

Introdução ao KNIME

Palestras do DCC

Geraldo Xexéo^{1,2}

¹Departamento de Ciências da Computação

²Programa de Engenharia de Sistemas e Computação

Palestra Remota para o DCC, Junho 2020

Agenda

- 1 O que é o KNIME
- 2 Workflows
- 3 Exemplos
- 4 BIRT – Basic Intelligence Report Tool
- 5 Conclusão

Onde Estamos?

1 O que é o KNIME

2 Workflows

3 Exemplos

4 BIRT – Basic Intelligence Report Tool

5 Conclusão

Apresentação do KNIME

- Pronúncia “NAIME”
- 2 Ferramentas de *Data Science*
 - KNIME Analytics Platform
 - Open Source
 - Desenvolvimentos de sistemas de *Data Science*
 - KNIME Server
 - Comercial
 - Colaboração, gerência, automação e *deployment* de *workflows* de *Data Science* como aplicações e serviços

KNIME Analytics Platform

- Baseada no Eclipse
- Implementa vários algoritmos
- Programável e extensível
 - Java, Python, R
- Gera relatórios e gráficos
- Possui plug-ins de terceiros

O que é o KNIME
Workflows
Exemplos
BIRT – Basic Intelligence Report Tool
Conclusão

Apresentação do KNIME
KNIME Analytics Platform
A interface do KNIME
O que é um workflow

A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow titled "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the nodes, there are labels: "Ler arquivo MS-Covid" under the Excel Reader, "Filtrar apenas dados do Brasil" under the Row Filter, and "Salva em arquivo CSV" under the CSV Writer.

The left sidebar contains the "KNIME Explorer" pane, showing a tree view of the project structure. The "Favorite Nodes" pane is also visible, showing a list of favorite nodes. The "Node Repository" pane is at the bottom left, showing a list of nodes categorized by type (IO, Manipulation, Views, Analytics, DB, Other Data Types, Structured Data, Scripting, Tools & Services, Community Nodes, KNIME Labs, Workflow Control, Workflow Abstraction, Social Media).

The right sidebar contains the "Description" pane, which provides metadata for the selected workflow. The metadata includes:

- Title:** Palestra-Exemplo-Workflow-1
- Description:** No description has been set yet.
- Tags:** No tags have been added yet.
- Links:** No links have been added yet.
- Creation Date:** 2020-5-28
- Author:** Geraldo

The bottom of the interface features an "Outline" pane on the left, showing a hierarchical view of the workflow nodes, and a "Console" pane on the right, displaying the KNIME log output. The log output includes the following text:

```
KNIME Console
*** welcome to KNIME Analytics Platform v4.1.3.v202005121100 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: F:\Users\Gerald\knime-workspace\.metadata\knime.log
WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\Data\HIST_PAINEI_COVID08_27mai2020.xls"
WARN Excel Reader (XLS) 0:4 failed to apply settings: Loading input file...
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\Data\Resultado-1.csv" exists and will be overw
```

O que é o KNIME
Workflows
Exemplos
BIRT – Basic Intelligence Report Tool
Conclusão

Apresentação do KNIME
KNIME Analytics Platform
A interface do KNIME
O que é um workflow

A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow titled "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the nodes, there are labels: "Lar arquivo MS-Covid", "Papa apenas dados do Brasil", and "Salva em arquivo CSV". A large blue box with the text "Área do Workflow" is overlaid on the workflow area.

The interface includes several panels:

- KNIME Explorer:** Shows the project structure, including "My-KNIME-Hub", "EXAMPLES", "LOCAL (Local Workspace)", and "Palestra Knime".
- Favorite Nodes:** Lists "Personal favorite nodes" and "Most frequently used nodes".
- Node Repository:** Provides a search bar and a list of node categories such as "IO", "Manipulation", "Views", "Analytics", "DB", "Other Data Types", "Structured Data", "Scripting", "Tools & Services", "Community Nodes", "KNIME Labs", "Workflow Control", "Workflow Abstraction", and "Social Media".
- Description:** Displays metadata for the selected workflow, including Title, Description, Tags, Links, Creation Date, and Author.
- Console:** Shows the KNIME Console output, including a welcome message and log file information.

