
Processamento de Linguagem Natural

Aula 06

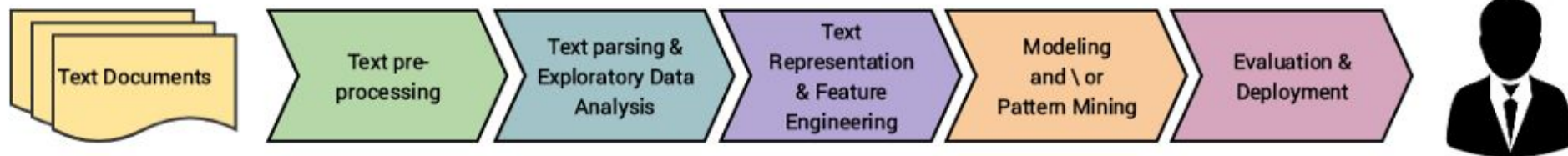
Arquitetura de aplicação para
Processamento de Linguagem Natural



PUC Minas

Workflow do PLN

- Ao longo das aulas vimos as diversas tarefas relacionadas ao processamento de linguagem natural



- Iniciamos com um *corpus* de documentos de texto
- Processos padrão de seleção e pré-processamento dos texto seguidos de análise básica de dados exploratórios.

Workflow do PLN

- Com base nos insights iniciais, geralmente representamos o texto usando técnicas como frequência de termos.
- Dependendo do problema em questão, nos concentramos na criação de modelos preditivos supervisionados ou modelos não supervisionados, que geralmente se concentram mais na mineração de padrões e no agrupamento.
- Por fim, avaliamos o modelo e os critérios gerais de sucesso com partes interessadas ou clientes relevantes e implantamos o modelo final para uso futuro.

Corpus - Documentos de Texto

- Seleção dos textos
- Web-scrapping de páginas
- Carregamento de arquivos

Limpeza e Pré-Processamento

- Remoção de tags HTML
- Tratamento de caracteres especiais
- Expansão de formas contraídas de palavras
- ... as várias outras etapas de pré-processamento necessárias para que o texto possa ser melhor tratado

Parsing do Texto

- Nesta etapa entendemos a estrutura e a sintaxe do texto
- As quebras facilitam o entendimento do texto e mostram como algumas palavras são utilizadas de forma agrupada
 - POS Tagging - Tagueamento de partes do discurso
 - N-Gramm Tagger - Quebra do texto em pedaços

Representação Textual

- Através das entidades nomeadas começamos a ver alguns significados para cada uma das palavras ou grupo de palavras

Análise de Sentimento

- Análise de Sentimento é uma das aplicações mais populares de PLN
 - Muitos cursos e tutoriais
 - Datasets com análise de sentimento de pesquisas corporativas a comentários sobre filmes
- Aspecto chave é entender a opinião expressada por um texto
 - Geralmente a polaridade é quantificada com um valor positivo ou negativo
 - O sentimento geral inferido pode ser positivo, neutro ou negativo

Ferramentas Comerciais

- Dificuldade de desenvolvimento específico
- Ferramentas comerciais customizáveis
 - IBM Watson
 - Microsoft
 - Google
 - Amazon

Amazon AI Services

- Com o AI Services da AWS, você pode usar recursos como análise de imagem e vídeo, linguagem natural, recomendações personalizadas, assistentes virtuais e previsão para seus aplicativos sem um profundo conhecimento em aprendizado de máquina..

<https://aws.amazon.com/pt/machine-learning/ai-services/>



Amazon AI Services

- Amazon Comprehend
 - Serviço de Natural Language Processing (NLP – Processamento de linguagem natural) que usa Machine Learning para encontrar insights e relações dentro de documentos em texto. Nenhuma experiência de Machine Learning necessária.
- Amazon Lex
 - Serviço para a criação de interfaces de conversa com funcionalidades de aprendizado e Natural Language Understanding, para o reconhecimento da intenção do texto
- Amazon Textract
 - Serviço que extrai automaticamente texto e dados de documentos digitalizados.

IBM Watson

- IBM Watson® para Processamento de Linguagem Natural, Reconhecimento de Imagens e Aprendizado de Máquina

<https://www.ibm.com/cloud/ai>

- Conjunto de várias APIs com propósitos e valores diferente

IBM Watson

- Watson Discovery
 - Uncover connections in data by combining automated ingestion with advanced AI functions.
- Watson Natural Language Classifier
 - Interpret and classify natural language with confidence.
- Watson Natural Language Understanding
 - Analyze text to extract metadata from content such as concepts, entities and sentiment.
- Watson Tone Analyzer
 - Analyze emotions and tones in written content.

Google Cognitive Services

- Hub de AI com diversas APIs que poderão ser conectadas entre si

<https://cloud.google.com/products/ai/>

Google Cognitive Services

- AutoML Natural Language
 - Treine modelos de ML para classificar, extrair e detectar sentimentos com mínimo esforço e especialização de machine learning usando o AutoML Natural Language. Você pode usar a IU do AutoML Natural Language para fazer upload de seus dados de treinamento e testar seu modelo personalizado.
- API Natural Language
 - Modelos pré-treinados da API Natural Language permitiram recursos de processamento de linguagem natural, incluindo análise de sentimento, análise de entidade, análise de sentimento de entidade, classificação de conteúdo e análise de sintaxe.

Microsoft Cognitive Services

- Serviços de linguagem que garantem que aplicativos e serviços possam compreender o significado de texto não estruturado ou reconhecer a intenção por trás das declarações de um locutor.

<https://azure.microsoft.com/pt-br/services/cognitive-services/>

Microsoft Cognitive Services

- Análises de texto
 - Reconhecimento de Entidade Nomeada
 - Extração de expressão chave
 - Análise de sentimento em textos
- QnA Maker
 - Extração de Perguntas e Respostas de texto não estruturado
 - Criação de bases de dados de conhecimento com base em coleções de Perguntas e Respostas
 - Combinação semântica para bases de dados de conhecimento



Microsoft Cognitive Services

- LUIS

- <https://azure.microsoft.com/en-us/services/cognitive-services/language-understanding-intelligent-service/>



Feedbacks

<http://bit.ly/nlp20192>