



Processamento de Linguagem Natural (PLN)

Introdução

Prof. Julio Cesar Soares dos Reis

Motivação

- Processamento de linguagem natural (PLN) se ocupa das **atividades que envolvem a linguagem humana.**



Motivação

- Linguagem Artificial:
 - Java, Python, C#, Javascript, C++
- Linguagem Natural:
 - Linguagem falada/escrita pelos seres humanos
- *Objetivo*

“Construir mecanismos artificiais (computacionais) que permitam processar, entender e gerar a linguagem que os humanos utilizam naturalmente”.

Exemplo de um Processador de Linguagem

NAME

`wc` - print newline, word, and byte counts for each file

SYNOPSIS

`wc` [OPTION]... [FILE]...

`wc` [OPTION]... --files0-from=F

DESCRIPTION

Print newline, word, and byte counts for each FILE, and a total line if more than one FILE is specified. A word is a non-zero-length sequence of characters delimited by white space.

With no FILE, or when FILE is -, read standard input.

The options below may be used to select which counts are printed, always in the following order: newline, word, character, byte, maximum line length.

-c, --bytes

print the byte counts

-m, --chars

print the character counts

-l, --lines

print the newline counts

--files0-from=F

read input from the files specified by NUL-terminated names in file F; If F is - then read names from standard input

-L, --max-line-length

print the maximum display width

-w, --words

print the word counts

Exemplos de Sistemas



deep neural networks



All

Images

Videos

News

Maps

More

Settings

Tools

About 59,300,000 results (0.57 seconds)

Scholarly articles for deep neural networks

... classification with **deep** convolutional **neural networks** - **Krizhevsky** - Cited by 35680

... game of Go with **deep neural networks** and tree search - **Silver** - Cited by 4256

Deep neural networks for acoustic modeling in speech ... - **Hinton** - Cited by 5358

Deep learning - Wikipedia

https://en.wikipedia.org/wiki/Deep_learning

Deep learning is part of a broader family of machine learning approaches based on learning representations, as opposed to task-specific algorithms. Deep learning is a subset of machine learning, which is a subset of artificial intelligence. Deep learning is a type of machine learning that uses deep neural networks to learn from data. Deep learning is a type of machine learning that uses deep neural networks to learn from data.

A Beginner's Guide to Neural Networks

<https://skymind.ai/wiki/neural-networks>

The three pseudo-mathematical formulas for neural networks: scoring input, calculating output, and updating weights. A Beginner's Guide to Neural Networks · Spiking Neural Networks



who discovered usa



All

Images

Videos

News

Maps

More

Settings

Tools

About 638,000,000 results (0.58 seconds)

Christopher Columbus

The history of the United States began with the settlement of Indigenous people before 15,000 BC. Numerous cultures formed. The arrival of **Christopher Columbus** in the year of 1492 started the European colonization of the **Americas**.

History of the United States - Wikipedia

https://en.wikipedia.org/wiki/History_of_the_United_States

Exemplos de Sistemas



em que ano a UFV foi fundada?



Todas

Notícias

Shopping

Imagens

Maps

Mais

Ferramentas

Aproximadamente 176.000 resultados (0,55 segundos)

UFV / Fundação

30 de março de 1922

Pesquisas relacionadas



Universidade
Federal de
Mina...

7 de setembro
de 1927



Universidade
Federal de
Juiz d...

1960



Universidade
Federal de
Lavras

1908

Exemplos de Sistemas



em que ano a UFV foi fundada?



