

# Named-Entity Recognition [NER]

# Named-Entity Recognition

Named-Entity Recognition ou no português, Reconhecimento de Entidades Mencionadas (REM) é a tarefa aplicada na extração e classificação de entidades mencionadas em um texto produzido através de linguagem natural.

# Named-Entity Recognition

Named-Entity Recognition ou no português, Reconhecimento de Entidades Mencionadas (REM) é a tarefa aplicada na extração e classificação de entidades mencionadas em um texto produzido através de linguagem natural.

Neste tipo de tarefa o objetivo é encontrar os limites das classes contidas no texto analisado.

# Named-Entity Recognition

Named-Entity Recognition ou no português, Reconhecimento de Entidades Mencionadas (REM) é a tarefa aplicada na extração e classificação de entidades mencionadas em um texto produzido através de linguagem natural.

Neste tipo de tarefa o objetivo é encontrar os limites das classes contidas no texto analisado.

Através do processamento de linguagem natural o texto original é manipulado, inferências são realizadas para então, classificar cada palavra com sua classe relativa.

# Named-Entity Recognition

ORG  
Manchester United produced a stunning comeback from 2-0  
down to beat ORG  
Newcastle United - and potentially save  
manager PERSON  
José Mourinho from the sack. Second half strikes  
from PERSON  
Juan Mata and PERSON  
Anthony Martial levelled proceedings,  
before an injury-time winner from PERSON  
Alex Sánchez sent the LOC  
Old  
Trafford crowd into raptures. After the international break,

Aplicações

# Named-Entity Recognition

## **Indexação de documentos**

Podemos aplicar o reconhecimento de entidades para a busca de metadados em dados não estruturados. Identificando médico solicitante, data de realização, dados do paciente e conclusões em um laudo médico.

# Named-Entity Recognition

## Algoritmos de busca

Imagine construir um algoritmo para realizar buscas em uma quantidade significativa de documentos. Buscar por um termo de interesse em textos livres pode ser relativamente demorado dependendo do número de documentos em um corpus.

O reconhecimento de entidades pode ser aplicado no objetivo de extrair tags desses documentos visando uma performance maior na tarefa. A busca apenas ocorre na lista de entidades extraídas de cada documento.



# Named-Entity Recognition

## **Recomendação de conteúdo**

Muito utilizado em portais de notícias, onde o foco é manter o visitante entretido e assim navegando pelo maior tempo possível no portal.

Notícias com entidades similares são exibidas para criar interesse ao leitor.

# Named-Entity Recognition

## **Suporte ao cliente**

Imagine uma empresa que possui uma quantidade significativa de produtos. Um cliente acaba de postar uma mensagem na página inicial do Facebook da empresa relatando um problema com um produto específico.

O reconhecimento de entidades pode ser utilizado para, direcionar o responsável pelo SAC do produto em questão para de maneira assertiva resolver o problema deste cliente.

# Named-Entity Recognition

## **Extração de conteúdo textual**

Em um documento com certa padronização, podemos utilizar do reconhecimento de entidades para extrair informações contidas no texto, assim como na identificação de metadados, porém agora estamos descrevendo algo considerando suas entidades vizinhas.

# Estudo de caso

# NER aplicada em laudos radiológicos

Nossa tarefa é através do emprego do reconhecimento de entidades, extrair informações de um laudo de angiotomografia de coronárias.

# NER aplicada em laudos radiológicos

Nossa tarefa é através do emprego do reconhecimento de entidades, extrair informações de um laudo de angiotomografia de coronárias.

A angiotomografia de coronárias é uma tomografia da região torácica onde o objetivo é a análise dos vasos presentes no coração.

# NER aplicada em laudos radiológicos

Nossa tarefa é através do emprego do reconhecimento de entidades, extrair informações de um laudo de angiotomografia de coronárias.

A angiotomografia de coronárias é uma tomografia da região torácica onde o objetivo é a análise dos vasos presentes no coração.

Para cada vaso sumarizado no documento, esperamos extrair suas qualidades descritas e expressar essa informação através de um dado estruturado.

# Descrição do problema

Um número representativo de laudos de angiotomografia de coronárias foi selecionado para ser utilizado como corpus para este problema.



# Descrição do problema

Um número representativo de laudos de angiotomografia de coronárias foi selecionado para ser utilizado como corpus para este problema.

Uma parcela dos documentos foi separada e analisada por radiologistas, o objetivo dessa análise foi identificar os principais termos utilizados para descrever cada qualidade de interesse.

# Descrição do problema

Um número representativo de laudos de angiotomografia de coronárias foi selecionado para ser utilizado como corpus para este problema.

Uma parcela dos documentos foi separada e analisada por radiologistas, o objetivo dessa análise foi identificar os principais termos utilizados para descrever cada qualidade de interesse.

