **Analisa Pragmatik.**

Struktur yang merepresentasikan apa yang telah dikatakan harus ditafsirkan kembali untuk mengetahui apa yang dimaksudkan sesungguhnya. Sebagai contoh kalimat 'Do you know what time

it is?' maksud sesungguhnya adalah suatu permintaan untuk menunjukkan jam berapa saat ini. Dari sudut pandang yang lain, penganalisaan suatu bahasa dapat dibagi menjadi beberapa bagian sebagai berikut:

- **Knowledge Base.**

Bagian ini bertugas menyimpan entitas dan predikat-predikat yang berkaitan dengan suatu toko tertentu. Predikat digunakan untuk memberikan Karakteristik entitas dan untuk pengujung antar entitas. Bagian ini dapat juga menggunakan aturan inferensi untuk menghasilkan fakta-fakta dan informasi baru dari data yang telah ada.

- **Text Analysis.**

Bagian ini berisi komponen pengolah sintaks yang bertugas melakukan dekomposisi struktur teks. Selain itu dapat juga ditambahkan modul pemahaman semantik yang dapat melengkapi konstruksi sintaks dengan makna semantik dari teks tersebut.

- **Pencarian dan Perbandingan.**

Bagian ini mengambil bagian tertentu dari suatu teks, melengkapi bagian tersebut, atau mengubah teks tersebut menggunakan basis pengetahuan yang sudah disimpan sebelumnya.

- **Problem Solving.**

Bagian ini mengambil makna yang sudah diturunkan dari suatu teks berbahasa alami dengan rujukan Knowledge Base untuk melakukan tugas praktis tertentu misalkan membuat *query formal* untuk dikirimkan ke modul pengelola basis data atau sistem *text-retrieval*.

Klasifikasi Gramatika dan Bahasa Formal

Setiap bahasa didefinisikan oleh suatu gramatika yang menentukan bagaimana penyusunan kata-kata pembentuk kalimat

dalam bahasa tersebut. Secara formal, gramatika suatu bahasa dituliskan sebagai pasangan empat elemen:

$$G = (V_n, V_t, P, S)$$

Rancangan Penerjemah Bahasa
(Yohanes Bowo W.)

mengembangkan mesin penterjemah dan berbagai proyek penterjemahan elektronik sudah dilaksanakan.

Pada saat pekerjaan diatas berjalan, ternyata bahwa sistem kamus (kata per kata) maupun pendekatan sintaks masih belum memadai untuk menangani kompleksitas teks berbahasa alami. Pada tahun 1966 Komite 'National Academy of Science' meragukan keseluruhan usaha pembuatan mesin otomatis. Bersamaan dengan makin jarang nya sumber biaya penelitian, perhatian beralih dari pembuatan program penterjemah praktis ke arah sistem penganalisa teks teoritis yang diharapkan dapat meningkatkan mutu penterjemahan otomatis dalam jangka panjang.

Pada masa kini, perhatian terhadap penterjemahan otomatis muncul kembali, bersamaan dengan meningkatnya jumlah teks yang tersimpan dalam bentuk elektronik dan merebaknya jaringan komunikasi yang memungkinkan pentransferan teks. Penelitian tentang pemahaman teks dan sistem penterjemah bahasa merupakan topik utama proyek Jepang yang terkenal, yaitu 'Fifth Generation'.

Alternatif Metodologi

Ada beberapa pendekatan cara kerja mesin penterjemah yaitu:

- **Sistem Penterjemahan Langsung** yang berdasarkan metoda tertentu, menterjemahkan masukan dalam bahasa tertentu menjadi bahasa target yang diinginkan.
- **Sistem Penterjemahan Tidak Langsung**, yang menganalisa bahasa

sumber terlepas dari proses sintesa bahasa target. Dalam penterjemahan tidak langsung, proses analisa bahasa sumber tidak tergantung dari karakteristik bahasa target, dan proses sintesa bahasa target tidak tergantung dari bahasa sumber. Pada sistem tidak langsung ini ada dua kemungkinan utama yaitu:

Metoda Interlingual, yaitu menggunakan representasi formal yang tidak tergantung dari bahasa sumber. Dari representasi formal ini dapat dilakukan sintesa bahasa target.

