

Natural Language Processing



Ali Ridho Barakbah

Apa itu NLP?

- Proses pembuatan model komputasi dari bahasa sehingga memungkinkan terjadinya interaksi antara manusia dan komputer dengan perantara bahasa alami yang dipakai oleh manusia.
- NLP memodelkan pengetahuan terhadap bahasa, baik dari segi kata, bagaimana kata-kata bergabung menjadi suatu kalimat dan konteks kata dalam kalimat.

Disiplin ilmu dari NLP

- Fonetik / fonologi
- Morfologi
- Sintaksis
- Semantik
- Pragmatik
- Discourse knowledge
- World knowledge

Fonetik / Fonologi

- Berhubungan dengan suara yang menghasilkan kata yang dapat dikenali.
- Bidang ini dipakai dalam aplikasi-aplikasi speech based system

Morfologi

- Pengetahuan tentang kata dan bentuknya sehingga bisa dibedakan antara yang satu dengan yang lain.
- Contoh:
 - membangunkan
 - bangun (kata dasar)
 - mem (prefix)
 - kan (suffix)

Sintaktis

- Pengetahuan tentang urutan kata dalam pembentukan kalimat.
- Contoh:
 - Kalimat → Subyek, Predikat
 - Subyek → Determinan, KataBenda
 - Predikat → KataKerja, KataBenda

Semantik

- Mempelajari arti suatu kata dan bagaimana arti kata-arti kata tersebut membentuk suatu arti kata dari kalimat yang utuh.
- Contoh:
Ayahku datang membawa buah tangan

Pragmatik

- Pengetahuan tentang konteks kata/kalimat yang berhubungan erat keadaan atau situasi kata/kalimat tersebut dipakai.
- Contoh:
 - Ayah datang (diucapkan dengan nada datar)
 - Ayah datang! (diucapkan dengan nada tinggi)
 - Ayah datang? (diucapkan dengan tempo cepat)

Discourse Knowledge

- Pengetahuan tentang hubungan antar kalimat.
- Melakukan pengenalan apakah suatu kalimat yang telah dikenali mempengaruhi kalimat selanjutnya.
- Penting untuk identifikasi kata ganti orang, keterangan tempat atau aspek sementara dari informasi.

- Contoh:

Ibu pergi ke pasar. Ia membeli makanan disana.



World Knowledge

- Mencakup arti sebuah kata secara umum dan apakah arti khusus bagi suatu kata dalam suatu percakapan dengan konteks tertentu

Aplikasi NLP

- Text-based application
- Speech-based application

Text-based application

- Aplikasi yang melakukan memprosesan terhadap teks tertulis
- Contoh:
 - Mencari topik tertentu dari buku di perpustakaan
 - Mencari isi dari suatu berita atau artikel
 - Mencari isi dari email
 - Menterjemahkan dokumen dari suatu bahasa ke bahasa lain

Speech-based application

- Aplikasi yang melakukan memprosesan dari bahasa lisan atau pengenalan suara.
- Contoh:
 - Sistem otomatis pelayanan melalui telepon
 - Control suara pada peralatan elektronik
 - Aplikasi peningkatan kemampuan berbahasa

Contoh aplikasi NLP

- ELIZA yang dibekali pengetahuan psikologi, sehingga beberapa orang terdorong untuk mampu merubah sikap dan perilakunya.
- Jupiter yang mampu memberikan informasi cuaca melalui telepon.
- ALVIN yang mampu menjawab pertanyaan mengenai DOS.
- SEXPERT yang dirancang untuk perbincangan mengenai pendidikan seksual.
- Email translator
- Web translator
- World translator

Jupiter

A Conversation with Jupiter

The following is a transcription of a telephone conversation between Victor Zue and the Jupiter system for weather information. Jupiter's toll-free number is 1-822-573-8756.

[Phone ringing]

Jupiter: Hi. Welcome to Jupiter, the M.I.T. Lab for Computer Science Weather Information System. This call is being recorded for system development. You may hang up or ask for help at any time. How can I help you?

V.Z.: What is the forecast for Boston today?

[The theme song of the television show *Jeopardy* plays during a pause of seven seconds. Each of Zue's subsequent questions is followed by a pause of one to three seconds.]

Jupiter: In Boston this afternoon, partly sunny and windy. High around 50. Tonight, clear, lows around 35. What other information can I give you?

V.Z.: Will it rain tomorrow?

Jupiter: To my knowledge, the forecast calls for no rain in Boston tomorrow. Is there anything else?

V.Z.: How about Seattle?

Jupiter: To my knowledge, the forecast calls for no rain in Seattle tomorrow. Is there something else you would like to know?

V.Z.: Is it sunny anywhere in the Caribbean?

Jupiter: Sorry, I am not sure what you said.

V.Z.: Is it sunny anywhere in the Caribbean?

Jupiter: I am sorry, I may not know the city you are asking about. Try asking for the state or country. For example, what cities do you know about in Massachusetts?