Com essa análise feita, a equipe se reuniu para discutir sobre a padronização dos documentos e qual forma seria mais adequada para processar o texto buscando uma forma automatizada de anotação no momento inicial.

# Descrição do problema

...paciente com escore de cálcio de zero fase com contraste **coronária direita** dominante tronco da **coronária esquerda** com trajeto e calibre normais sem redução luminal bifurca-se em **descendente anterior** e **circunflexa** artéria **descendente anterior** com trajeto e calibre normais **primeiro e segundo ramos diagonais** de moderado calibre ambos sem redução luminal **terceiro ramo diagonal** de pequeno calibre de avaliação limitada ao método artéria **circunflexa** continuando-se com o **segundo ramo marginal** com pequena **placa** **parcialmente calcificada** ostial determinando **irregularidade parietal** **primeiro ramo marginal** de moderado calibre sem redução luminal **coronária direita** com trajeto e calibres normais artérias **descendente posterior** e **ventricular posterior** com trajeto e calibre normais...

# Descrição do problema

Com os termos separados, foram criadas as seguintes classes:

- Vaso
- Trajeto
- Placa
- Composição
- Grau
- Modificador V
- Stent
- Redução Stent
- Enxerto

# Descrição do problema

Cada termo foi organizado em sua classe correspondente criando assim uma lista de termos por classe.

# Descrição do problema

Cada termo foi organizado em sua classe correspondente criando assim uma lista de termos por classe.

Para cada classe então, os termos listados serviram para a criação de uma expressão regular capaz de identificar o padrão em um documento.

# Descrição do problema

Cada termo foi organizado em sua classe correspondente criando assim uma lista de termos por classe.

Para cada classe então, os termos listados serviram para a criação de uma expressão regular capaz de identificar o padrão em um documento.

Todo o corpus foi submetido a um processamento textual:

1. O documento é normalizado;
2. Apenas a seção que descreve a condição dos vasos foi mantida no conteúdo do documento;
3. O documento é submetido às regras criadas para cada classe;
4. Todo termo encontrado é salvo em uma lista, com suas propriedades registradas em uma estrutura tipo chave/valor.

# Descrição do problema

Após todas as entidades computadas, os radiologistas revisaram todo o corpus através de uma plataforma criada para reproduzir as anotações geradas pelo processamento.



# Descrição do problema

Após todas as entidades computadas, os radiologistas revisaram todo o corpus através de uma plataforma criada para reproduzir as anotações geradas pelo processamento.

Após o corpus ser homologado, todos os documentos foram novamente processados para padronizar o formato do dataset:

identificador do documento	palavra	tag
943	calibre	O
943	normais	O
943	bifurcando	adj_posicao
943	em	adj_posicao
943	descendente	adj_posicao

Modelo

# Conditional Random Fields (CRF)

CRF é um modelo estatístico com aplicabilidade para reconhecimento de padrões em previsões estruturadas sequenciais.

# Conditional Random Fields (CRF)

CRF é um modelo estatístico com aplicabilidade para reconhecimento de padrões em previsões estruturadas sequenciais.

Leva em consideração a vizinhança de um dado sequencial para prever a qual classe cada parcela pertence.

# Conditional Random Fields (CRF)

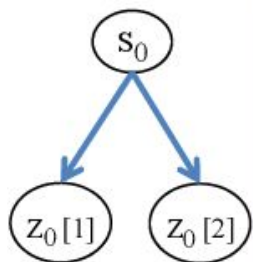
CRF é um modelo estatístico com aplicabilidade para reconhecimento de padrões em previsões estruturadas sequenciais.

Leva em consideração a vizinhança de um dado sequencial para prever a qual classe cada parcela pertence.

Utilizado nos campos de processamento de linguagem natural e sequenciamento em bioinformática.