Metoda Transfer, pertama-tama mengubah teks sumber menjadi representasi internal yang bentuk dan strukturnya tergantung dari bahasa masukan. Kemudian, mentransfer representasi tersebut menjadi representasi keluaran yang tergantung dari bahasa target, yang kemudian digunakan untuk mensintesa keluaran.

Pada penelitian ini, metoda yang dipilih adalah terjemahan langsung, karena dapat mengakomodasi bentuk-bentuk spesifik yang merupakan karakteristik pemetaan Bahasa Indonesia dan Bahasa Inggris yang mungkin tidak dapat diturunkan dari pemetaan yang umum. Selain itu juga dengan pertimbangan bahwa permasalahan kita hanya pada penterjemahan dwi bahasa yaitu Inggris-Indonesia sehingga tidak dibutuhkan adanya representasi perantara.

Tahapan Pengembangan

Beberapa tahapan yang harus dilaksanakan untuk membangun pengolah bahasa alami (dalam hal ini penterjemah bahasa alami), adalah sebagai berikut:

1. Membangun Kamus. Tahapan ini dimaksudkan untuk menyimpan data tentang arti suatu kata yang berdiri sendiri, tanpa melihat dimana kata tersebut digunakan dalam kalimat dan bagaimana konteks

pembicaraan. Apabila modul ini sudah dikembangkan, pemakai sudah dapat memperoleh penterjemah sederhana yang memetakan satu kata Bahasa Inggris dengan satu kata Bahasa Indonesia secara sangat sederhana.

2. Membangun Basis Pengetahuan. Tahap ini dimaksudkan untuk menyimpan semua aturan tata bahasa yang mungkin, terutama aturan sintaks (Bahasa Inggris maupun Bahasa Indonesia). Dalam membangun basis pengetahuan ini mungkin harus memperhatikan temuan ilmu komputer yang mutakhir seperti pemrograman berorientasi objek, fuzzy logic, kecerdasan buatan, pemrograman paralel, bahkan jaringan syaraf, agar dapat dibentuk basis pengetahuan yang handal.

3. Membangun Inference Engine dan User Interface.

Tahap ini dimaksudkan untuk menyediakan fasilitas untuk mengekstraksi pengetahuan berdasarkan kebutuhan dari pemakai. Disini harus dapat diusahakan suatu interaksi yang semudah mungkin dengan manfaat yang maksimal agar pemakai dapat memperoleh manfaat yang sebesar-besarnya dari Basis Pengetahuan yang ada. Dalam tahap ini juga mungkin harus menggunakan penemuan ilmu komputer yang mutakhir misalkan Hypertext.

Pada bagian ini juga harus disediakan modul tambahan yaitu *preediting* untuk memperbaiki dan menstandarkan masukan, serta modul *post editing* untuk memperbaiki dan menstandarkan keluaran.

10. Masukan-Keluaran

Masukan untuk program Penterjemah Teks dapat terdiri dari beberapa bentuk: a. **File Teks Bahasa Inggris dalam format ASCII.** Teks ini akan diterjemahkan langsung menjadi bentuk file teks berbahasa Indonesia. File hasil tersebut dapat disunting

terlebih dahulu maupun langsung dicetak.

b. Pengetikan langsung melalui editor.
 Jika pemakai memilih bentuk masukan ini, akan disediakan dua window yaitu window teks bahasa Inggris untuk mengedit masukan dalam bahasa Inggris, dan window keluaran untuk menampilkan hasil terjemahannya.

Diharapkan keluaran dapat disajikan dalam bentuk interaktif, yaitu bahwa setiap perubahan teks di window masukan akan mengakibatkan perubahan yang bersesuaian di window keluaran. Window Editor direncanakan memiliki bentuk sebagai berikut:

Tabel 2. Bentuk Window Editor

BAHASA INGGRIS 1 SAW A LITTLE CAR PASSED THEY STUDY SERIOUSLY JOHN MET ME YESTERDAY TECHNICAL SUPPORT & MESSAGE	BAHASA INDONESIA SAYA MELIHAT MOBIL KECIL LEWAT MEREKA BELAJAR BERSUNGGUH- SUNGGUH JOHN MENJUMPAI SAYA KEMADIN
---	--