Todas

Notícias

Shopping

Imagens

Maps

Mais

Ferramentas

Aproximadamente 176.000 resultados (0,55 segundos)

UFV / Fundação

30 de março de 1922

Pesquisas re



Universidade
Federal de
Minas Gerais
7 de setembro
de 1927

ufv data de fundação



Todas

Notícias

Imagens

Maps

Shopping

Mais

Ferramentas

Aproximadamente 502.000 resultados (0,65 segundos)

UFV / Fundação

30 de março de 1922

Pesquisas relacionadas



Universidade
Federal de
Minas Gerais
7 de setembro
de 1927



Universidade
Federal de
Juiz de Fora
1960



Universidade
Federal de
Lavras
1908

Exemplos de Sistemas

The screenshot shows the iFeel web application running in Google Chrome. The browser's address bar displays the URL `blackbird.dcc.ufmg.br:1210/analyseFiles`. The application has a dark blue sidebar with navigation links: Home, Analyse File Texts (active), and Publications. The main content area is titled "Analyse the Sentiment of your files" and includes a user login status "You're logged as Matheus GMAIL" with "Your Files" and "Logout" buttons.

On the left, the "Upload a text file:" section contains an "Upload File" area with a "Click to upload" button and a "Select File Language" dropdown menu currently set to "English". A green "Submit" button is located below these options.

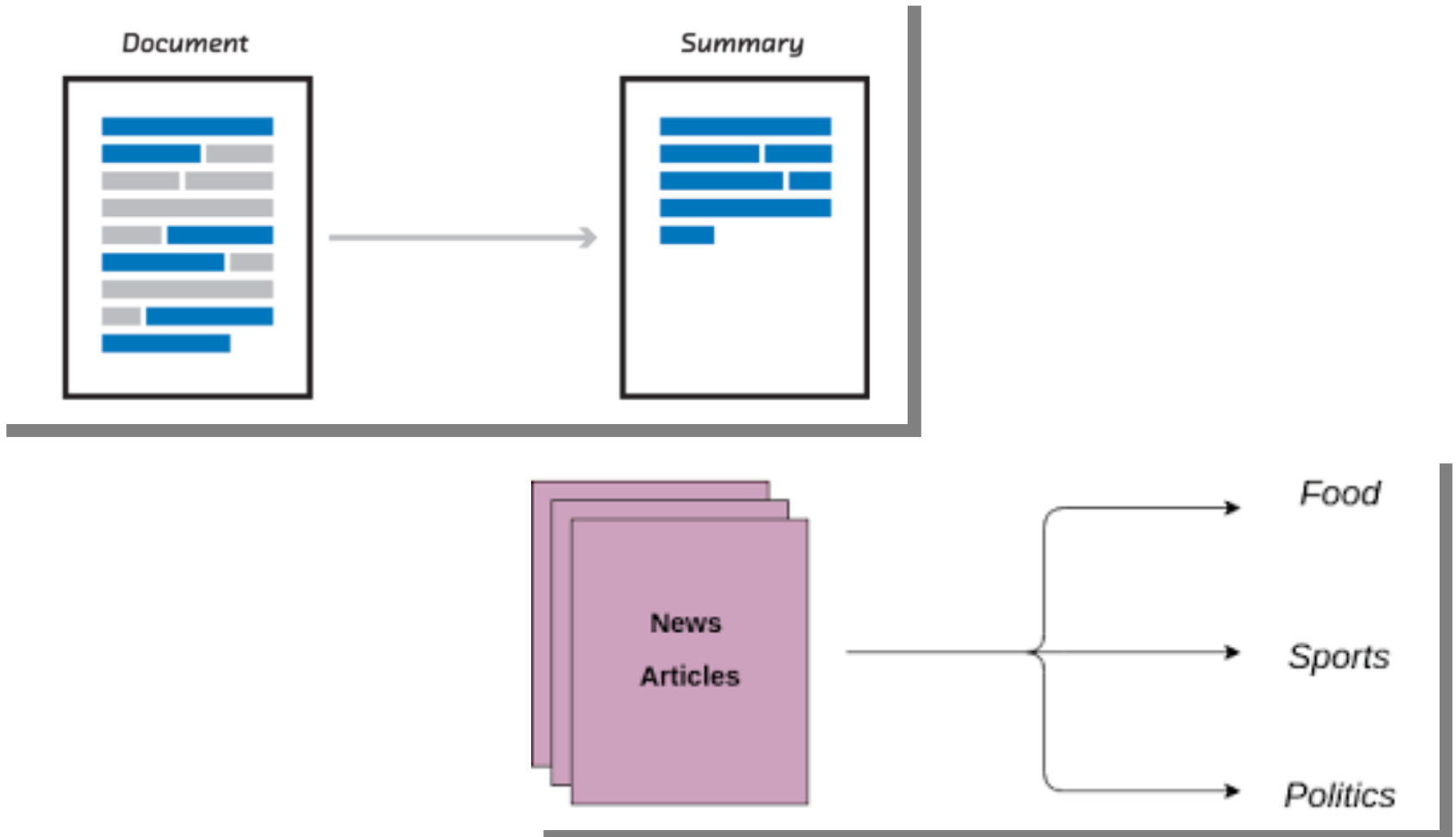
On the right, the "Files Uploaded" section displays a table of analyzed files. A status message above the table reads: "iFeel is analysing: 4 lines right now. Be patient." The table has columns for Download, Name, Status, Language, #Lines, Date, and Delete. It lists ten files, all named "file.txt", with a status of "Complete" and language of "English". Each row includes download links for Excel and XML formats and a red "Delete" button.