The console output includes the following text:

```
KNIME Console
*** welcome to KNIME Analytics Platform v4.1.3.v202005121100 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: F:\Users\Gerald\knime-workspace\metadata\knime.log
WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\Data\HIST_PAINEI_COVID08_27mai2020.xls"
WARN Excel Reader (XLS) 0:4 failed to apply settings: Loading input file...
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\Data\Resultado-1.csv" exists and will be over
```

O que é o KNIME
Workflows
Exemplos
BIRT – Basic Intelligence Report Tool
Conclusão

Apresentação do KNIME
KNIME Analytics Platform
A interface do KNIME
O que é um workflow

A interface do KNIME

Explorador

The screenshot displays the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' pane shows a project tree with 'Palestra-Exemplo-Workflow-1' selected. A blue box with the word 'Explorador' is overlaid on this pane. The main workspace shows a workflow with three nodes: 'Excel Reader (XLS)', 'Row Filter', and 'CSV Writer'. Below the nodes, descriptive text reads: 'Ler arquivo MS-Covid', 'Fazer apenas dados do Brasil', and 'Salva em arquivo CSV'. On the right, the 'Description' pane shows details for 'Palestra-Exemplo-Workflow-1', including title, description, tags, links, creation date (2020-5-28), and author (Geraldo). At the bottom, the 'Console' pane displays the KNIME log, which includes a welcome message and warnings about file paths and output file existence.

KNIME Analytics Platform

File Edit View Node Help

KNIME Explorer

Palestra-Exemplo-Workflow-1

Example Workflow

KNIME_project

KNIME_project2

Simple Regression Tree

Favorite Nodes

Personal favorite nodes

Most frequently used nodes

Node Repository

IO

Manipulation

Views

Analytics

DB

Other Data Types

Structured Data

Tools & Services

Community Nodes

KNIME Labs

Workflow Control

Workflow Abstraction

Social Media

Excel Reader (XLS)

Row Filter

CSV Writer

Ler arquivo MS-Covid

Fazer apenas dados do Brasil

Salva em arquivo CSV

Description

Palestra-Exemplo-Workflow-1

Title Palestra-Exemplo-Workflow-1

Description No description has been set yet.

Tags No tags have been added yet.

Links No links have been added yet.

Creation Date 2020-5-28

Author Geraldo

Outline

Console

KNIME Console

*** welcome to KNIME Analytics Platform v4.1.3.v202005121100 ***

*** Copyright by KNIME AG, Zurich, Switzerland ***

Log file is located at: F:\Users\Geraldo\knime-workspace\.metadata\knime.log

WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\data\HIST_PAINEI_COVID08_27mai2020.xls" failed to apply settings: Loading input file...

WARN Excel Reader (XLS) 0:4 Directory "F:\Github\Palestra KNIME\data" of output file does not exist

WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\data" of output file does not exist

WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\data" of output file does not exist

WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\data\Resultado-1.csv" exists and will be overw...

O que é o KNIME
Workflows
Exemplos
BIRT – Basic Intelligence Report Tool
Conclusão

Apresentação do KNIME
KNIME Analytics Platform
A interface do KNIME
O que é um workflow

A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' pane shows a project tree with 'Palestra-Exemplo-Workflow-1' selected. Below it, the 'Favorite Nodes' pane lists 'Personal favorite nodes' and 'Most frequently used nodes'. The main workspace shows a workflow with three nodes: 'Excel Reader (XLS)', 'Row Filter', and 'CSV Writer'. The 'Excel Reader (XLS)' node is connected to the 'Row Filter' node, which is connected to the 'CSV Writer' node. The 'Excel Reader (XLS)' node has a description 'Lar arquivo MS-Covid'. The 'Row Filter' node has a description 'Papa apenas dados do Brasil'. The 'CSV Writer' node has a description 'Salva em arquivo CSV'. On the right, the 'Description' pane shows the workflow's title 'Palestra-Exemplo-Workflow-1', description, tags, links, creation date (2020-5-28), and author (Geraldo). At the bottom, the 'Console' pane shows the KNIME Console output, including a welcome message and log file information.