Dengan model interaksi seperti diatas, pemakai dapat berkomunikasi dengan program secara interaktif, sehingga hasil dari suatu kata terjemahan dapat dilihat langsung oleh pemakai, dan pesan dari basis pengetahuan dapat dilihat melalui Window Technical Support & Message. Pada window tersebut diharapkan dapat disajikan informasi yang sesuai berkenaan dengan aturan struktur kalimat, tata bahasa, serta informasi tentang

hasil parsing dan pencarian kamus. Program dikembangkan dengan bahasa Pemrograman C++, dengan memanfaatkan fasilitas MDI (Multiple Document Interface), dimana pengguna dapat mengaktifkan beberapa window sekaligus, dan dapat dengan mudah berpindah dari satu window aktif ke window yang lainnya. Direncanakan bentuk menu yang disediakan seperti pada table3.

Tabel 3. Bentuk Menu

TRSTAR- PROGRAM PENTERJEMAH BAHASA INGGRIS - INDONESIA					
FILE	EDIT	OPERATION	KNOWLEDGE	THESAUR US	
New	Open	Cut	Copy	Alternatives	Add Knowledge
Save	Save	Paste	Option	Choose	Disp Knowledge
As	Print	Delete	Expert	Choose	Delete Knowledge
Exit	Clear	Style	Choose	Word	KS Sentence
	PerWord	Level		KS	Semantic KS

Menu FILE dengan submenu-submenunya digunakan untuk operasi file baik membuka file (sekali membuka dua buah window akan terbuka sekaligus), menutup file, menyimpan file, mencetak file, dan lain-lain. Bentuk menu ini serupa dengan program-program berbasis window yang lain. Menu EDIT dan submenu-submenunya digunakan untuk mengedit file, meliputi menyalin, menghapus, menyunting per kata, menyunting perkalimat, dan menyunting perteks (dengan memperhatikan aspek semantik).

Menu OPERATION dan submenu-submenunya digunakan untuk pemilihan setting penerjemahan, misalkan untuk memilih *xpert*, memilih *style* terjemahan, memilih level terjemahan, dan lain-lain. Selain itu melalui menu ini juga dapat disajikan beberapa alternatif terjemahan jika hasil terjemahan yang telah diberikan tidak memuaskan pemakai.

Menu KNOWLEDGE dan submenu-submenunya digunakan untuk memelihara basis pengetahuan yang disimpan dalam program, meliputi penambahan basis pengetahuan baru, menghapus basis pengetahuan baru, termasuk di dalamnya menambah kata baru maupun menghapusnya, mengelola basis pengetahuan kata, pengetahuan kalimat, dan pengetahuan semantik.

Menu THESAURUS disediakan bagi pemakai apabila pemakai ingin mendapat penjelasan mengenai suatu kata tertentu, baik artinya, sinonimnya, antonimnya, maupun informasi yang lain. Diharapkan menu ini juga dapat menyediakan fasilitas navigasi bagi pemakai apabila pemakai ingin mengetahui tentang suatu konsep tertentu, beserta istilah-istilah yang berkaitan dengan konsep tersebut.

Basis Pengetahuan

Agar pengetahuan tentang bahasa dapat dipergunakan untuk penterjemahan, pengetahuan tersebut harus disimpan dalam bentuk tertentu, yang dikenal sebagai *representasi pengetahuan*. Kumpulan dan representasi pengetahuan disebut *Basis Pengetahuan*. Dalam merepresentasikan pengetahuan ini terdapat dua entitas yang berbeda yaitu:

- fakta, yaitu hal-hal yang ingin direpresentasikan
- representasi fakta dalam bentuk yang dipilih dan dapat dimanipulasi. Dalam representasi pengetahuan harus terdapat pemetaan dari fakta ke representasi fakta, dan pemetaan dari representasi fakta ke fakta. Representasi fakta yang paling sering dipergunakan dalam kehidupan sehari-hari adalah kalimat bahasa alami.

