

Introducción a Natural Language Processing

Módulo 2

Ronald Cárdenas Acosta

Setiembre, 2016

Outline

- 1 Inteligencia Artificial y Lenguaje Humano
- 2 Sistemas de interacción Humano-Computador via voz
- 3 Tecnologías de Lenguaje
- 4 ¿Porqué es NLP difícil?

Outline

- 1 Inteligencia Artificial y Lenguaje Humano
- 2 Sistemas de interacción Humano-Computador via voz
- 3 Tecnologías de Lenguaje
- 4 ¿Porqué es NLP difícil?

Inteligencia Artificial y el Lenguaje Humano

- Comprender y generar lenguaje natural (humano) es uno de los enfoques principales de IA desde sus inicios:
Interacción Humano-Máquina a nivel conversacional
- La tarea es considerada AI-complete:
Se requeriría de todos los diferentes tipos de conocimientos poseidos por un humano (gramática, semántica, hechos del mundo real, etc)
- Principales dificultades:
 - Ambigüedad del lenguaje humano.
 - Vocabularios cambian (evolucionan) rápidamente.
 - Riqueza de expresión: varias formas semánticamente correctas.
- Actualmente, a estos programas se les conoce como Tecnologías de Lenguaje (*Language Technologies*)

Evolución de métodos aplicados

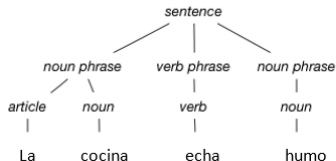
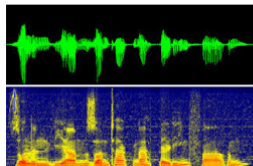
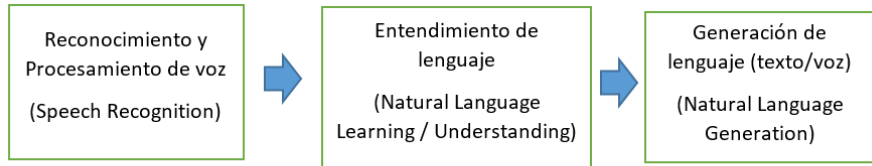
- 1960s: Uso de ontologías (estructuras semánticas de hechos/objetos del mundo real) redactadas a mano.
- 1970s: Traducción y sistemas básicos de conversación basados en reglas lógicas hechas a mano.
- 1980s: Uso de técnicas de Machine Learning estadístico para producir reglas difusas.

Outline

- 1 Inteligencia Artificial y Lenguaje Humano
- 2 Sistemas de interacción Humano-Computador via voz**
- 3 Tecnologías de Lenguaje
- 4 ¿Porqué es NLP difícil?

Ejemplo: Sistema de Interacción Humano-Computadora (Siri, Amazon Echo, Cortana)

Figure: Pipeline de sistema HCI con interfaz de voz.



"¿Debo llamar a los bomberos?"

Outline

- 1 Inteligencia Artificial y Lenguaje Humano
- 2 Sistemas de interacción Humano-Computador via voz
- 3 Tecnologías de Lenguaje**
- 4 ¿Porqué es NLP difícil?

Tecnologías de Lenguaje: Aplicaciones de NLP

- Reconocimiento de entidades (*Named Entity Recognition*)
- Inferencia de categoría gramatical (*Part-of-speech Tagging*)
- Generación de resumen (*Summarization*)
- Extracción de información (*Information Extraction*)
- Traducción automática (*Machine Translation*)
- Sistemas de Pregunta-Respuesta (*Question Answering*)
- Minería de opiniones / análisis de sentimientos (*Sentiment Analysis*)
- Desambigüación de palabras (*Word Sense Disambiguation*)
- Y muchas más...

Outline

- 1 Inteligencia Artificial y Lenguaje Humano
- 2 Sistemas de interacción Humano-Computador via voz
- 3 Tecnologías de Lenguaje
- 4 ¿Porqué es NLP difícil?

¿Cuán difícil es NLP?

Spam detection

Let's go to Agra! ✓

Buy V1AGRA ... ✗

Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

Sentiment analysis

Best roast chicken in San Francisco! 👍

The waiter ignored us for 20 minutes. 👎

Coreference resolution

Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)

I need new batteries for my *mouse*.

Parsing

I can see Alcatraz from the window!

Machine translation (MT)

第13届上海国际电影节开幕... →

The 13th Shanghai International Film Festival...

Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30

Party May 27 add

Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose

⇒ Economy is good

Dialog

Where is Citizen Kane playing in SF?

Castro Theatre at 7:30. Do you want a ticket?

(a) Campos con mejores resultados

(b) Campos con buen progreso

(c) Campos aun muy difíciles

Ambigüedad del lenguaje natural

Caso 1

Juan vio a Pedro en el parque con un vinocular

Caso 2

Juan vio a Pedro en el parque con un vinocular

No está claro a que persona esta ligado el modificador *con un vinocular*.

Evolución del vocabulario

- Importación de palabras en otros idiomas:
chat, bartender, blog, gym,...
- Deformalización del lenguaje cotidiano (texto informal en websites, redes sociales):
wow, aww, lol,:),:D,:((
- Desuso de nombres de objetos obsoletos:
beepers, vhs, diskette, tocadiscos,..

Paráfrasis

- Diferentes formas de expresar una misma idea
- Cada frase puede tener diferente estructura sintáctica
- El modelo debe ser capaz de comparar las frases semánticamente

Ejemplo

El presidente del Perú terminó su mandato en fines del mes de julio
La gestión de Humala finalizó en julio.