# Conditional Random Fields (CRF)

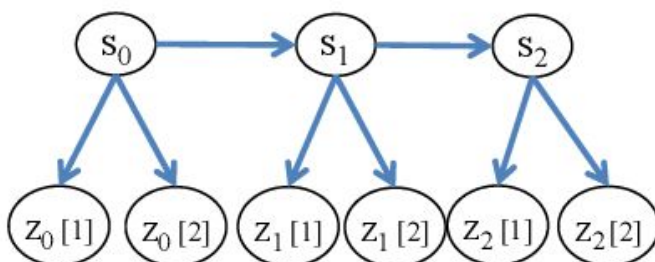
Bayes Model



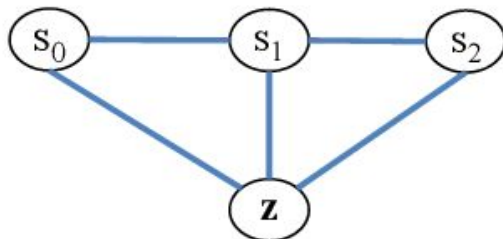
Max. Entropy Model



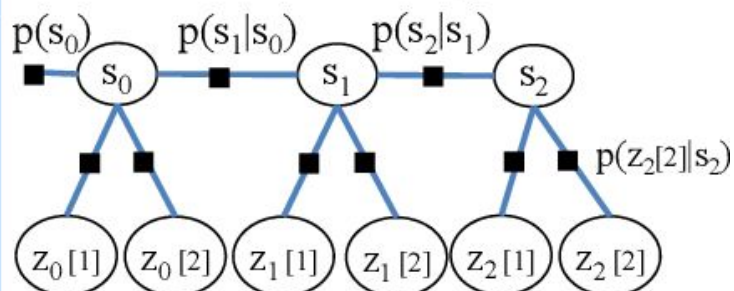
Hidden Markov Model



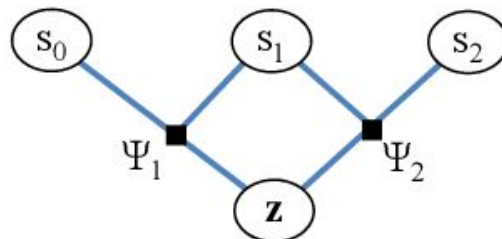
Linear Chain Conditional Random Field



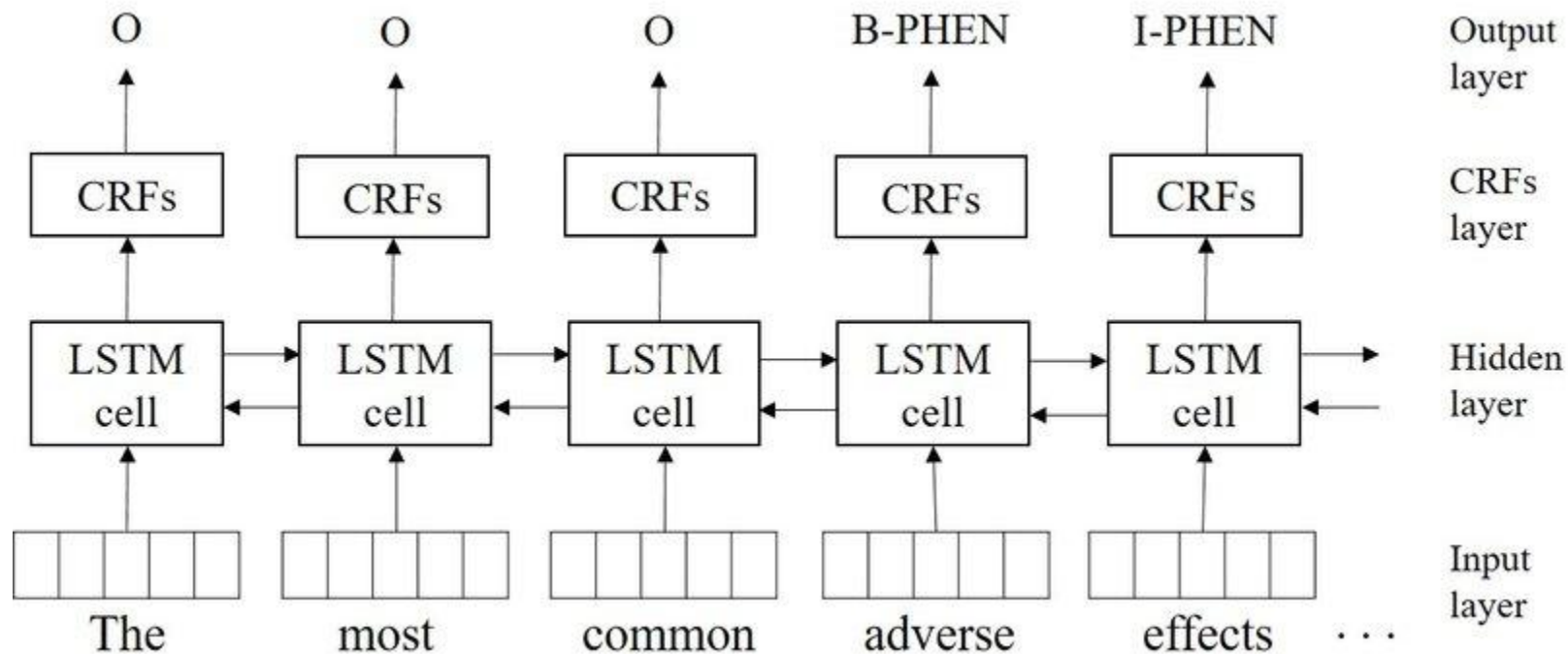
HMM Factor Graph



Linear Chain CRF Factor Graph



# LSTM Bidirecional + CRF



Laboratório proposto, clique [aqui!](#)