Nós disponíveis

- Analytics
- DB
- Other Data Types
- Structured Data
- Tools & Services
- Community Nodes
- KNIME Labs
- Workflow Control
- Workflow Abstraction
- Social Media

```
KNIME Console
*** welcome to KNIME Analytics Platform v4.1.3.v202005121100 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: F:\Users\Geraldo\knime-workspace\.metadata\knime.log
WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\data\HIST_PAINEI_COVID08_27mai2020.xls"
WARN Excel Reader (XLS) 0:4 failed to apply settings: Loading input file...
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\data" of output file does not exist
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\data" of output file does not exist
WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\data\Resultado-1.csv" exists and will be over
```

O que é o KNIME
Workflows
Exemplos
BIRT – Basic Intelligence Report Tool
Conclusão

Apresentação do KNIME
KNIME Analytics Platform
A interface do KNIME
O que é um workflow

A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow titled "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the nodes, there are labels: "Lar arquivo MS-Covid" under Excel Reader, "Papa apenas dados do Brasil" under Row Filter, and "Salva em arquivo CSV" under CSV Writer.

The left sidebar contains the "KNIME Explorer" pane, showing a tree view of the workspace. It includes folders for "My KNIME-Hub", "EXAMPLES", "LOCAL", and "Palestra Knime". The "Palestra Knime" folder is expanded, showing "Palestra-Exemplo-Workflow-1", "Example Workflow", "KNIME_project", "KNIME_project2", and "Simple Regression Tree".

The right sidebar contains the "Description" pane, which is currently empty. It has a title bar "Palestra-Exemplo-Workflow-1" and a description field. Below the description field, there are sections for "Tags", "Links", "Creation Date", and "Author".

The bottom of the interface features the "Outline" pane, which shows a small diagram of the workflow, and the "Console" pane, which displays the KNIME log. The log includes the following text:

```
KNIME Console
*** welcome to KNIME Analytics Platform v4.1.3.v202005121100 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: F:\Users\Gerald\knime-workspace\.metadata\knime.log
WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\Data\HIST_PAINEI_COVID08_27mai2020.xls"
WARN Excel Reader (XLS) 0:4 failed to apply settings: Loading input file...
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\Data\Resultado-1.csv" exists and will be over
```

O que é o KNIME
Workflows
Exemplos
BIRT – Basic Intelligence Report Tool
Conclusão

Apresentação do KNIME
KNIME Analytics Platform
A interface do KNIME
O que é um workflow

A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow titled "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the nodes, there are labels: "Ler arquivo MS-Covid", "Fazer apenas dados do Brasil", and "Salva em arquivo CSV".

On the left side, there is a "KNIME Explorer" pane showing a tree view of the project structure, including "My-KNIME-Hub", "EXAMPLES", "LOCAL", and "Palestra Knime". Below this is a "Favorite Nodes" pane and a "Node Repository" pane.

On the right side, there is a "Description" pane for the selected workflow, showing details like Title, Description, Tags, Links, Creation Date, and Author.

At the bottom, there is a "Console" pane displaying the KNIME log output, which includes the version number (v4.1.3.v202005121100) and the path to the log file.

A blue box with the word "Outline" is overlaid on the bottom left of the screenshot.