Download	Name	Status	Language	#Lines	Date	Delete
Excel XML	file.txt	Complete	English	3 / 3	24-03-17 15:09	Delete
Excel XML	file.txt	Complete	English	3 / 3	25-03-17 02:48	Delete
Excel XML	file.txt	Complete	English	3 / 3	24-03-17 15:03	Delete
Excel XML	file.txt	Complete	English	3 / 3	24-03-17 15:03	Delete
Excel XML	file.txt	Complete	English	3 / 3	24-03-17 15:04	Delete
Excel XML	file.txt	Complete	English	3 / 3	24-03-17 15:04	Delete
Excel XML	file.txt	Complete	English	3 / 3	24-03-17 15:04	Delete
Excel XML	file.txt	Complete	English	3 / 3	26-04-17 13:51	Delete
Excel XML	file.txt	Complete	English	3 / 3	24-03-17 15:05	Delete
Excel XML	file.txt	Complete	English	3 / 3	26-04-17 14:25	Delete

Exemplos de Sistemas

- **Busca usando termos ou palavras-chave:**
 - Relativamente fácil para as plataformas atuais.
 - Os sistemas clássicos usam termos simples.
- **Busca usando perguntas completas:**
 - Relativamente difícil (complexo)
 - São necessários resolução de inferências, capacidade de síntese de informações oriundas de diferentes fontes.

Exemplos de Sistemas



Análise Semântica



what year was newton born?



[Browse Examples](#) [Surprise Me](#)

Input interpretation:

Isaac Newton year of birth

Result:

1642

Time from today:

The year 1642



MD5 "Julio Cesar Soares dos Reis."



[Extended Keyboard](#)

[Upload](#)

[Examples](#)

[Random](#)

Input interpretation:

MD5 Julio Cesar Soares dos Reis.

Message digest:

integer form	54599 140 789 309 490 176 751 787 972 423 334 837
hexadecimal form	2913 697a 099a 05e4 cd7d 5fd8 12b4 3bb5

Análise Semântica



Extended Keyboard

Upload

Examples

Random

Assuming "do you like me" is a phrase | Use as [a question about Alpha](#) instead

Assuming Do you like me? | Use [Do you like people \(humans, children, ...\)?](#) instead

Input interpretation:

Do you like me?

Response:

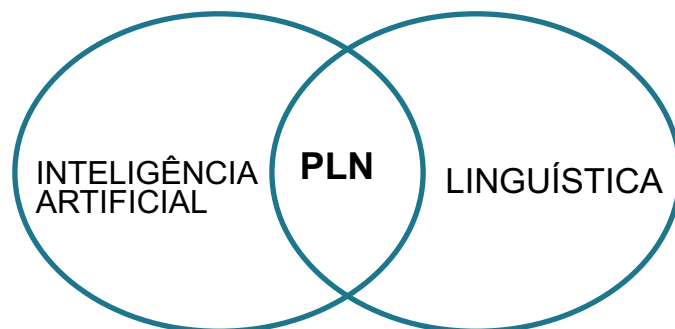
Of course; I like all humans who ask me questions I can answer.

