

Jeancarlo Campos Leão; Abdiel Batista Santos; Joice Gabrielle Souza Silva
Poster, resultado inédito.

Uma Plataforma Web Escalável para Mineração de Dados

Resumo

A análise de dados é uma etapa fundamental da pesquisa científica e geralmente envolve grande volume de dados. Neste contexto, o uso de técnicas adequadas e eficientes de mineração de dados para extrair informações específicas exige habilidades de programação e de gerenciamento de infraestrutura que nem sempre estão previstos no cronograma, custo e escopo de um projeto de pesquisa. O objetivo deste trabalho é apresentar o JKnowledgeFlow (JKF), uma plataforma de gerenciamento de experimentos científicos que possui uma arquitetura escalável e uma interface simples e de fácil utilização. Nesta plataforma, podem ser feitas requisições para análise de quantidade massiva de dados, por um conjunto predefinido de técnicas de mineração e sem a necessidade de programação ou conhecimento específico sobre as tecnologias necessárias para uso da técnica. Uma plataforma como essa pode contribuir em acessibilidade e produtividade ao permitir que diferentes áreas do conhecimento possam abstrair algumas habilidades e conhecimentos específicos da computação. Para atingir o objetivo de simplicidade e escalabilidade, a plataforma JKF foi concebida com base no uso dos seguintes componentes arquiteturais e de software: computação na nuvem, para dispensar o usuário de adquirir, implantar e manter infraestrutura tecnológica; processamento e armazenamento distribuídos (opcionalmente na máquina do usuário), para maior aproveitamento do hardware de múltiplos computadores de pequeno porte; deduplicação automática de dados armazenados; painel de gerenciamento de experimentos, centralizado na Web com controle de acesso a dados e módulos, além do monitoramento individual das replicações de cada experimento; flexibilidade para serem acoplados novos módulos, com técnicas de mineração que atendam a novas demandas; gerenciamento semi automatizado de hosts multiplataforma e de configuração do software utilizado nos experimentos. A plataforma tem desempenhado um papel importante como interface para prospecção científica por múltiplas áreas do conhecimento e da utilização de diferentes técnicas para mineração de dados. Em sua primeira versão, foi utilizada na mineração de grafos e na extração de padrões em enquetes acadêmicas, com o intuito melhorar a interpretação de respostas a questionários. Além disso, diversas outras técnicas estão sendo utilizadas nessa ferramenta. Nesses casos, verificamos a facilidade de uso da plataforma e avaliamos a qualidade dos resultados obtidos. Ademais, a plataforma se apresentou apta a satisfazer demandas reais de computação científica de múltiplas áreas do conhecimento. Em trabalhos futuros pretendemos viabilizar o uso da plataforma pela comunidade de pesquisadores do IFNMG.

Referências

LEÃO, J. C. et al. Arquitetura de uma Plataforma de Computação em Nuvem para Prospecção Científica. Anais do VIII Seminário de Iniciação Científica do IFNMG, 03 2019. ISSN 2238-085X. Disponível em: <https://jcloud.net.br/url/2019SIC_JKF/>.

LEÃO, J. C. et al. Classificação de Relações Sociais para Melhorar a Detecção de Comunidades. In: Proc. VI Brazilian Work. Soc. Netw. Anal. Min. São Paulo, SP, Brazil: [s.n.], 2017. v. 6, n. 1/2017, p. 647–657. ISSN 2595-6094. Disponível em: <<https://jcloud.net.br/url/LBDL2017VIBraSNAM/>>.

LEÃO, J. C. et al. Mineração de Perfis Sociais em Redes Temporais. In: XXXII Simpósio Brasileiro de Banco de Dados - SBBD 2017. Uberlândia-MG: [s.n.], 2017. p. 264–269. ISSN 2316-5170. Disponível em: <<https://jcloud.net.br/url/LBDL2017SBBD/>>.

LEÃO, J. C. et al. Mineração de Dados Demográficos de Minas Gerais : Uma Proposta para Clusterização de Perfis do Arranjo Familiar. Anais dos Simpósios de Informática do IFNMG - Campus Januária, n. 2, 06 2016. ISSN 2447-3847. Disponível em: <<https://jcloud.net.br/url/Leao2016CENSOjanuaria/>>.

MENEZES, L. S. et al. Um Arcabouço para Integração de Microdados da Pesquisa Nacional por Amostra de Domicílios: Aplicação na Análise dos Fatores de Influência sobre a Renda do Trabalho na Região Sul do Brasil. Anais dos Simpósios de Informática do IFNMG - Campus Januária, n. 3, 06 2018. ISSN 2447-3847. Disponível em: <<https://jcloud.net.br/url/Lara2018PNADXSimpósioJanuaria/>>.