O que é o KNIME
Workflows
Exemplos
BIRT – Basic Intelligence Report Tool
Conclusão

Apresentação do KNIME
KNIME Analytics Platform
A interface do KNIME
O que é um workflow

A interface do KNIME

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow titled "Palestra-Exemplo-Workflow-1" with three nodes: "Excel Reader (XLS)", "Row Filter", and "CSV Writer". The workflow is connected by arrows, indicating a sequential process. Below the nodes, there are labels: "Lar arquivo MS-Covid" under Excel Reader, "Paga apenas dados do Brasil" under Row Filter, and "Salva em arquivo CSV" under CSV Writer.

The left sidebar contains the "KNIME Explorer" pane, showing a tree view of the project structure. The "Favorite Nodes" pane is also visible, showing a list of favorite nodes. The "Node Repository" pane is at the bottom left, showing a list of nodes categorized by type (IO, Manipulation, Views, Analytics, DB, Other Data Types, Structured Data, Scripting, Tools & Services, Community Nodes, KNIME Labs, Workflow Control, Workflow Abstraction, Social Media).

The right sidebar contains the "Description" pane, showing the title "Palestra-Exemplo-Workflow-1", description, tags, links, creation date (2020-5-28), and author (Geraldo).

The bottom of the interface features the "Outline" pane, which shows a small thumbnail of the workflow, and the "Console" pane, which displays the execution log. The console log shows the following output:

```
Log file is located at: F:\Users\Geraldo\knime-workspace\workspace\knime.log
WARN Excel Reader (XLS) 0:4 The file/directory "F:\Github\Palestra KNIME\Data\HIST_PAINEI_COVID08_27mai2020.xls"
WARN Excel Reader (XLS) 0:4 Failed to apply settings: Loading input file...
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Directory "F:\Github\Palestra KNIME\Data" of output file does not exist
WARN CSV Writer 0:2 Output file "F:\Github\Palestra-KNIME\Data\Resultado-1.csv" exists and will be over
```

O que é um workflow

Um workflow, ou fluxo de trabalho, é uma forma de descrever uma cadeia de processamento por meio de um grafo onde nós representam processos e arestas representam o fluxo de dados entre esses processos.

Excel Reader (XLS)



Ler arquivo MS-Covid

Row Filter



Pega apenas
dados do Brasil

CSV Writer



Salva em arquivo CSV

Onde Estamos?

1 O que é o KNIME

2 Workflows

3 Exemplos

4 BIRT – Basic Intelligence Report Tool

5 Conclusão

Entendendo um Workflow

- Um workflow é composto de nós
- Cada nó faz uma atividade
- Nós são divididos em tipos
- Nós tem portões de entrada e saída, de tipos diferentes
- Portões são conectados por fluxos, passando os dados e informações sobre eles

Excel Reader (XLS)



Ler arquivo MS-Covid

Row Filter



Pega apenas
dados do Brasil

CSV Writer



Salva em arquivo CSV

Um nó básico

- Tipo do nó
- Desenho indicativo do nó
- Porta de entrada
- Porta de saída
- Status do nó
- Nome do nó



Status do Nó

- Não configurado
- Pronto para executar
- Executado

Configuração de um Nó

Filter Criteria | Flow Variables | Memory Policy

☒ Include rows by attribute value
☐ Exclude rows by attribute value
☐ Include rows by number
☐ Exclude rows by number
☐ Include rows by row ID
☐ Exclude rows by row ID

Column value matching
Column to test:

☐ filter based on collection elements

Matching criteria
☒ use pattern matching

☐ case sensitive match ☐ contains wild cards
☐ regular expression

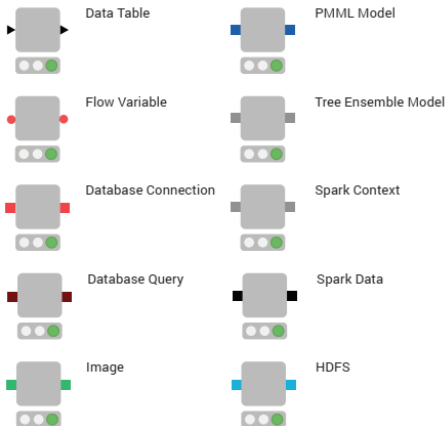
☐ use range checking
lower bound:
upper bound:

☐ only missing values match

Tipos de Nós

- Leitura
- Escrita / *Deploy*
- Transformação
- Análise, Aprendizado
- Exploração
- Predição
- Controle

Tipos de Porta (mais comuns)



Construindo um Workflow

- Coloque nós no mapa
- Configure os nós
- Execute
- Repita até alcançar o desejado

Onde Estamos?

