



# Sumarização Automática

*Prof. Marcelo Pita*

Pós em Inteligência Artificial  
Processamento de Linguagem Natural



# Sumarização automática

**Sumarização automática** consiste na geração automática de resumos a partir de textos mais longos.

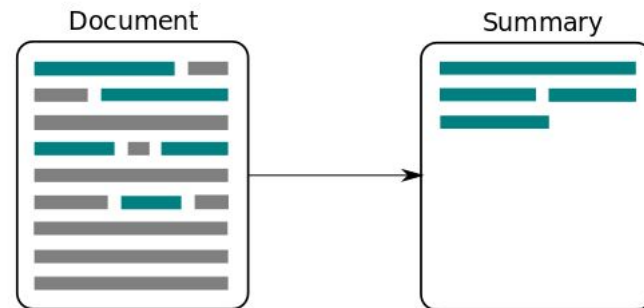
Duas principais abordagens:

- Extrativa
- Abstrativa

# Sumarização automática

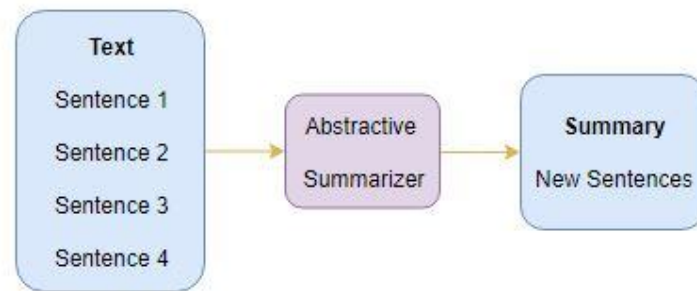
## Abordagem extrativa

Consiste em pontuar sentenças dentro do texto e selecionar as de maior importância para o sumário.



## Abordagem abstrativa

Consiste em gerar novas sentenças com base na representação semântica do texto original.

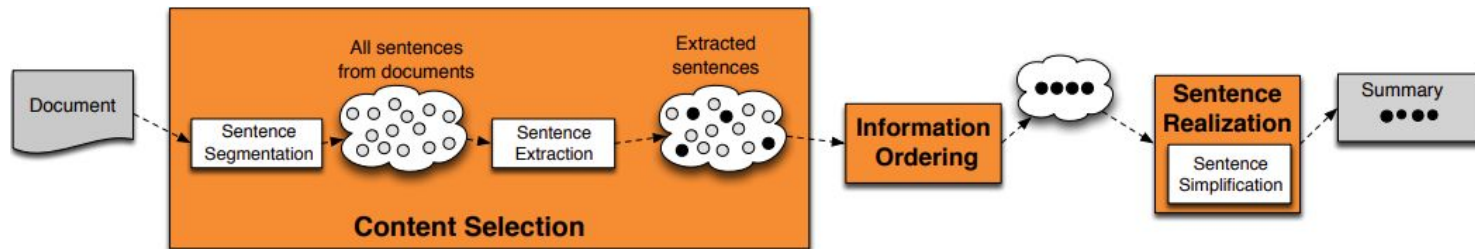


# Sumarização automática

## Abordagem extrativa

Três passos principais:

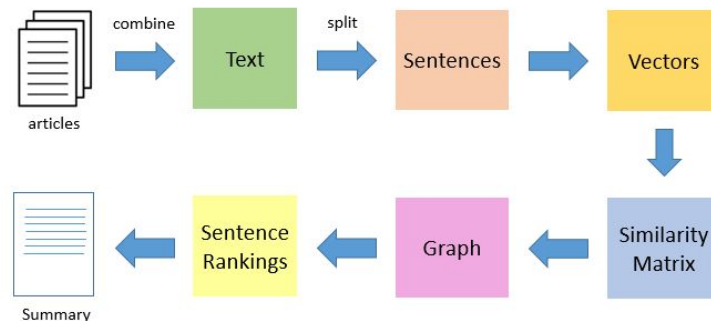
1. Seleção de conteúdo
2. Ordenação da informação
3. Realização da sentença



# Sumarização automática

## O algoritmo TextRank

1. Separar o texto em sentenças (frases)
2. Criar uma matriz TF-IDF para toda a coleção de sentenças
3. Gerar um grafo de similaridade entre sentenças com base na matriz
4. Usar PageRank para avaliar sentenças
5. Selecionar sentenças com maior pontuação

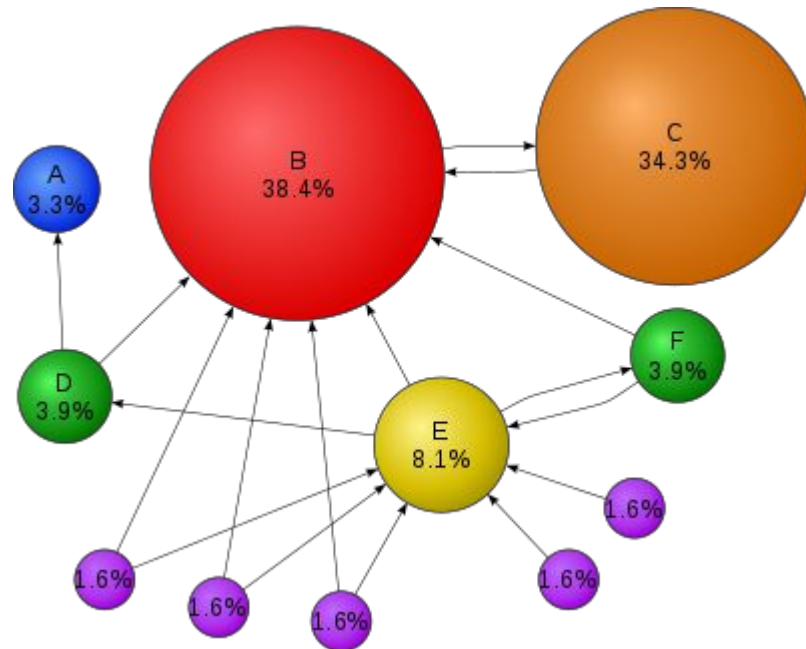


# Sumarização automática

## O algoritmo TextRank

### *PageRank*

- Algoritmo iterativo
- Todas as sentenças começam com a mesma importância
- Valoriza páginas com textos mais referenciados



# Sumarização automática



<Prática>