R-Shiny as an interface for Data Visualization and Data Analysis on the Brazilian Digital Library of Thesis and Dissertations (BDTD)

Open Repositories 2019
Hamburg, Germany

Lucca F. Ramalho, Leonard R. Campelo

Washington Segundo
BDTD – Brazilian Digital Library of Theses and Dissertations

- Created in 2002
- Supported by IBICT – Brazilian Institute of Information in Science and Technology
- +540K full-text documents
- 114 Brazilian institutions
BDTD – Brazilian Digital Library of Theses and Dissertations

- Is is part of the oasisbr – Brazilian Portal of Open Access Publications

- An window to:
  - NDLTD (Network Digital Library of TDs)
  - LA Referencia
  - OpenAIRE
BDTD – Brazilian Digital Library of Theses and Dissertations

ACESSO E VISIBILIDADE ÀS TESES E DISSERTAÇÕES BRASILEIRAS

114 Instituições
397.082 Dissertações
143.573 Teses
540.653 Documentos

Sobre a BDTD

O Ibict desenvolveu e coordena a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), que integra os sistemas de informação de teses e dissertações existentes nas instituições de ensino e pesquisa do Brasil, e também estimula o registro e a publicação de teses e dissertações em meio eletrônico. A BDTD, em parceria com as instituições brasileiras de ensino e pesquisa, possibilita que a comunidade brasileira de C&T...
BDTD – Brazilian Digital Library of Theses and Dissertations
BDTD – Brazilian Digital Library of Theses and Dissertations

- Meta-searcher
  - Built over the LA Referencia software
    - VuFind (Solr search engine)
    - LRHarvester
    - OAI-PMH

- Local repositories are using different platforms:
  - DSpace 4, 5 and 6
  - A minor amount is using locally developed platform
BDTD – Brazilian Digital Library of Theses and Dissertations

- Metadata schema
  - Qualified Dublin Core (with modifications)
  - Examples:
    - dc.contributor.authorID
    - dc.creator.authorLattes
    - dc.contributor.advisor1ID
    - dc.contributor.advisor1Lattes
Lattes Research Profile Platform

- +6mi records
- Created in 1999
- Supported by the National Council of Scientific and Technological Development (CNPq)
- 99.9% of the researchers in Brazil have a profile in this platform
- Lattes ID → ORCID
Lattes Research Profile Platform

<CURRICULO-VITAE SISTEMA-ORIGEM-XML="LATTES_OFFLINE" NUMERO-IDENTIFICADOR="9453481318889500" DATA-ATUALIZACAO="29052019" HORA-ATUALIZACAO="114525">

<DADOS-GERAIS NOME-COMPLETO="Washington Luís Ribeiro de Carvalho Segundo" NOME-EM-CITACOES-BIBLIOGRAFICAS="CARVALHO-SEGUNDO, W.; SEGUNDO, W.; RIBEIRO, W.; CARVALHO-SEGUNDO, WASHINGTON; DE CARVALHO-SEGUNDO, WASHINGTON" NACIONALIDADE="B" PAIS-DE-NASCIMENTO="Brasil" UF-NASCIMENTO="SP" CIDADE-NASCIMENTO="Piracicaba" PERMISSAO-DE-DIVULGACAO="NAO" DATA-FALECIMENTO="" SIGLA-PAIS-NACIONALIDADE="BRA" PAIS-DES-NACIONALIDADE="Brasil">

<RESUMO-CV TEXTO-RESUMO-CV-RH="É Doutor e Mestre em Informática pela Universidade de Brasília (UnB). Possui graduação em Matemática (Bacharelado e Licenciatura) pela mesma universidade. É Coordenador do Laboratório de Metodologias de Tratamento e Disseminação da Informação (COLAB) no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT / MCTIC). Áreas de interesse em pesquisa: Teoria da Computação, Lógica, Sistemas Algébricos, Verificação Formal, Assistentes de Prova, Lógica Nominal, Unificação, Repositórios Digitais e Sistemas Agregadores de Acesso Aberto." TEXTO-RESUMO-CV-RH-EN=""/>

</DADOS-GERAIS>
</CURRICULO-VITAE>
Formação acadêmica/titulação

**Doutorado em Informática (Conceito CAPES 5).**
Universidade de Brasília, UnB, Brasil.
com período sanduíche em King’s College London (Orientador: Maribel Fernández).
Título: Nominal Equational Problems Modulo Associativity, Commutativity and Associativity-Commutativity,
Ano de obtenção: 2019.

