



A UNIVERSIDADE DE BRASÍLIA

APRESENTA UMA INOVAÇÃO

EXTRAÇÃO DE ENTIDADES NOMEADAS DO DIÁRIO OFICIAL DA UNIÃO (DOU) COM PROCESSAMENTO DE LINGUAGEM NATURAL (PLN)

BR 51 2019 002939 9

Grupo: Ciências exatas e informática

Subgrupo: Inteligência artificial

INTELIGÊNCIA ARTIFICIAL PARA PROCESSAMENTO DE LINGUAGEM NATURAL NA ADMINISTRAÇÃO PÚBLICA

O processamento de linguagem natural (PLN) é uma área da inteligência artificial que estuda como os computadores podem entender e gerar textos em linguagem humana. Essa tecnologia pode ser útil na administração pública, pois permite automatizar tarefas que envolvem o processamento de documentos, como classificação, extração de informações, resumo, tradução e geração de textos, além de possibilitar a análise de grandes volumes de dados textuais. As principais vantagens do PLN são a rapidez, a precisão e a escalabilidade, pois as máquinas podem processar grandes volumes de dados em pouco tempo e com baixo custo. No entanto, o PLN também enfrenta desafios, como a diversidade e a complexidade das línguas humanas, a necessidade de dados de qualidade e de ética no uso da tecnologia, e a garantia de segurança e privacidade dos dados processados.

Sob essa perspectiva, pesquisadores da Universidade de Brasília (UnB), desenvolveram um programa de computador intitulado “Construção de um Corpus para extrair entidades nomeadas do Diário Oficial da União utilizando Aprendizado Supervisionado”. O programa é uma ferramenta de processamento de linguagem natural (PLN) que utiliza o Diário Oficial da União (DOU) como fonte de informação para identificar e classificar entidades nomeadas, tais como pessoas, organizações, locais, datas e valores. O programa também propõe a construção de um corpus anotado manualmente com as entidades nomeadas extraídas do DOU, e avalia o desempenho do método de extração utilizando esse corpus como referência. Além disso, o



A UNIVERSIDADE DE BRASÍLIA **APRESENTA UMA INOVAÇÃO**

programa implementa uma aplicação que mostra os resultados das extrações do corpus utilizado, facilitando a verificação e a análise dos dados.

VANTAGENS

- **Segurança:** Utiliza o Diário Oficial da União como fonte de informação confiável e atualizada para identificar e classificar entidades nomeadas;
- **Eficiência:** Avalia o desempenho do método de extração utilizando o corpus do DOU, verificando a qualidade e a precisão dos resultados; Demonstra os resultados das extrações do corpus utilizado, facilitando a verificação e a análise dos dados; Contribui para o avanço do conhecimento sobre o reconhecimento de entidades nomeadas em português.
- **Escalabilidade:** Processamento de grandes volumes de dados em pouco tempo.

Agenda 2030 da ONU:



Gostou dessa tecnologia?

Para obter mais informações entre em contato com a Agência de Comercialização de Tecnologias (ACT) da Coordenação de Inovação e Transferência de Tecnologia (CITT), por meio dos contatos a seguir:

E-mail: act@cdt.unb.br

Telefone: (61) 3107-4116