Hubungan antara fakta, representasi bahasa alami, dan representasi internal diilustrasikan sebagai berikut:

- Fakta dapat direpresentasikan dalam kalimat-kalimat bahasa alami. Proses ini ditemui pada seorang penutur yang sedang menceritakan suatu fakta atau kejadian.
- Rangkaian kalimat bahasa alami merupakan representasi dari fakta, misalnya seorang pendengar cerita dapat membayangkan fakta atau kejadian yang diceritakan padanya.
- Proses pemahaman bahasa alami merupakan upaya untuk mentransformasikan fakta yang direpresentasikan dalam bahasa alami ke bentuk representasi internal perangkat lunak.
- Fakta dalam representasi internal perangkat lunak dapat dikembalikan ke

bentuk representasi bahasa alami melalui proses pembangkitan bahasa alami.

- Untuk kebutuhan tertentu, fakta dapat langsung ditransformasikan ke dalam representasi internal perangkat lunak, seperti dilakukan dalam penisian basis pengetahuan sistem pakar.
- Representasi internal dapat diinterpretasikan dengan kaidah tertentu sehingga diperoleh kembali fakta yang direpresentasikannya, seperti pada pemakaian sistem pakar.
- Representasi internal inilah yang diperoleh oleh perangkat lunak sesuai dengan kebutuhan pengolahan. Fakta dalam representasi internal ini diolah oleh program penalaran untuk menghasilkan representasi internal lain yang dapat dipergunakan lebih lanjut.

Agar pengetahuan dapat diolah dan dipergunakan oleh komputer, pengetahuan tersebut harus disimpan dalam bentuk representasi pengetahuan tertentu. Salah satu diantara representasi pengetahuan yang dikenal pada bidang intelegensia buatan adalah *Conceptual Dependency* (CD).

Primitif konseptual secara umum dapat dikelompokkan menjadi empat kategori yaitu:

- Kategori aksi (ACT)
- Kategori objek (PP)
- Kategori keterangan aksi (AA)
- Kategori keterangan objek (PA)

Skenario (*script*) merupakan suatu struktur representasi pengetahuan yang menjabarkan urutan kejadian dalam lingkup tertentu, Skenario terdiri atas sekumpulan alur (slot) yang masing-masing berisi informasi tentang nilai yang mungkin diisikan ke dalam alur atau suatu nilai baku. Jika tidak ada informasi yang tersedia.

Dalam dunia nyata, skenario dapat dipandang sebagai pola kemunculan kejadian. Pola ini

muncul karena adanya hubungan sebab akibat antar kejadian. Pelaku akan melakukan aksi tertentu yang memungkinkannya melakukan aksi yang lain. Kejadian-kejadian yang dijabarkan dalam skenario membentuk sebuah rantai sebab-akibat (causal chain). Jika suatu skenario diketahui berhubungan dengan situasi yang diberikan, maka skenario ini dapat digunakan untuk meramalkan kemunculan kejadian yang tidak disebutkan secara eksplisit dalam teks. Skenario juga dapat digunakan untuk menunjukkan bagaimana suatu kejadian berinteraksi dengan kejadian lainnya.

Skenario yang dipergunakan harus dipilih sesuai dengan situasi yang ditemui dalam wacana. Menurut cara pengaktifannya, skenario dapat dibedakan atas dua jenis, yaitu:

- Skenario yang dipergunakan pada keseluruhan teks. Setelah situasi dalam teks diketahui, skenario yang sesuai diaktifkan, dan alur-alur dalam skenario dicoba diisi dengan objek yang sesuai.
- Skenario yang hanya diperhatikan sepintas dan bukan merupakan skenario pusat situasi. Satu skenario tidak dapat dipergunakan dalam keseluruhan wacana, tetapi beberapa skenario diaktifkan secara bergantian untuk melengkapi pengetahuan wacana yang diperoleh.

Proses

Preediting merupakan proses *optional* (boleh dilakukan boleh juga tidak) yang dimaksudkan untuk mengatur penulisan teks masukan dalam bahasa sumber sesuai dengan aturan penulisan yang berlaku bagi bahasa sumber tersebut. Preediting dilakukan jika pemakai ingin mendapatkan hasil terjemahan yang optimal, sesuai dengan kaidah bahasa yang berlaku.

Analisa Prefix dan Suffix merupakan proses *morfologi* yang menganalisa pembentukan

suatu kata yang telah muncul yaitu menentukan apa kata dasarnya, apa awalannya, apa sisipannya, apa akhirnya dan bagaimana perubahan makna yang mungkin terjadi setelah kata dasar ditambah dengan imbuhan-imbuhan tersebut.

