

Belo Horizonte - MG



38º Simpósio Brasileiro de Bancos de Dados

38th Brazilian Symposium on Databases

Anais do SBBD 2023

Proceedings of SBBD 2023

De 25 a 29 de setembro de 2023

Execução:



Realização:



Apoio Acadêmico:



Patrocínio:



Agências de fomento:



Organização do SBBD 2023

SBBD 2023 Organization

Program Chair

Daniel Kaster (UEL, Brazil)

Short Papers Chair

Eduardo Ogasawara (CEFET/RJ, Brazil)

Demos and Applications Chair

Marcos Bedo (UFF, Brazil)

WTDBD Chair

Renata Galante (UFRGS, Brazil)

CTDBD Chair

Eduardo C. de Almeida (UFPR, Brazil)

Short Courses Chair

Humberto Razente (UFU, Brazil)

Tutorials Chair

Marcos André Gonçalves (UFMG, Brazil)

Workshops Chair

Elaine Sousa (USP, Brazil)

WTAG Chair

Marcelo Iury (UFPB, Brazil)

Execução:



Apoio Acadêmico:



Patrocínio:



Agências de fomento:

Proceedings Chair

Ticiania L. Coelho da Silva (UFC, Brazil)

Sponsorships and Institutional Contacts Chair

Wladimir Cardoso Brandão (PUC Minas, Brazil)

Organização Geral

General Organization

Chairs

Michele Amaral Brandão (IFMG, Brazil)

Felipe Domingos da Cunha (PUC Minas, Brazil)

Execução:



Realização:



Apoio Acadêmico:



Patrocínio:



Agências de fomento:



Os Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados (SBBB 2023) incluem os artigos selecionados e apresentados na edição do evento realizado de 25 a 29 de setembro de 2023. Nesta edição, os anais incluem 25 artigos completos e 24 artigos curtos. Esses trabalhos foram selecionados por meio de um processo de revisão por pares do tipo double-anonymized, com um total de 51 artigos completos e 55 artigos curtos submetidos, o que implicou em taxas de aceitação de 49% (25/51) e 42,5% (23/54), respectivamente. Um artigo originalmente submetido para a trilha de artigos completos foi aceito como artigos curtos. Tanto os artigos completos quanto os artigos curtos incluem trabalhos das categorias pesquisa, indústria e visão, em diversas áreas de bancos de dados, ciência de dados e big data. O comitê de programa de Artigos Completos foi coordenado pelo Prof. Daniel Kaster (UEL, Brasil) e o referente aos Artigos Curtos foi coordenado pelo Prof. Eduardo Ogasawara (CEFET/RJ, Brasil) que organizaram este volume. Os artigos completos que integram essa série de anais foram submetidos até a data 29/05/2023, aceitos para publicação em 06/07/2023 e 19/07/2023, tendo a versão final submetida em 07/08/2023. Os artigos curtos, por sua vez, foram submetidos em 26/06/2023, aceitos para publicação em 31/07/2023, e tiveram a versão final submetida em 15/08/2023.

A qualidade dessa edição é devida essencialmente aos autores e revisores dos artigos publicados. Expressamos nossos fortes agradecimentos pelas contribuições e discussões durante o SBBB 2023.

Para mais informações sobre o SBBB 2023, visite o site desta edição do evento em <https://sbbd.org.br/2023>.

Execução:**Realização:****Apoio Acadêmico:****Patrocínio:****Agências de fomento:**

The Proceedings of the 38th Brazilian Symposium on Databases (SBBDB 2023) include the papers selected and presented in the edition of the event from September 25th to September 29th, 2023. The proceedings contain 25 full papers and 24 short papers in this issue. These papers were selected through a double-anonymized review process, with a total of 51 full papers and 55 short papers submitted, which implied acceptance rates of 49% (25/51) and 42.5% (23/54), respectively. One paper originally submitted to the full paper track was accepted as short. Both full and short papers include works from the categories of research, industry or applications, and vision, in diverse areas of databases, data science, and big data. The Full Papers program committee was coordinated by Daniel Kaster (UEL, Brazil), and the Short Papers program committee was coordinated by Eduardo Ogasawara (CEFET/RJ, Brazil), who organized this volume. The full papers in this volume were submitted by May 29th, accepted for publication on July 6th and July 19th, with the final version submitted on August 8th. The short, vision, and industrial papers in this volume, in turn, were submitted on June 26th and accepted for publication on July 31st, with the final version submitted on August 15th.