<https://www.wolframalpha.com/>

Similares: Siri, Google Assistant, Cortana

Mas, o que de fato é PLN?

- É uma subárea da **Inteligência artificial** e **Linguística**.



- Definido como **processamento automático (ou semi-automático) da linguagem humana**.
- Nesta aula, o objetivo é apresentar uma visão geral da área.

Tópicos Relacionados

- Compiladores (Autômatos);
- Modelos Probabilísticos;
- Aprendizado de Máquina;
 - Aprendizado Profundo.
- Interação Homem-Máquina;
- Inteligência Artificial.

Porque ainda é um desafio?

- As frases de uma língua podem ser descrições incompletas das informações que pretendem transmitir

- Há alunos lá no DCE

- Há alguns alunos no DCE

- Há 4 alunos no DCE

- Ana, Mariana, Patrick e Igor estão no DCE

Porque ainda é um desafio?

Vou à praia frequentemente	
Vou à praia frequentemente	

Porque ainda é um desafio?

Vou à praia frequentemente	Carioca
Vou à praia frequentemente	Mineiro

Porque ainda é um desafio?

Vou à praia frequentemente	Se for carioca significa no mínimo toda semana
Vou à praia frequentemente	Se for mineiro significa no máximo uma vez mês.

Porque ainda é um desafio?

- A mesma expressão pode significar coisas diferentes em contextos diferentes

O servidor caiu



Porque ainda é um desafio?

- A mesma expressão pode significar coisas diferentes em contextos diferentes

Esta casa é uma zona



Porque ainda é um desafio?

- Alto poder de expressividade da linguagem humana (em alguns casos difícil de entender)
- Léxico complexo
- Gramática com incontáveis construções
- Polissemia
- Como lidar com o sarcasmo e ironia?
- Como lidar com trocadilhos?
- Línguas são aprendidas de forma intuitiva.
- Processamento perfeito é um problema IA-hard

Um erro de uma aplicação de PLN?

Tay: Twitter conseguiu corromper a IA da Microsoft em menos de 24 horas

24/03/2016 às 15:37 • 1 min de leitura



<https://www.tecmundo.com.br/inteligencia-artificial/102782-tay-twitter-conseguiu-corromper-ia-microsoft-24-horas.htm>

Um erro de uma aplicação de PLN?



UNIVERSIDADE FEDERAL
DE MINAS GERAIS

DECOM

RETORIA GERAL

gência de Notícias

Notícias

Boletim UFMG

Revista Diversa

otografia

ssessoria de Imprensa

omunicação Interna

riação Gráfica

ádio UFMG Educativa

/ UFMG

eb UFMG

UFMG Digital

Redes Sociais

Observatório da Web

cretaria Administrativa

pediente

scar notícias

05/set, 13h24 - Coral da OAP se apresenta no Conservatório, nesta quarta

05/set, 13h12 - Grupo de 'drag queens' evoca universo LGBT em show amanhã, na Praça de Serviços

05/set, 12h48 - 'Domingo no campus': décima edição em galeria de fotos

05/set, 9h24 - Faculdade de Medicina promove semana de prevenção ao suicídio

05/set, 9h18 - Pesquisador francês fará conferência sobre processos criativos na próxima semana

05/set, 9h01 - Encontro reunirá pesquisadores da memória e da história da UFMG

05/set, 8h17 - Sessões do CineCentro em setembro têm musical, comédia e ficção científica

05/set, 8h10 - Concerto 'Jovens e apaixonados' reúne obras de Mozart nesta noite, no Conservatório

04/set, 11h40 - Adriana Bogliolo toma posse como vice-diretora da Ciência da Informação

04/set, 8h45 - Nova edição do Boletim é dedicada aos 90 anos da UFMG

04/set, 8h34 - Pesquisador francês aborda diagnóstico de pressão intracraniana por meio de teste

Pesquisadores do DCC identificam em mecanismos de busca na web estereótipos que relacionam raça, idade e beleza feminina

quarta-feira, 24 de agosto de 2016, às 15h29

Seguir @ufmg

Recomendar 150

Curtir 152 mil

Lucas Braga

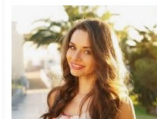


Computação (DCC) do Insti
da UFMG.