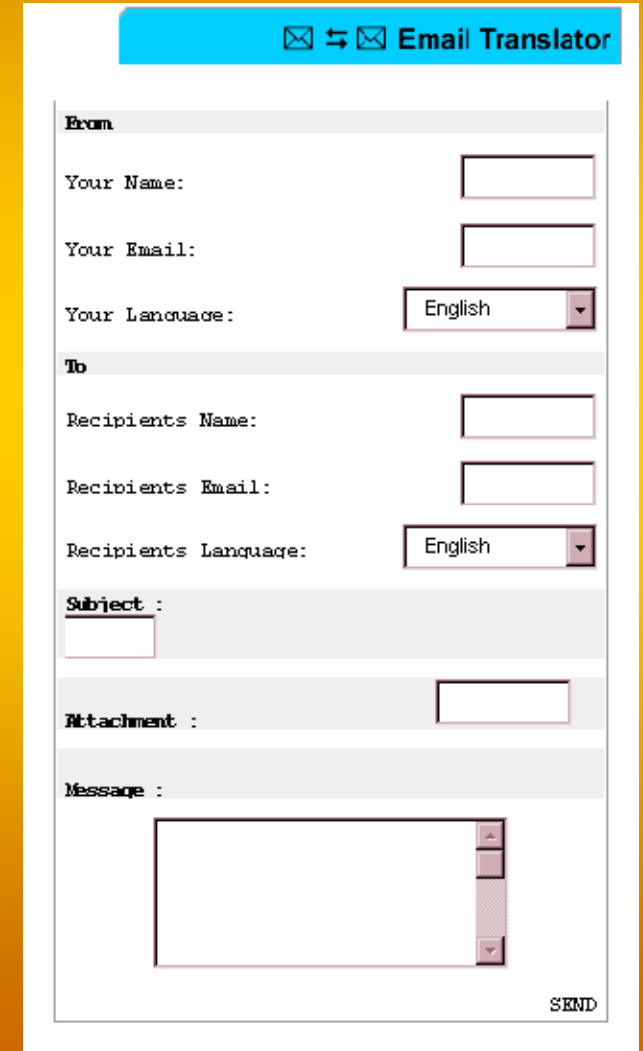
V.Z.: What is the temperature in Montego Bay?

Jupiter: In Montego Bay today, high 86 and low 73. Is there something else?

V.Z.: Good-bye.

Email translator

- Alat yang akan menjawab masalah perbedaan bahasa, karena email translator mampu menterjemahkan bahasa, seperti yang kita inginkan.
- Email Translator akan menterjemahkan kalimat-kalimat di dalam mail box, jika email yang kita terima tidak sesuai dengan bahasa kita sehari-hari.



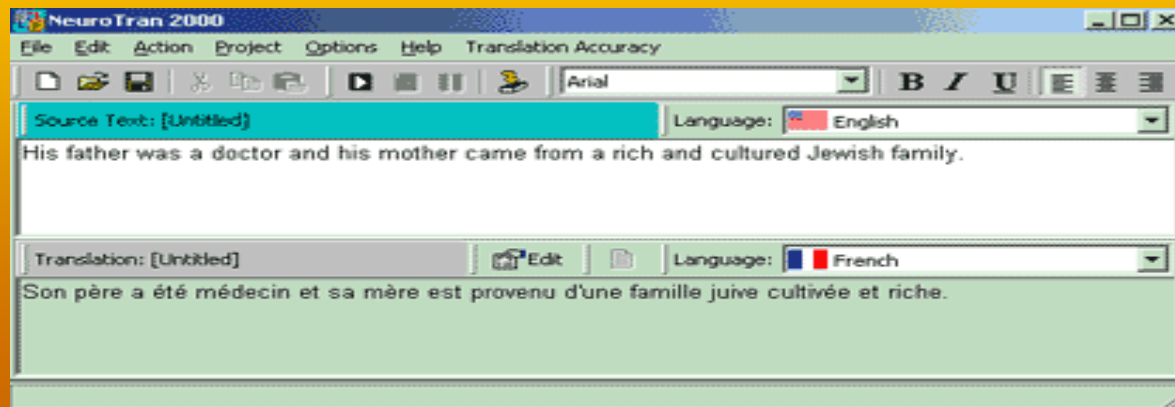
The screenshot shows a web interface for an email translator. At the top, there is a blue header with the text "Email Translator" and icons of two mailboxes with a double-headed arrow between them. Below the header, the form is organized into sections:

- From:** This section contains three input fields: "Your Name:" with an empty text box, "Your Email:" with an empty text box, and "Your Language:" with a dropdown menu currently set to "English".
- To:** This section contains three input fields: "Recipients Name:" with an empty text box, "Recipients Email:" with an empty text box, and "Recipients Language:" with a dropdown menu currently set to "English".
- Subject :** This section has a single empty text box.
- Attachment :** This section has a single empty text box.
- Message :** This section features a large, empty text area with a vertical scrollbar on the right side.

At the bottom right corner of the form, there is a "SEND" button.

Web translator

- Suatu mesin aplikasi berbasis World Wide Web yang dapat menterjemahkan bahasa dalam suatu web site.
- Web Translator akan menterjemahkan bahasa di dalam semua link-link, page per page menjadi bahasa seperti yang kita inginkan.



