Plan Overview

A Data Management Plan created using DMPTool

Title: Extração de conhecimento em Big Data

Creator: Carlos roberto Valêncio

Affiliation: São Paulo State University (unesp.br)

Principal Investigator: Angelo Cesar Colombini

Funder: Universidade Federal Fluminense

Template: Digital Curation Centre

Project abstract:

A agressividade com que o volume de dados cresce tem desafiado as tecnologias no contexto de manipular, tratar e extrair conhecimento a um custo computacional e operacional capaz de atender de forma efetiva a demanda. O objeto de estudos deste projeto, assenta-se sobre dois pontos básicos: manipular e extrair - a busca de algoritmos e processos capazes de tornar eficiente e minimamente viável a gestão do conhecimento; tratar - a detecção de dados atípicos ou distantes da média *outliers,* que podem adicionar distorções e custos na fase de limpeza e a redução da dimensionalidade, espera-se com isso um aumento na precisão dos resultados e na confiabilidade, reduzindo problemas de ajustes, conhecidos como *overfitting*.

Start date: 01-02-2023

End date: 01-02-2024

Last modified: 12-06-2022

Extração de conhecimento em Big Data

Data Collection

What data will you collect or create?

Os dados utilizados são de banco de dados abertos, ou seja, coleções de dados disponíveis gratuitamente.

Este trabalho efetuará também o levantamento dos principais trabalhos da literatura científica e por meio de bibliotecas digitais.

How will the data be collected or created?

A coleta dos dados será efetuada pelos links que disponibilizam os bancos de dados abertos, como por exemplo: Dados Abertos Linguísticos Vinculados (linguistic-lod.org).

Os textos científicos serão obtidos por meio bibliotecas digitais disponíveis pelas instituição em que os pesquisadores estão vinculados, como por exemplo: http://www-periodicos-capes-gov-br.ez87.periodicos.capes.gov.br/index.php?

Documentation and Metadata

What documentation and metadata will accompany the data?

Todos os dados utilizados estarão acompanhados por documento que descrevem as tabelas, atributos e respectivos tipos de dados. A documentação e os dados ficarão disponíveis nos repositórios institucionais das Entidades envolvidas neste projeto.

Ethics and Legal Compliance

How will you manage any ethical issues?

Não se aplicam as questões éticas para estes bancos de dados abertos e objetos desta proposta. Em sendo pertinente, será efetuada a devida solicitação de autorização para o Comitê de Ética do âmbito recomendado.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

As Instituições de ensino e pesquisa envolvidas neste projeto tem as respectivas agências de inovações e que tratam do suporte necessário e amparada em toda a regulamentação já existente.

Storage and Backup

How will the data be stored and backed up during the research?

Será mantido uma cópia do conjunto de dados abertos nos repositórios das Instituições envolvidas.

How will you manage access and security?

O acesso aos bancos de dados abertos a serem utilizados é livre. O acesso as respectivas cópias estarão disponíveis aos pesquisadores envolvidos durante o período da execução do projeto.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Todos os dados utilizados por este trabalho estarão disponíveis por meio das respectivas publicações científicas e demais relatórios nos repositórios institucionais, de modo a atender as restrições de compartilhamento e divulgação.

What is the long-term preservation plan for the dataset?

As Instituições envolvidas disponibilizam repositórios para o armazenamento e manutenção dos dados utilizados pelos projetos desenvolvidos em seus respectivos âmbitos. Deste modo, este trabalho contará com estes suportes institucionais e respectivas políticas de preservação.

Data Sharing

How will you share the data?

Por meio do armazenamento nos repositórios institucionais das entidades envolvidas e, durante a execução do projeto, por meio de permissão de acesso ao servidor do Grupo de Banco de Dados/Ibilce/Unesp.

Are any restrictions on data sharing required?

Durante o desenvolvimento deste projeto, a restrição de acesso ao servidor do Grupo de Banco de Dados/Ibilce/Unesp, onde estarão alocados os dados utilizados, se fará por fornecimento de senhas que habilitam utilização por meio do respectivo software de gerenciamento.

Responsibilities and Resources

Who will be responsible for data management?

O pesquisador coordenador do projeto e o pesquisador responsável pela execução do mesmo.

What resources will you require to deliver your plan?

As instituições envolvidas possuem todos os equipamentos físicos, infraestrutura de comunicação de dados, recursos humanos técnico-especialistas e restrições de segurança necessários e exigidos para o gerenciamento do plano de dados deste projeto.