Google

mulher bonita

Todas Imagens Vídeos Notícias Maps Mails Configurações Ferramentas



Mulher bonita Fotografias de Ba...
br.depositphotos.com



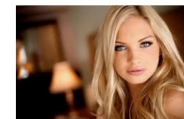
Mulher Bonita Para A Câ...
pixabay.com



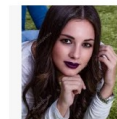
Mulher bonita na rua | Baixar foto...
br.freepik.com



mulher-bonita - Método Reconq...
reconquistar.com



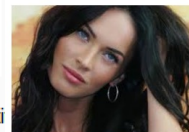
Como chamar a atenção de mulheres bo...
santopappo.com.br



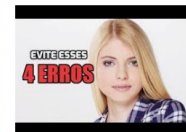
mulher bonita da Costa Ri...
br.depositphotos.com



Essa mulher é m...
entarte.com.br



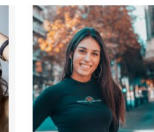
Mariana Kalli: mulher bonita sofre | Donna...
gauchazh.clicrbs.com.br



4 Erros ao Conversar com uma Mulher ...
youtube.com



Jovem, mulher bonita, reti...
br.freepik.com



Barco de imagens - akimb...
pafene.com



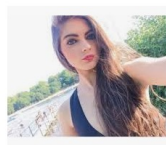
Mulher Bonita Banco de Imagens e Fo...
istockphoto.com



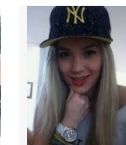
10 coisas que apenas mulheres bonit...
planetappp.com.br



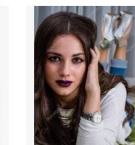
Ex-Panicat Carol Dias des...
cempop.ufol.com.br



Esta mulher foi demitida por ser 'boni...
paraoscuriosos.com



Bruna Santana posa de bo...
diversao.7.com



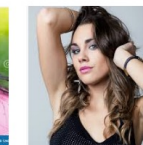
mulher bonita da Costa Ri...
br.depositphotos.com



O que é uma mulher bonit...
duplanoticia.com



Jovem Mulher Bonita Encantado ...
pt.dreamstime.com



Jovem, mulher bonita, reti...
br.freepik.com

Um erro de uma aplicação de PLN?