World translator

- Suatu pengembangan dari word translator yang sudah ada. Diharapkan dengan teknik ini hasil terjemahan bahasa akan menjadi lebih sempurna, mengikuti kaidah tata bahasa.
- Terjemahan akan lebih cepat, akurat bukan lagi dengan sistem menterjemahkan per kata, tapi per kalimat dengan melihat Subjek-Predikat-Objek.
- Pengembangan ini diharapkan mampu menjawab tantangan dari word translator yang sudah ada di pasaran.

Gramatika

- Suatu aturan yang menentukan apakah suatu kumpulan kata dapat diterima sebagai kalimat dalam suatu bahasa.
- Context Free Grammar (CFG) adalah representasi grammar dari Chomsky Hierarchy yang mudah dipahami dan diolah dalam bentuk program.
- CFG merepresentasikan bahwa suatu grammar itu dapat dibentuk dari 4 elemen, yaitu:
 - Simbol awal
 - Aturan penulisan
 - Simbol non terminal
 - Simbol terminal

Contoh

simbol awal

aturan penulisan

simbol non terminal

Kalimat → KataBenda KataKerja

KataBenda → {ayam, kucing}

KataKerja → {makan, minum}

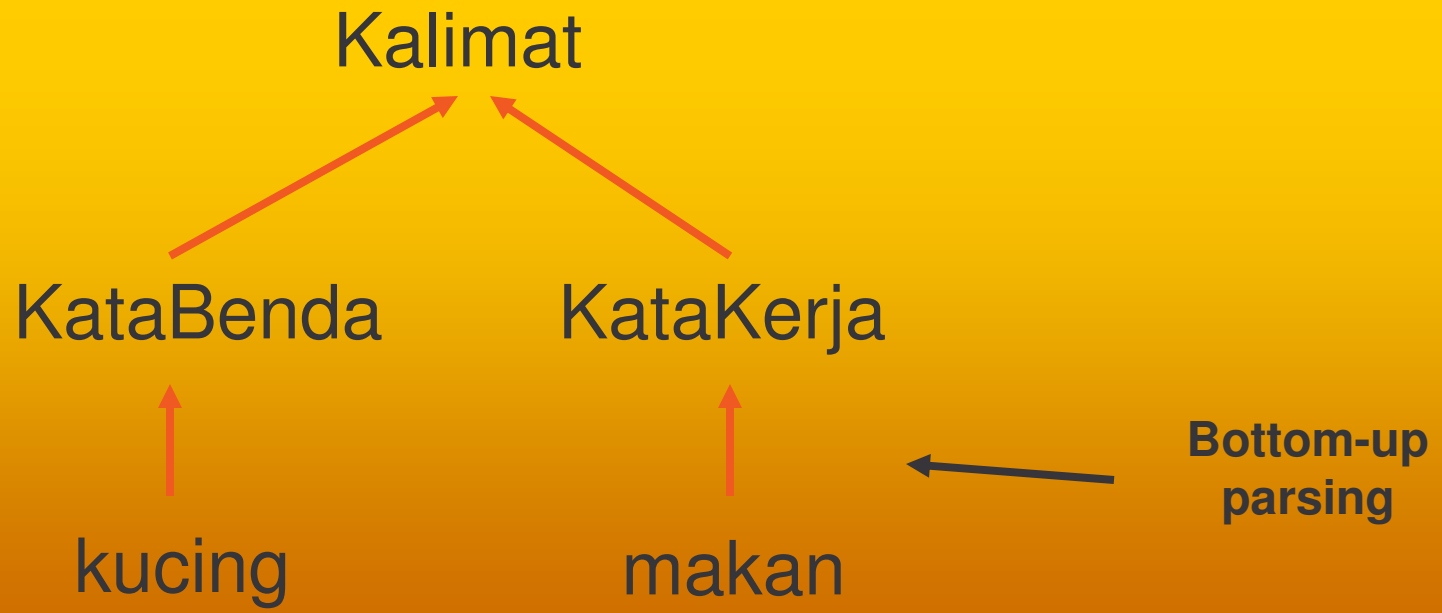
simbol terminal

Parsing

- Suatu proses menganalisa suatu kumpulan kata dengan memisahkan kata-kata itu dan menentukan struktur sintaktis dari tiap kata tersebut.
- Mempunyai 2 pendekatan:
 - Top-down parsing
 - Bottom-up parsing

Contoh

kucing makan



Referensi

- Modul Ajar Kecerdasan Buatan, Entin Martiana, Tessy Badriyah, Riyanto Sigit, Politeknik Elektronika Negeri Surabaya, 2005.
- Artificial Intelligence with Turbo Prolog, Keith Weiskamp and Terry Hengl.
- Talking with your computer, Victor Zue, Scientific American, August 1999.
- Tuntunan Praktis Pemrograman Bahasa Prolog, Andrey Andoko, cetakan kedua, Penerbit elex Media Komputindo, 1991.