The richness of this issue can be mainly credited to the authors and reviewers. We thank them a lot for their insightful contributions and discussions.

You can find more information about SBBDB 2023 by visiting the website of the event at <https://sbbdb.org.br/2023>.

Execução:**Realização:****Apoio Acadêmico:****Patrocínio:****Agências de fomento:**

FULL

A New Data Modeling Approach for Alignment-free Biological Applications	sbbd:01 1
<i>Diogo Munaro Vieira, Elvismary Molina de Armas, Maria Leandra Guateque Jaramillo, Marcos Catanho, Antonio Basílio de Miranda, Edward Hermann Haeusler, Sérgio Lifschitz</i>	
Database Support for Online Social Network Analysis	sbbd:02 14
<i>Paulo Sérgio Pintor, Mariana Salgueiro, Sérgio Lifschitz, Daniel Schwabe, José Manuel Moreira, Rogério Luís Costa</i>	
Análise de Desempenho em Banco de Dados NoSQL Orientado a Documentos: Um Índice para Comparação de Modelos de Dados	sbbd:03 26
<i>Harley Vera-Oliveira, Edwin Alvarez-Mamani, Maristela Holanda</i>	
Análise e Classificação de Gêneros Musicais com Base em Letras de Músicas . . .	sbbd:04 39
<i>Barbara Poliana Jaber Martins Ferreira, Daniel Dalip Hasan, Ismael Santana Silva</i>	
Approximate Similarity Joins over Dense Vector Embeddings	sbbd:05 51
<i>Douglas Rolins Santana, Leonardo Andrade Ribeiro</i>	
Avaliando Fatores de Influência sobre Algoritmos de Aprendizado de Máquina na Etapa de Classificação da Resolução de Entidades	sbbd:06 63
<i>Milena Santos Macedo, Dimas Cassimiro Nascimento</i>	
Big Data Architectures for FAIR-compliant Repositories: A Systematic Review	sbbd:07 76
<i>João Pedro de Carvalho Castro, Cristina Dutra Aguiar</i>	
Calibração de Distância em Métodos de Acesso Métrico por meio de Realimentação de Relevância	sbbd:08 89
<i>Renato Gomes Marcacini, Willian Dener de Oliveira, Agma Juci Machado Traina</i>	
Impacto do Pré-processamento e Representação Textual na Classificação de Documentos de Licitações	sbbd:09 102
<i>Michele A. Brandão, Mariana Silva, Gabriel P. Oliveira, Henrique Hott, Anisio Lacerda, Gisele L. Pappa</i>	
Improving Interoperability between Relational and Blockchain-based Database Systems: A Middleware Approach	sbbd:10 115
<i>Rafael Avilar Sá, Leonardo Moreira, Javam de Castro Machado</i>	
Modelagem de Tópicos para a Tarefa de Recuperação de Casos Legais	sbbd:11 128
<i>Altigran Soares da Silva, Luisa Pereira Novaes, Daniela Vianna</i>	
Predição de Tempo Restante para Conclusão de Processos de Negócio Utilizando Aprendizado Profundo	sbbd:12 141
<i>Ronildo Oliveira da Silva, Regis Magalhães, Lívia Almada Cruz, Criston P. de Souza, Davi Romero, José Macedo</i>	
Reptilerecon: Um Arcabouço para Extração e Análise de Sinais de Lagartos . .	sbbd:13 154
<i>João Gabriel Fernandes Zenóbio, Pedro Henrique Lopes Silva, Eduardo José Da Silva Luz, Gladston Juliano Prates Moreira, Conrado Aleksander B. Galdino, Jadson Castro Gertrudes</i>	
Spending pattern visualization using unsupervised machine learning	sbbd:14 167
<i>Gabriel Porto Oliveira, Roberta Barbosa Oliveira, Jadson Castro Gertrudes</i>	
SQL2Neo: Uma Interface de Acesso SQL para o Neo4j	sbbd:15 179