Pencarian Kamus merupakan proses untuk mencari apakah kata yang muncul sudah terekam dalam kamus bahasa sumber. Apabila kata tersebut sudah direkam, maka proses dapat dilanjutkan. Apabila kata tersebut belum direkam, maka pemakai dapat menambahkan pengetahuan tentang kata tersebut, dan memperbaiki kamus. Setelah kamus diperbaiki dan kata baru telah masuk, maka dapat dilanjutkan proses **Pencarian kamus kedua** yaitu pencarian kamus setelah kamus diperbaiki. **Analisa Sintaks** merupakan proses dekomposisi kalimat menjadi elemen-elemennya yaitu Subjek, Predikat, Objek, Keterangan Waktu, Keterangan Tempat, Adjektiva, dan lain-lain. Analisa ini juga menentukan manakah bagian inti dari suatu kalimat majemuk, dan mana bagian bukan intinya. Analisa ini dapat menentukan apakah suatu kalimat telah tersusun berdasarkan susunan baku kalimat Bahasa Sumber. Apabila belum sesuai, dapat dilakukan penyesuaian secara otomatis dengan memilih struktur sintaks yang ditawarkan oleh program. **Transfer Leksik** merupakan proses mencari padanan (pemetaan) kata per kata dari bahasa sumber ke bahasa sasaran. Pencarian tersebut melibatkan kamus bahasa sumber, kamus bahasa sasaran dan pemetaannya. Proses ini masih belum memperhitungkan struktur kalimat bahasa sumber dan struktur kalimat bahasa sasaran.

Transfer Sintaks merupakan proses penyesuaian aturan tata bahasa antara tata bahasa sumber dengan tata bahasa sasaran. Proses ini memanfaatkan basis pengetahuan tata bahasa sumber dan basis pengetahuan tata bahasa sasaran untuk menetapkan mana aturan produksi yang cocok. Apabila aturan

produksi yang cocok sudah ditemukan, maka dapat dilakukan transfer aturan produksi bahasa sumber menjadi aturan produksi bahasa tujuan yang bersesuaian. Pada proses ini juga sudah mempertimbangkan aspek bebas konteks dari teks masukan.

Pembentukan Sintaks merupakan proses pembentukan struktur kalimat bahasa sasaran, berdasarkan aturan produksi yang sudah ditemukan pada tahap sebelumnya. Dalam proses ini belum diperhitungkan aspek morfologi (imbuhan dari tiap kata yang muncul).

Pembentukan Morfologi merupakan proses menambahkan imbuhan yang cocok untuk setiap kata yang muncul dalam kalimat bahasa sasaran. Proses ini mempergunakan basis pengetahuan tentang aspek morfologis bahasa sasaran. **Post Editing** merupakan proses optional yang dilakukan jika pemakai kurang merasa puas terhadap hasil terjemahan, atau pemakai ingin menggunakan gaya terjemahannya sendiri.

Kesimpulan

- Penterjemahan bahasa alami merupakan permasalahan yang masih sulit dipecahkan secara total oleh komputer. Sejauh ini yang bisa dilakukan adalah memanfaatkan keunggulan komputer secara maksimal, misalkan dalam hal kemampuan penyimpanan pengetahuan, sedangkan hal yang tidak dapat diselesaikan oleh komputer tetap diserahkan kepada pakar penterjemah. Untuk membangun perangkat lunak penterjemah teks perlu beberapa modul utama yaitu, modul penanganan kamus, modul penanganan 'Knowledge Base', dan modul 'Inference Engine.' dan 'User Interface'
- Teknik-teknik baru dalam ilmu komputer seperti pemrograman berorientasi objek, logika fuzzy, *hypertext*, dan lain-lain dapat meningkatkan kualitas terjemahan.
- Kualitas terjemahan juga ditentukan oleh

banyaknya aturan yang disimpan dalam basis pengetahuan, meliputi aturan sintaks, morfologi, dan semantik. Selain dari jumlah aturan yang disimpan, aturan-aturan tersebut harus akurat, sesuai dengan karakteristik bahasa sumber dan bahasa sasaran.