UFMG

RETORIA GERAL

Agência de Notícias

Notícias

Boletim UFMG

Revista Diversa

Fotografia

Assessoria de Imprensa

Comunicação Interna

Relação Gráfica

Rádio UFMG Educativa

TV UFMG

Web UFMG

UFMG Digital

Redes Sociais

Observatório da Web

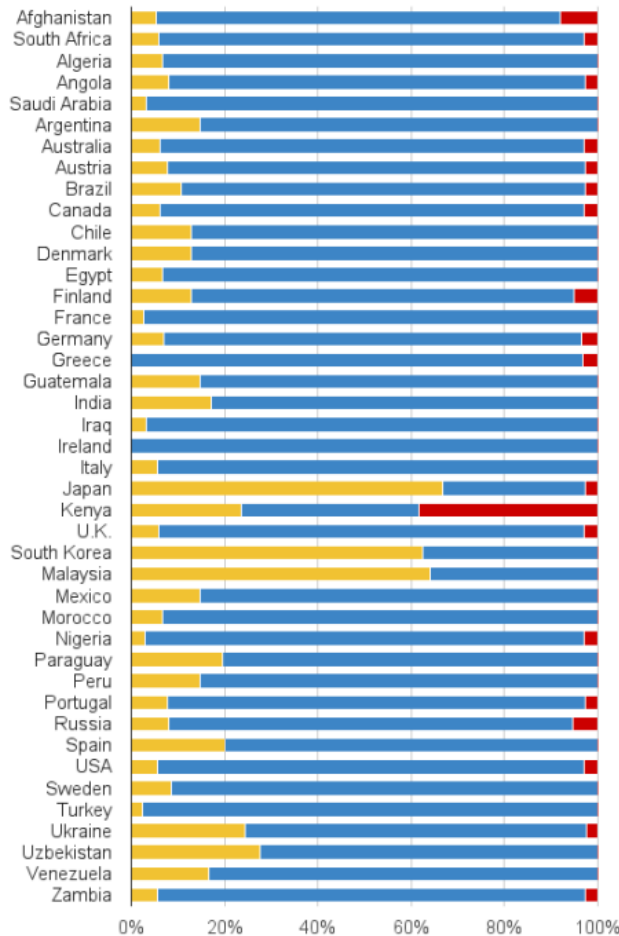
Secretaria Administrativa

Compêndio

Recursos

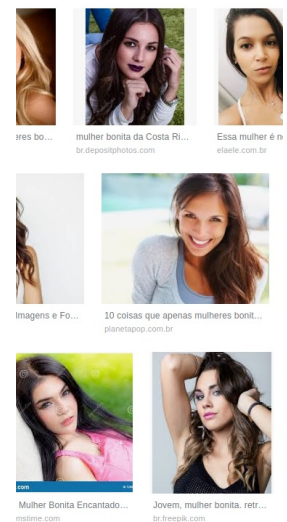
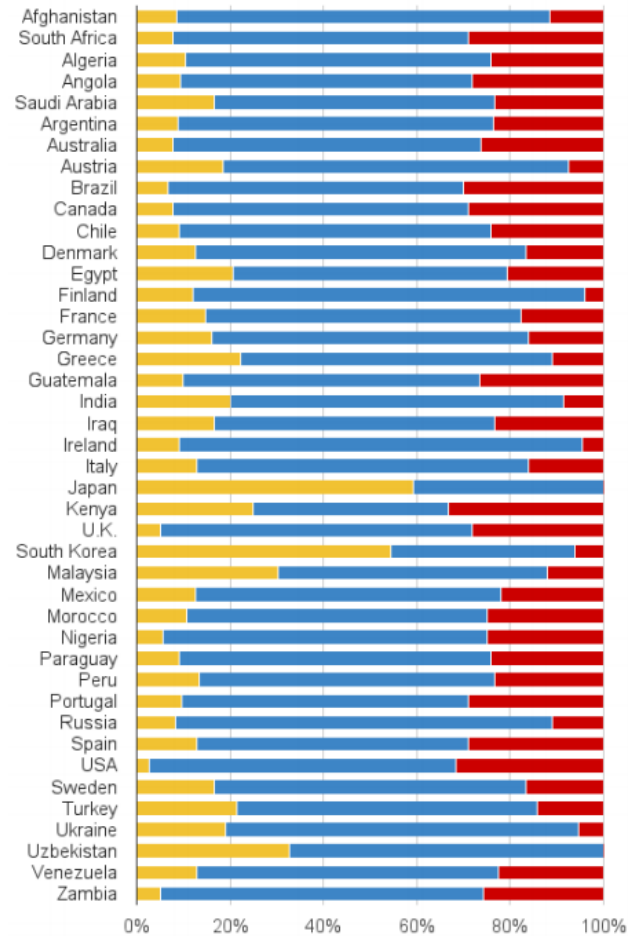
Race Fractions for the query: 'beautiful woman'

