



Inteligencia Artificial

Unidad 4

Introducción

Procesamiento de Lenguaje Natural

- Hugo David Calderón
- Willy Ugarte Rojas
- Jorge Valverde Rebaza

Logro



- Al finalizar la unidad, el estudiante conoce las técnicas de Procesamiento de Lenguaje Natural.





Lenguaje Natural

Es el lenguaje hablado y/o escrito por humanos para propósitos generales de comunicación, para distinguirlo de otros como puedan ser una lengua construida, los lenguajes de programación o los lenguajes usados en el estudio de la lógica formal, especialmente la lógica matemática.

El término lenguaje natural se refiere al estudio de las propiedades computacionales y de otro tipo implicadas en la comprensión, producción y uso de las lenguas naturales.

100 mil años ?





Lenguaje Natural

Resumiendo los lenguajes naturales se distinguen por las siguientes propiedades:

- Desarrollados por enriquecimiento progresivo antes de cualquier intento de formación de una teoría.
- La importancia de su carácter expresivo debido grandemente a la riqueza del componente semántico
- Dificultad o imposibilidad de una formalización completa.





Procesamiento de lenguaje natural

Se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas o entre personas y máquinas por medio de programas que ejecuten o simulen la comunicación.

Modelos



comprensión del lenguaje



Aplicaciones PLN





Componentes de PLN



Análisis morfológico



Análisis sintáctico



Análisis pragmático



Análisis semántico



Lingüística Computacional

- Campo multidisciplinar de la lingüística y la informática.
- Utiliza la informática para estudiar y tratar el lenguaje humano.

Modela



Forma lógica el lenguaje natural

Participan lingüistas, informáticos especializados en inteligencia artificial, psicólogos cognoscitivos y expertos en lógica, entre otros



Herramientas disponibles

Natural Language Toolkit (NLTK) is an open source platform for performing NLP tasks including tokenization, stemming, POS tagging, parsing, and semantic reasoning. It provides interfaces for many corpora and lexicons which are useful for opinion mining and sentiment analysis.

<http://www.nltk.org/>

Python



Herramientas disponibles

The Apache OpenNLP is a JAVA library for the processing of natural language texts, which supports common tasks including tokenization, sentence segmentation, POS tagging, named entity recognition, parsing, and coreference resolution.

<https://opennlp.apache.org>

Java



Herramientas disponibles

Stanford CoreNLP is a framework which supports not only basic NLP task, such as POS tagging, named entity recognition, parsing, coreference resolution, but also advanced sentiment analysis.
<http://stanfordnlp.github.io/CoreNLP/>

Java



Herramientas disponibles

Gensim is an open source library for topic modeling which includes online Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), Random Projection, Hierarchical Dirichlet Process and word2vec. All implemented algorithms support large scale corpora. LSA and LDA have distributed parallel versions.

<http://radimrehurek.com/gensim/>

Python



Herramientas disponibles

FudanNLP is an open source toolkit for Chinese NLP, which supports word segmentation, POS tagging, named entity recognition, dependency parsing, coreference resolution and so on.

<https://code.google.com/archive/p/fudannlp/>

Java



Herramientas disponibles

The Language Technology Platform (LTP) is an open source NLP system for Chinese, including lexical analysis (word segmentation, POS tagging, named entity recognition), syntactic parsing and semantic parsing (word sense disambiguation, semantic role labeling) modules.

<http://www.ltp-cloud.com/intro/en/>

Python / C++



Herramientas disponibles

NiuParser is a Chinese Syntactic and Semantic Analysis Toolkit, which supports word segmentation, POS tagging, named entity recognition, constituent parsing, dependency parsing and semantic role labeling.

<http://www.niuparser.com/index.en.html>

C++

Fin