Prospek Pengembangan

Karena perangkat lunak penerjemah menyimpan pengetahuan tentang kamus dan struktur kalimat, perangkat lunak ini dapat dikembangkan menjadi *text formatter*, yaitu suatu alat untuk mengoreksi teks yang tidak tertulis dalam bentuk standar menjadi teks lain yang tertulis dalam bentuk standarta tanpa menghiian arti.

Penterjemah bahasa ini dapat dikembangkan lebih lanjut sehingga dapat digunakan untuk bahasa di semua kalangan, dengan menambah aturan tata bahasa dan basis pengetahuannya. Selanjutnya pengetahuan yang ada dapat dikelompokkan lagi menjadi beberapa domain pengetahuan misalnya domain Ekonomi, Politik, Fisika, Elektronika, Komputer, Hukum, Psikologi, dan lain-lain. Dengan cara tersebut seorang yang hendak menterjemahkan suatu teks dapat rnenyebutkan domain mana yang ingin digunakan, sehingga setiap kata dapat diterjemahkan secara lebih tepat. Selain dari pembagian berdasarkan domain, dapat juga dikembangkan pembagian berdasarkan gaya penterjemahan. Dalam kasus ini perangkat lunak dapat berfungsi sebagai Expert System dimana perangkat lunak tersebut dapat menyimpan pengetahuan dari beberapa orang pakar penterjemah dengan gaya masing-masing, dan dapat menyediakan pilihan bagi pemakai mengenai gaya mana yang diinginkan oleh pemakai tersebut, Konsep-konsep yang digunakan didaiaim pengelolaan bahasa alami juga dapat diterapkan untuk bahasa selain Inggris-Indonesia, misalnya Perancis, Spanyol, Cina, Jepang dan lain-lain. Apabila

dapat dibangun basis pengetahuan untuk semua bahasa diatas, niscaya akan dapat mempermudah masyarakat Internasional! untuk berkomunikasi antara negara yang satu dengan negara yang lainnya.

Perangkat lunak ini dapat dilengkapi dengan modul-modul pengolahan suara, baik mengolah suara menjadi teks maupun mengolah teks menjadi suara, agar didapat hasil yang lebih berguna bagi pemakai. Apabila hal ini dapat dilakukan maka dapat dihasilkan suatu perangkat yang menterjemahkan suara bahasa Inggris langsung menjadi suara bahasa Indonesia. Perangkat semacam tersebut tentu saja akan sangat bermanfaat bagi peradaban manusia. Selanjutnya apabila perangkat tersebut dapat dimodifikasi sehingga dapat dibawa kemana-mana dengan mudah (*portable*), seperti telpon genggam, maka kemanfaatannya akan lebih terasa lagi bagi masyarakat yang secara intensif berkomunikasi dengan orang luar negeri.

KESIMPULAN

Dengan menggunakan software penterjemah bahasa dapat meningkatkan produktivitas khususnya dalam berkomunikasi dan ternyata dapat menggunakan beberapa

DAFTAR PUSTAKA

- Aho, Alfred V. ; Sethi, Ravi, dan Ullman, Jeffrey D. . *Compilers Principles Techniques and Tools*. Addison Wisley, 1986.
- Booch, Grady, *Object-Oriented Analysis and Design with Application*. The Benjamin/Cummings Publishing Company Inc, 1994.
- CICC, Multi-lingual Machine Translation System, CICC (Center of the International Cooperation for Computerization, 1994. '

- Echols, John M., dan Shadily, Hassan, *Kamus Indonesia Inggris*, PT Gramedia, 1992.
- Echols, John M., dan Shadily, Hassan, *Kamus Inggris Indonesia*. PT Gramedia, 1992.
- Frakes, William B. , dan Baeza-Yates, Ricardo, *Information Retrieval – Data Structures and Algorithms* -, Prentice Hall.
- Klir, George J. , dan Folger, Tina A. *Fuzzy Sets. Uncertainly and Information*, Prentice-Hall Inc, 1992.
- Lippman, Stanley B. , *C++ Primer*. Addison Wesley Publishing Co, 1991.
- Rao, Valiuru B., dan Rao, Hayagriva V. , *C++ Neural Networks and Fuzzy Logic*. Management information Source Inc, 1993