Asian White Black

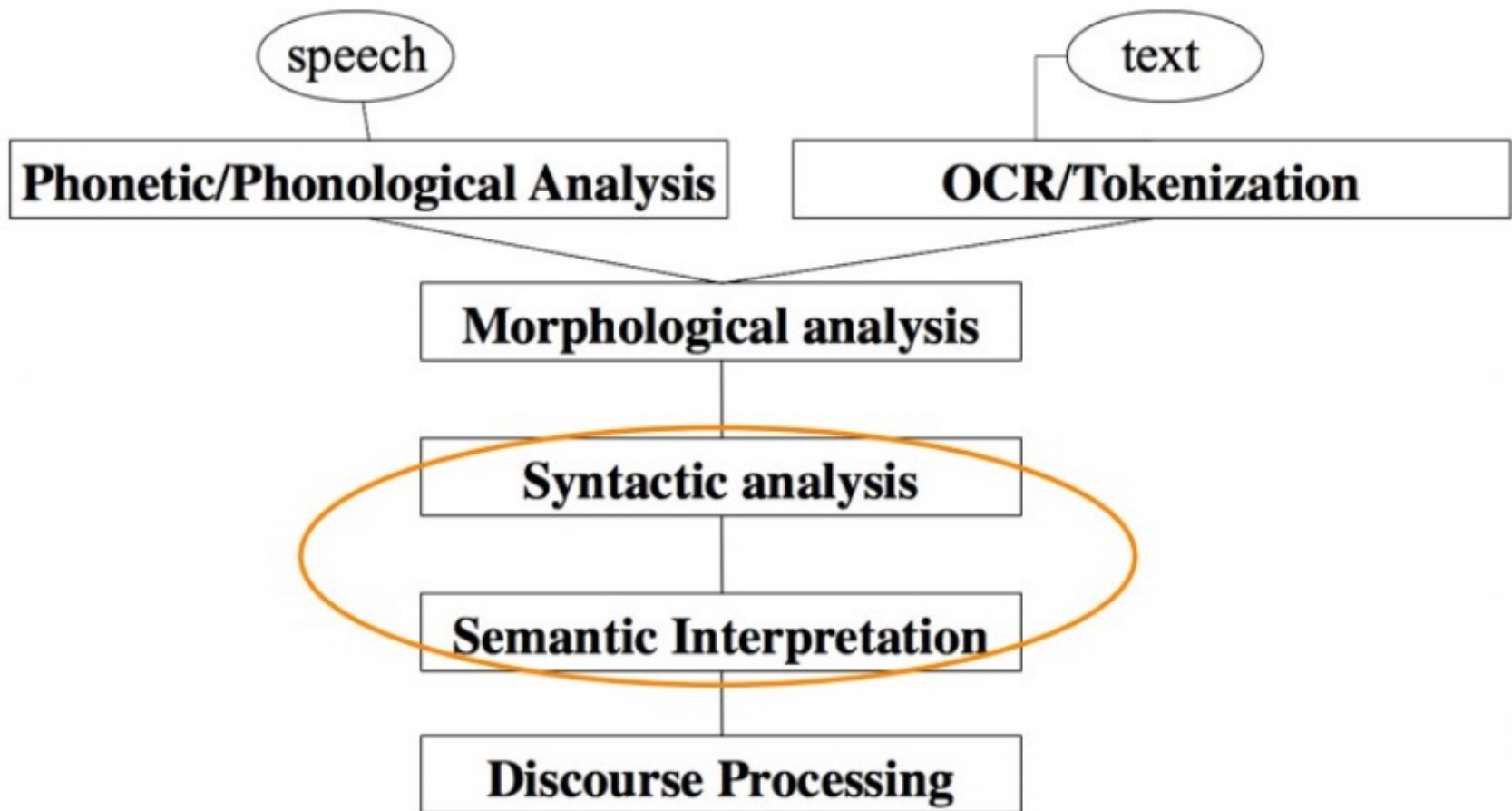


Race Fractions for the query: 'ugly woman'

Asian White Black



O que envolve?



PLN

- ... em Python podemos usar *frameworks* como NLTK, SpaCy, Gensim, OpenNLP
- Processamento de texto básico
- Aprendizagem de máquina
- Modelagem de linguagem
- Tarefas básicas do PLN
- Classificação de textos (análise de sentimentos)
- Descoberta de tópicos
- Panorama de tarefas complexas do PLN

Obrigado!