Orientador: Mauricio Ayala Rincón.
Coorientador: Maribel Fernández.
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.
Palavras-chave: Métodos Formais; Reescrita Nominal; Formal Methods; Nominal Rewriting.
Grandes áreas: Ciências Exatas e da Terra
Grande Área: Ciências Exatas e da Terra / Área: Matemática / Subárea: Álgebra / Especialidade: Lógica Matemática.

**Mestrado em Informática (Conceito CAPES 5).**
Universidade de Brasília, UnB, Brasil.
Título: Verificação de Propriedades do Cálculo Lambda_ex em Coq, Ano de Obtenção: 2011.

Orientador: Flávio Leonardo Cavalcanti de Moura.
Palavras-chave: verificação formal; cálculos de substituições explícitas; cálculo lambada ex.
Grandes áreas: Ciências Exatas e da Terra
Grande Área: Ciências Exatas e da Terra / Área: Ciência da Computação / Subárea: Computação.
Grande Área: Ciências Exatas e da Terra / Área: Matemática / Subárea: Verificação Formal.
R interface

- Language and environment for statistical computing and graphics
- Open Source
- Strong Community
- Supports multiple data types
- Works with Packages (libraries)
Packages (Libraries)

... and many more!

dplyr, forcats, ggplot2, haven, purrr,
readr, readxl, stringr, tibble, tidyr,
tidyverse

jsonlite, leaflet, networkD3
R-Shiny

- Easy to build web apps straight from R
- Let users interact with your data and your analysis
- Can be extended with CSS themes, htmlwidgets, and JavaScript actions.
How it works
Figure 1: Publications per year by type

Figure 2: Publications per year by geographic group
BDTD APP

Figure 3: Cartogram

Figure 4: Data export interface
Figure 5: Heatmap of number of documents

Figure 6: Simple network analysis on authors
specialists dashboard

A busca pelo termo: 'machine learning' resultou em:

- 1.022 documentos
- 878 autores
- 49 instituições

Tipo de documento:
- Dissertação de Mestrado: 411
- Artigo: 397
- Tese de Doutorado: 214

Publicações por ano:

Assuntos Relacionados:
- Biometrics
- Biometria
- Algoritmos
- Data Streams
- Leitura

Bioinformática
TOP 10 ORIENTADORES ESPECIALISTAS

ANDRÉ CARLOS PONCE DE LEON FERREIRA DE CARVALHO

- orientações de Doutorado: 10
- orientações de Mestrado: 9
- Artigos: 18

Assuntos relacionados: Aprendizado De Máquina, Bioinformática, Meta-Aprendizado, Algoritmos Genéticos, Classificação e Classificação Hierárquica.

GUILHERME DE ALENÇAR BARRETO

- orientações de Doutorado: 9
- orientações de Mestrado: 4
- Artigos: 0

Assuntos relacionados: Teleinformática, Redes Neurais, Curto Circuito, Inteligência Computacional, Aprendizado De Máquina e Aprendizagem Profunda.

MARIA CAROLINA MONARD

- orientações de Doutorado: 7
- orientações de Mestrado: 6
- Artigos: 0

Assuntos relacionados: Aprendizado De Máquina, Combinação De Classificadores, Computação Evolutiva, Descoberta De Conhecimento, Mineração De Dados e Programação Lógica Indutiva.

JOÃO PAULO PAPA

- orientações de Doutorado: 2
- orientações de Mestrado: 6
- Artigos: 3

Assuntos relacionados: Floresta De Caminhos Ótimos, Aprendizado De Máquina, Processamento De Imagens - Tecnicas Digitais, Reconhecimento D Padrões, Computação - Matematica e Computer Science Mathematics.
networkD3 package
Final considerations

- BDTD and Lattes Platform have considerable amount of data to build interesting analyzes

- R and R-Shiny were showed be scalable when operating over searching engines such as apache-Solr

- The huge community of R provides a plenty of well-established packages that facilitates the creation of R-Shiny interfaces (such as the ropensci, which provides open tools for open science)
Thank you!
Danke schön!

luccaramalho@ibict.br
leonardcampelo@ibict.br
washingtonsegundo@ibict.br

https://gitlab.com/luccaramalho/