1 O que é o KNIME

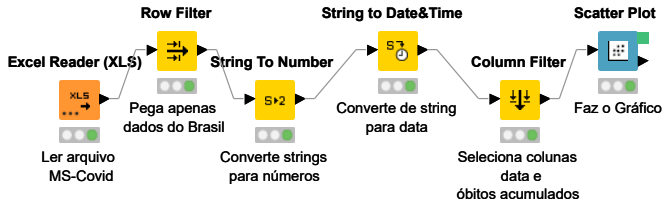
2 Workflows

3 Exemplos

4 BIRT – Basic Intelligence Report Tool

5 Conclusão

Visualização de Curvas



Metanodes

Excel Reader (XLS)



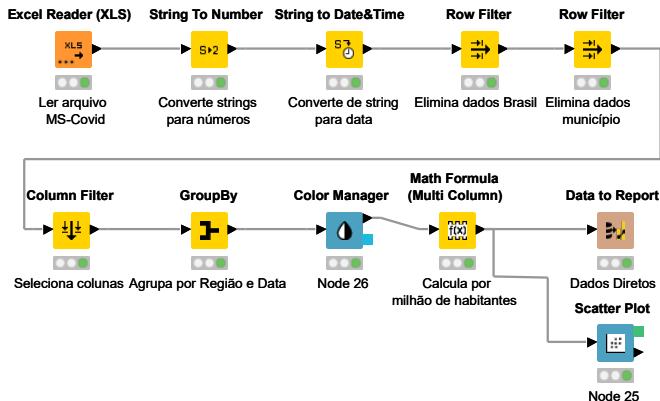
Seleciona Colunas



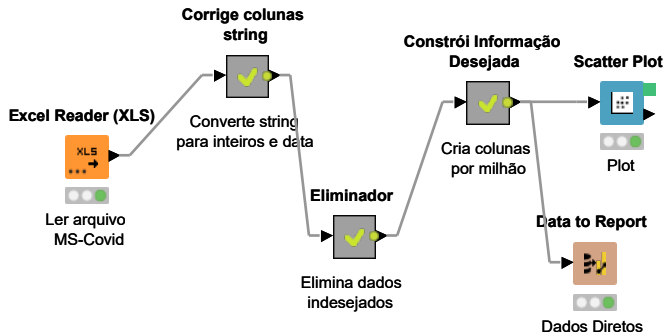
Scatter Plot



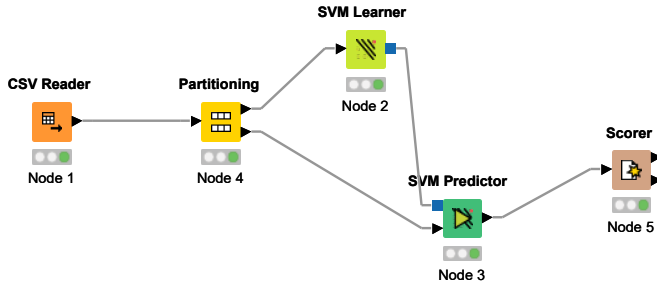
Estudando o Covid-19



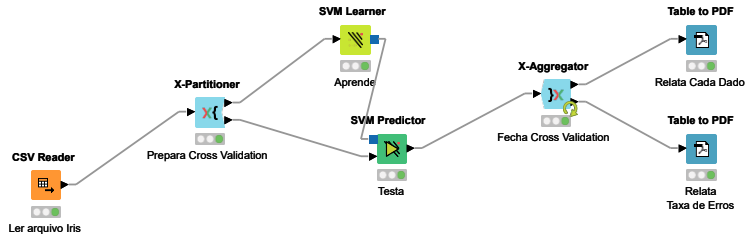
Simplificando o Modelo



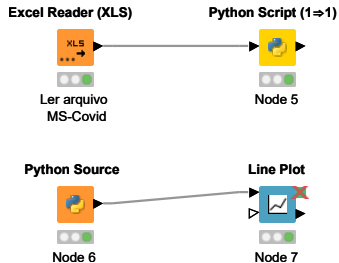
Aprendizado de Máquina



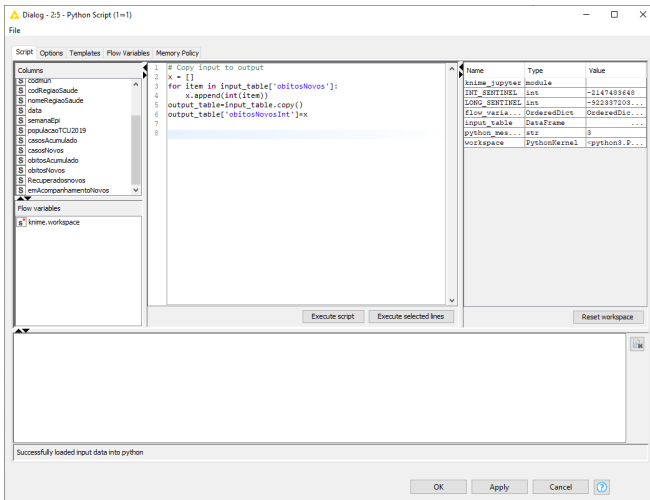
Cross Validation



Python Script



Janela do Python

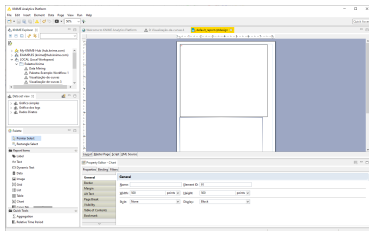


Onde Estamos?

- 1 O que é o KNIME
- 2 Workflows
- 3 Exemplos
- 4 BIRT – Basic Intelligence Report Tool
- 5 Conclusão

O que é BIRT

Software open source usado para criar relatórios, baseado no resultado dos workflows. Podem ser gerados em PDF, HTML e outros formatos.



Como usar BIRT

- instalar (instalar plug-in no)

- dentro de um workflow, clicar



Como usar BIRT II

- defina o layout da página
- defina a master page