<i>Pablo Vicente Passos, Lucas Santos de Oliveira, Geomar André Schreiner, Vanessa Lago Machado, Ronaldo dos Santos Mello</i>	
Aplicação de Técnicas de Confident Learning para Limpeza de Dados e Melhoria de Desempenho de Classificadores de Aprendizado de Máquina: um Estudo de Caso	sbbd:16 192
<i>Renato Okabayashi Miyaji, Felipe Valencia Almeida, Pedro Luiz Pizzigatti Corrêa</i>	sbbd:17
Database Access Control in a Database Keyword Search Tool	205
<i>Grettel Monteagudo Garcia, Javier Guillot Jiménez, Yenier Torres Izquierdo, Ângelo Batista Neves Júnior, Melissa Lemos, Marco Antonio Casanova, Ana Cristina Ferreira, Flavia Pacheco Texeira da Silva</i>	
Epiflow: a Hybrid Approach to Track Infectious Disease Spread in Brazil Based on Travel Data and Graph Databases	sbbd:18 218
<i>Mariama Oliveira, Andreza Alencar, Natalia Oliveira, Lucas Sales, Antônio Ricardo Khouri Cunha, Pablo Ramos</i>	
Fault Detection in Transmission Lines: a Denial Constraint Approach	sbbd:19 231
<i>Nicolas Tamalu, Leandro Augusto Ensina, Eduardo Cunha de Almeida, Eduardo Henrique Monteiro Pena, Luiz Eduardo Soares Oliveira</i>	
From Exploration to Exploitation: Understanding the Evolution of Music Careers through a Data-driven Approach	sbbd:20 244
<i>Danilo B Seufitelli, Mirella M Moro</i>	
MAESTRO: Uma Abordagem para a Composição e Análise de Workflows Baseados em Scripts por Meio de Ontologias	sbbd:21 256
<i>Luiz Gustavo Dias, Daniel de Oliveira, Bruno Lopes</i>	
PromptNER: Uma Abordagem para Reconhecimento de Entidades Nomeadas em Dados Sensíveis a Partir de Instâncias Rotuladas Automaticamente	sbbd:22 269
<i>Claudio Moisés Valiense Andrade, Celso França, Fabiano Muniz Belém, Gabriel Jalais, Marcelo Ganem, Gabriel Texeira, Alberto Laender, Marcos André Gonçalves</i>	
Replicação Descentralizada em Bancos de Dados Distribuídos Usando o Algoritmo Paxos	sbbd:23 282
<i>Danilo José Pereira Ferreira, Sahudy Montenegro González, Gustavo Maciel Dias Vieira</i>	
Um Estudo Sobre Métricas de Avaliação para Sumarização de Acórdãos	sbbd:24 295
<i>Gustavo Rufino Feltrin, Daniela Vianna, Altigran Soares da Silva</i>	
Uso de Regressores para a Geração de Conhecimento na Avaliação do Ensaio de Tração Ambiente	sbbd:25 306
<i>Keila de Carvalho Freitas, Kairo de Barros Guimarães, Luxmar Alves Xavier, Thiago Augusto de Oliveira Silva</i>	

SHORT

Subset Modelling: A Domain Partitioning Strategy for Data-efficient Machine-Learning sbbd:26 318	
<i>Vitor Ribeiro, Raphael de Freitas Saldanha, Eduardo Henrique Monteiro Pena, Reza Akbarinia, Julia Stoyanovich, Falaah Arif Khan, Patrick Valduriez, Fabio Porto</i>	
A Novel Graph-based Diversity-aware Rank Fusion Method Applied to Image Metasearch	sbbd:27 324
<i>José Solenir Lima Figuerêdo, Ana Lúcia Lima Marreiros Maia, Rodrigo Tripodi Calumby</i>	
Análise do Discurso de Ódio em Comentários de Vídeos no YouTube: Um Estudo de Caso da CPI da COVID-19 no Brasil	sbbd:28 330
<i>Marcela Cupertino Teixeira, Julio C. S. Reis</i>	
Ativismo da extrema direita brasileira no WhatsApp: O que mudou das eleições de 2018 para 2022?	sbbd:29 336
<i>José Maria da Silva Monteiro Filho, Javam de Castro Machado, Ivandro de Sá, Thiago Gadelha, Rangel dos Santos Jungles, Luciano Galick, Carlos Eduardo Santos</i>	
Avaliação Experimental de Detectores de Erros em Conjuntos de Dados Relacionais sbbd:30 342	
<i>William G. R. Medina, Eduardo Henrique Monteiro Pena, Daniel S. Kaster</i>	
Avaliando o Processo de Seleção de Características na Tarefa de Junção de Similaridade	sbbd:31 348
<i>Lucas Romeiro Silva, Dimas Cassimiro Nascimento</i>	
Construção de Banco de Dados do Mercado Imobiliário. Um estudo na Cidade de São Paulo	sbbd:32 354
<i>Celso Gabriel Azevedo Ribeiro, Flavio Falcao Helena, Flavio Almeida Magalhaes Cipparrone</i>	
Context-Aware Knowledge Graphs Exploratory Search	sbbd:33 360
<i>Veronica dos Santos, Sergio Lifschitz, Daniel Schwabe, Edward Hermann Haeusler</i>	
Linking Heterogeneous Health Data Sources in Brazil Centered on Drug Leaflet Processing	sbbd:34 366
<i>Márcia Jacobina Andrade Martins, Claudia Bauzer Medeiros</i>	
Métodos de Detecção de Fake News: Uma Comparação entre as Abordagens de Crowd Signals e Ensembles	sbbd:35 372
<i>Uriel Merola, Paulo Freire, Ronaldo Ribeiro Goldschmidt, Jorge de Abreu Soares</i>	
Mineração de Dados sobre Despesas Públicas de Municípios Mineiros para Gerar Alertas de Fraudes	sbbd:36 378
<i>Larissa D. Gomide, Guilherme Bezerra dos Santos, Lucas L. Costa, Michele A. Brandão, Anisio Lacerda, Gisele L. Pappa</i>	
Planos físicos de consulta para Divisão Relacional	sbbd:37 384
<i>Letícia de Campos Lima, Lucio F.D. Santos, Marcos Bedo, Enzo Seraphim, Luiz Olmes Carvalho</i>	
Projeção Eficiente de Violações de Restrições de Negação	sbbd:38 390
<i>Leonardo Ferreira Luciano, Wendel Caio Moro, Eduardo Cunha de Almeida, Edu-</i>	

