

RECUPERAÇÃO DE DADOS DE EVENTOS DE FOGO EM GRANDES BASES DE DADOS UTILIZANDO PROCESSAMENTO DE LINGUAGEM NATURAL

Matheus Venancio Moreira Sales¹; Fabrício Galende Marques de Carvalho²; William Max S S Silva³

FATEC JACAREÍ / INPE¹ (matheusvmsales@hotmail.com)

INPE / FATEC SÃO JOSÉ DOS CAMPOS / FATEC JACAREÍ². (fabricio.galende@inpe.br)

FATEC JACAREÍ / INPE³. (max.fisuefs@gmail.com).

Resumo: O monitoramento ambiental de eventos de fogo exige sistemas de consulta que combinem agilidade no acesso às informações e precisão na recuperação dos dados, uma vez que, tanto a complexidade como o volume de dados geoespaciais podem representar uma barreira significativa para usuários em contextos críticos. No Instituto Nacional de Pesquisas Espaciais (INPE), o projeto DescrEVE Fogo é o responsável por prover dados e informações de eventos de fogo para serem utilizados em contextos tanto de pesquisa como operacionais, tais como ações de combate aos incêndios florestais. Este artigo trata do desenvolvimento de estratégias para recuperação de dados e informações sobre eventos de fogo utilizando técnicas de Processamento de Linguagem Natural (PLN). O objetivo é desenvolver um serviço web capaz de converter consultas em linguagem natural para comandos estruturados de consulta a bases de dados, incluindo extensões geoespaciais, tais como PostGIS, utilizando a arquitetura Text-to-SQL determinística. A solução desenvolvida leva em conta a segurança da informação e exatidão matemática, empregando PLN clássico e templates pré-programados, reduzindo riscos de inconsistências ou alucinações de dados. Além dos aspectos já mencionados, que incluem tanto a recuperação de dados geoespaciais como temporais, o projeto do serviço supõe que as respostas às requisições serão dadas sob a forma de texto que deve ser claro, simples e bem estruturado, eliminando barreiras técnicas e permitindo acesso intuitivo com rigor técnico e integridade dos dados. O projeto de solução e o protótipo funcional, construído como uma prova de conceito, assegura que métricas de área e tempo correspondam de modo acurado aos registros armazenados, apoiando a gestão ambiental e a tomada de decisão. A metodologia utilizada baseia-se em um pipeline de PLN composta por vetorização semântica, classificação de intenções e extração de entidades. Para representação da terminologia especializada, utilizou-se Word2Vec no modelo Continuous Bag of Words (CBOW), implementado com Gensim, permitindo a conversão de termos textuais em vetores numéricos pertencentes ao espaço latente de representação. O reconhecimento de intenções ocorre por meio do cálculo da Similaridade de Cosseno entre a consulta do usuário e centróides previamente definidos. A tecnologia utilizada inclui a linguagem de programação python, bibliotecas tais como NumPy. A extração de entidades a partir do texto fornecido pelo usuário utiliza expressões regulares para identificar variáveis temporais e converter numerais extensos em valores inteiros aplicados às cláusulas SQL. Os dados são armazenados em PostgreSQL com extensão PostGIS, sendo a interface mediada pelo ORM SQLAlchemy. A avaliação da técnica empregada incluiu Dendrogramas de Ward e matrizes de confusão para cálculo de métricas como precisão, sensibilidade e *F1-score*. Além dos aspectos já citados, o serviço desenvolvido gera saídas estruturadas com visualização clara dos atributos dos eventos de fogo, permitindo ordenação e filtragem por critérios geográficos, magnitude em ares e organização cronológica. Portanto, pode-se afirmar que o serviço desenvolvido além de viável e eficaz para recuperação de dados e informações em grandes bases de dados

geoespaciais, é uma ferramenta importante e aplicável à tomada de decisão crítica em contextos de gestão ambiental, em especial aquele envolvendo eventos de fogo

Palavras-chave: Monitoramento Ambiental; Processamento de Linguagem Natural; PostGIS; Sistema de Apoio à Decisão.