- usos mais avançados com scripts...

Onde Estamos?

- 1 O que é o KNIME
- 2 Workflows
- 3 Exemplos
- 4 BIRT – Basic Intelligence Report Tool
- 5 Conclusão**

Conclusão

KNIME é um ambiente de desenvolvimento (e execução) de processos de Data Science de alta flexibilidade, podendo ser extendido por meio de plug-ins, scripts ou programas externos. Aviso: a documentação é de qualidade média. Mas existem muitos vídeos de treinamento.

É uma ótima ferramenta para Projetos Finais de Data Science.

Obrigado!

Contato



UNIVERSIDADE FEDERAL
DO RIO DE JANEIRO



INSTITUTO DE MATEMÁTICA
Universidade Federal do Rio de Janeiro



DEPARTAMENTO DE
CIÊNCIA DA
COMPUTAÇÃO
- UFRJ -



LINE

Laboratório de Tratamento da
Informação Não Estruturada



COPPE
UFRJ
Instituto Alberto Luis Coimbra de
Pós-Graduação e Pesquisa de Engenharia



PESCE
Programa de Engenharia
de Sistemas e Computação



LUDS
Ludologia, Engenharia e Simulação



Geraldo Xexéo

xexeo@cos.ufrj.br

gxexeo@gmail.com

Este obra está licenciado com uma Licença [Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 Internacional](#).