Ferramentas para o Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) consiste na aplicação de métodos e técnicas que possibilitem ao computador extrair a semântica da linguagem humana expressa em textos e voz. Tem aplicações em várias áreas, como a recuperação da informação, a mineração de texto e o reconhecimento de voz.

Esta lista descreve suscintamente algumas ferramentas úteis para processamento de linguagem natural em textos escritos no idioma inglês.

1. CoreNLP

O Stanford CoreNLP fornece um conjunto de ferramentas de análise de linguagem natural. Ele pode interpretar um texto e dar as formas básicas das palavras, suas partes no discurso, classificar se são nomes de empresas, pessoas, locais e etc., normalizar datas, horários e quantidades numéricas, marcar a estrutura de sentenças em termos de frases e dependências das palavras, indicar a qual substantivo e entidades a frase de refere, indicar sentimento, extrair as relações entre as classes mencionadas, dentre outras funcionalidades.

Conhecendo melhor

http://stanfordnlp.github.io/CoreNLP/index.html

Download

http://stanfordnlp.github.io/CoreNLP/download.html

Instalação

http://stanfordnlp.github.io/CoreNLP/download.html

Testando a ferramenta

http://corenlp.run/

http://nlp.stanford.edu:8080/corenlp/

2. Ilinnois Curator

O Ilinnois Curator é um sistema de gerenciamento de componentes NLP projetado para simplificar o uso e agregação de componentes de PNL, como parte de taggers de fala, taggers de entidades nomeadas, etiquetadores de função semântica e analisadores sintáticos para uso de outras aplicações - incluindo dependências satisfatórias desses mesmos PNL Componentes. O Ilinnois Curator foi desenvolvido em um ambiente Linux. Para usá-lo em um sistema Windows atualmente exige que ele seja instalado em uma máquina virtual.

Conhecendo melhor

https://cogcomp.cs.illinois.edu/page/software_view/Curator

Download

https://cogcomp.cs.illinois.edu/page/download_view/Curator

Instalação

https://cogcomp.cs.illinois.edu/page/download_view/Curator

Testando a ferramenta

http://cogcomp.cs.illinois.edu/curator/demo/index.html

3. GATE

É um software de código aberto capaz de resolver problemas de processamento de texto. Possui um processo definido e repetível para a criação de fluxos de trabalho de processamento de texto robustas e sustentáveis. Pode ter uso ativo para todos os tipos de tarefas de processamento de linguagem e aplicações, incluindo: a voz do cliente; Pesquisa sobre câncer; pesquisa de drogas; apoio à decisão; recrutamento; mineração de web; extração de informação; anotação semântica, dentre outras. É utilizado por empresas, laboratórios de pesquisa e universidades em todo o mundo. Se você precisa resolver um problema com análise de texto ou linguagem de processamento humana, você estará utilizando a ferramenta correta.

Conhecendo melhor

https://gate.ac.uk/

Download

https://gate.ac.uk/download/

Instalação

https://gate.ac.uk/download/

Testando a ferramenta

https://cloud.gate.ac.uk/shopfront/sampleServices

4. SEMAFOR

O SEMAFOR é um analisador semântica de código aberto desenvolvido para fins de pesquisa, o SEMAFOR processa automaticamente frases em inglês de acordo com a forma de análise semântica em Berkeley FrameNet.

Conhecendo melhor

http://www.cs.cmu.edu/~ark/SEMAFOR/

Download

https://github.com/Noahs-ARK/semafor

Instalação

https://github.com/Noahs-ARK/semafor

Testando a ferramenta

http://demo.ark.cs.cmu.edu/parse

5. **OpenNLP**

A biblioteca Apache OpenNLP é um toolkit baseado em aprendizagem de máquinas para o processamento de texto em linguagem natural. Suporta as tarefas de PNL mais comuns, tais como tokenização, segmentação de sentenças, marcação de parte de fala, extração de entidade nomeada, fragmentação, análise e resolução de correferência. Essas tarefas são geralmente necessárias para criar serviços de processamento de texto mais avançados. OpenNLP também inclui entropia máxima e aprendizado de máquina baseada em perceptron.

Conhecendo melhor

https://opennlp.apache.org/index.html

Download

https://opennlp.apache.org/download.html

Instalação

https://opennlp.apache.org/documentation/1.7.2/manual/opennlp.html

6. **LingPipe**

O LingPipe é um conjunto de ferramentas para o processamento de texto usando linguística computacional. O LingPipe é usado para fazer tarefas como: Encontrar os nomes de pessoas, organizações ou locais nas notícias; Classificar automaticamente os resultados de pesquisa do Twitter em categorias; Sugerir ortografia correta das consultas.

Conhecendo melhor

http://alias-i.com/lingpipe/index.html

Download

http://alias-i.com/lingpipe/web/download.html

Instalação

http://alias-i.com/lingpipe/web/install.html

7. FreeLing

O Freeling é uma biblioteca C++ de disponibilização de serviços para análise e processamento de linguagens, sejam estas análises morfológicas, de reconhecimento de dados, tagging, PoS tagging, parsing, dentre outras. Se desenvolve-se, por exemplo, um sistema de tradução automática, e se precisa de algum tipo de processamento linguístico do texto de origem, a aplicação pode chamar módulos Freeling para fazer as análises necessárias.

Conhecendo melhor

http://nlp.cs.upc.edu/freeling/

Download

http://nlp.cs.upc.edu/freeling/node/30

Instalação

http://nlp.cs.upc.edu/freeling/node/8

Testando a ferramenta

http://nlp.lsi.upc.edu/freeling/demo/demo.php

8. NLTK

NLTK é uma plataforma em Python e construída para dar suporte a aplicações em Python para trabalhar com dados de linguagem humana. Ele fornece interfaces fáceis de usar para mais de 50 corpora e recursos lexicais como o WordNet, juntamente com um conjunto de bibliotecas de processamento de texto para classificação, tokenização, stemming, tagging, análise e raciocínio semântico, wrappers para bibliotecas PNL e possui um fórum de discussão ativo.

Conhecendo melhor

http://www.nltk.org/

Download

https://pypi.python.org/pypi/nltk

Instalação

http://www.nltk.org/install.html

9. SpaCy

O spaCy se destaca em tarefas de extração de informações de grande escala. É escrito desde o início em Cython cuidadosamente administrado pela memória. Pesquisa independente confirmou que spaCy é o mais rápido do mundo. Se seu aplicativo precisa processar lixeiras inteiras, spaCy é a biblioteca que você deseja usar.

Conhecendo melhor

https://spacy.io/

Download

https://spacy.io/docs/usage/

Instalação

https://spacy.io/docs/usage/

Testando a ferramenta

https://spacy.io/docs/usage/showcase

10. CLiPS

CLiPS (Linguística Computacional e Psicolinguística) é um centro de investigação associado ao Departamento de Lingüística da Faculdade de Letras da Universidade de Antuérpia e é o resultado da fusão dos centros de investigação CNTS e CPL. O objetivo do CLiPS é produzir pesquisas e recursos reconhecidos internacionalmente em psicolingüística (de desenvolvimento), lingüística (corpus) e linguística computacional, e investigar as combinações interdisciplinares dessas disciplinas.

Conhecendo melhor

http://www.clips.ua.ac.be/

Download

http://www.clips.ua.ac.be/

Instalação

http://www.clips.ua.ac.be/

Testando a ferramenta

http://www.clips.ua.ac.be/demos