<i>ardo Henrique Monteiro Pena</i>	
Quantile Symbolic Aggregate approXimation: A guaranteed equiprobable SAX	sbbd:39 396
<i>Eduardo de Medeiros Silveira, Joaquim Vinicius Carvalho Assunção, Leonardo Ramos Emmendorfer</i>	
Similarity Grouping by Influence: Exploring Result Diversification in Similarity Group-by Operators	sbbd:40 402
<i>Willian Dener de Oliveira, Caetano Traina Jr., Lucio F.D. Santos, Anna Júlia Costa Lauton</i>	
Uma Metodologia para Tratamento do Viés da Maioria em Modelos de Stacking via Identificação de Documentos Difíceis	sbbd:41 408
<i>Welton Augusto Santos, Washington Cunha, Celso França, Guilherme Fonseca Nascimento, Sergio Canuto, Leonardo Rocha, Marcos Gonçalves</i>	
Vetor de sufixos métrico compacto	sbbd:42 414
<i>Frederico Rezende Rosa, Felipe Alves da Louza, Humberto Razente</i>	
Impacto de Doações Eleitorais no Faturamento de Empresas: Um Estudo nas Eleições Municipais em Minas Gerais	sbbd:43 420
<i>Bárbara M. A. Mendes, Camila Braz, Lucas L. Costa, Gabriel P. Oliveira, Henrique Hott, Mariana Silva, Gisele L. Pappa</i>	
DRE-CVM: Explorando Dados Financeiros Enriquecidos com Proveniência e Princípios FAIR de Companhias Brasileiras de Capital Aberto	sbbd:44 426
<i>Saulo Almeida, Gilberto Gil F. Gomes Passos, Valquire da Silva de Jesus, Jorge Juan Zavaleta, Sérgio Manuel Serra da Cruz1</i>	
Ensino de Engenharia de Dados nas Universidades Brasileiras: Estado Atual e Perspectivas de Mercado	sbbd:45 432
<i>Tarsis Figueredo Azevedo, Altigran Soares da Silva</i>	
Identificação de Falhas em Turbinas Eólicas Utilizando Abordagens de Aprendizado de Máquina	sbbd:46 439
<i>Danielle Pinna Rodrigues, Diego Nunes Brandão, Rodrigo Toso Franco, Kele Belloze, Raphael Guerra Pereira de Oliveira, Fernando de Sá Pereira Gonçalves</i>	
Integrador de metadados modular para aprendizado de máquina com visualização em tempo de execução	sbbd:47 445
<i>Filipe Augusto da Silva, Marta Mattoso</i>	
Um caso de Democratização de Dados na Indústria de Óleo e Gás	sbbd:48 451
<i>Kellyne Marques Santos, Dean Melo, Nathalia Marassi Cianni</i>	
Uso de Grafos de Proveniência para Análise Temporal de Uso do Solo em Centros Urbanos: uma Abordagem Prática	sbbd:49 457
<i>Raffael Paranhos, Marcos Lage, Daniel de Oliveira</i>	