

Trabajo 2

Introducción a RapidMiner

Facultad De Ingeniería, Universidad De Cuenca
TEXT MINING
Freddy L. Abad L.
freddy.abadl@ucuenca.edu.ec

Opción 1

The screenshot shows the RapidMiner Studio interface in Design view. The main canvas displays a process flow with a 'Create Document' operator. The left sidebar contains the 'Repository' and 'Operators' panels. The 'Repository' panel shows a tree structure with 'Training Resources' (connected), 'Samples', 'Community Samples' (connected), 'DB', 'Local Repository' (Freddy), and 'Cloud Repository' (disconnected). The 'Operators' panel shows a search for 'create doc' with 'Create Document' selected under 'Text Processing'. The right sidebar contains the 'Parameters' and 'Help' panels. The 'Parameters' panel shows the 'Create Document' operator with an 'add label' checkbox. The 'Help' panel shows the 'Create Document' operator's synopsis and description.

The screenshot shows the RapidMiner Studio interface in Results view. The main canvas displays the output of the 'Create Document' operator. The left sidebar contains the 'Result History' and 'Repository' panels. The 'Result History' panel shows a list of documents with their metadata, including timestamps, file names, and license information. The 'Repository' panel shows the same tree structure as in the Design view.

Document
6/12 17:46:12.287 (2758) StarBurn Development Kit for Windows Version V15.4 'Patriot' (0x20150306)
Copyright (c) Rocket Division Software 2001-2016. All rights reserved.
Copyright (c) StarBurn Software 2009-2016. All rights reserved.
Started - Sun Jun 12 17:46:12 2016
6/12 17:46:12.288 (2758) License number: 7651, licensed to Desmond_Wondershare_Software_Co.,Ltd.,_SDK_PRO_15.4_StarBurnSDK_v15_License_Key on Fri Jun 12 13:21:47 2015 Please report all licensing violations to info@starburnsoftware.com software piracy fighting team.
10/17 15:35:29.973 (15cc) StarBurn Development Kit for Windows Version V15.4 'Patriot' (0x20150306)
Copyright (c) Rocket Division Software 2001-2016. All rights reserved.
Copyright (c) StarBurn Software 2009-2016. All rights reserved.
Started - Sun Jun 12 17:46:12 2016
6/12 17:46:12.288 (2758) License number: 7651, licensed to Desmond_Wondershare_Software_Co.,Ltd.,_SDK_PRO_15.4_StarBurnSDK_v15_License_Key on Fri Jun 12 13:21:47 2015 Please report all licensing violations to info@starburnsoftware.com software piracy fighting team.
10/17 15:35:29.973 (15cc) StarBurn Development Kit for Windows Version V15.4 'Patriot' (0x20150306)
Copyright (c) Rocket Division Software 2001-2016. All rights reserved.
Copyright (c) StarBurn Software 2009-2016. All rights reserved.

Local Repository/Practica 2 TM FREDDY ABAD FINAL* - RapidMiner Studio Trial 9.2.001 @ DESKTOP-PAKQTSI

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Repository

Import Data

Training Resources (connected)

Samples

Community Samples (connected)

DB

Local Repository (Freddy)

Cloud Repository (disconnected)

Operators

read doc

Extensions (2)

Text Processing (2)

Read Document

Read Documents (Mail)

No results were found.

Process

Process

100%

Create Document

Read Document

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

Parameters

Read Document

file ractica 2\bank.xml

extract text only

use file extension as type

encoding UTF-8

Hide advanced parameters

Help

Read Document

Text Processing

Tags: Text Processing

Input

file (File)

The file port.

Local Repository/Practica 2 TM FREDDY ABAD FINAL* - RapidMiner Studio Trial 9.2.001 @ DESKTOP-PAKQTSI

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Result History

Document (Create Document)

Document (Read Document)

Document

```
<bank>
<accounts>
<savings_accounts>
<savings_account id="a1" interest="0.03">
  <balance>2500</balance>
</savings_account>
<savings_account id="a2" interest="0.03">
  <balance>15075</balance>
</savings_account>
</savings_accounts>
<checking_accounts>
<checking_account id="a3">
  <balance>4025</balance>
</checking_account>
<checking_account id="a4">
  <balance>-125</balance>
</checking_account>
</checking_accounts>
</accounts>
</bank>
```

file	bank.xml
file_type	xml
path	C:\Users...
date	Sep 25, ...
size	1381

Repository

Import Data

Training Resources (connected)

Samples

Community Samples (connected)

DB

Local Repository (Freddy)

Cloud Repository (disconnected)

Local Repository/Practica 2 TM FREDDY ABAD FINAL* – RapidMiner Studio Trial 9.2.001 @ DESKTOP-PAKQTSI

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Result History Document (Create Document) Document (Read Document)

Document

2500

15075

4025

file	bank.xml
file_type	xml
path	C:\Users...
date	Sep 25, ...
size	1381

Repository

Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- DB
- Local Repository (Freddy)
- Cloud Repository (disconnected)

Opción 2



Views: Design Results Turbo Prep Auto Model

Find data, operators...etc. All Studio

Repository

Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- DB
- Local Repository (Freddy)
- Temporary Repository (Freddy)
- Cloud Repository (disconnected)

Operators

Search for Operators

- Services (1)
- HTML Processing (2)
- Utility (4)
- Get Page
- Get Pages
- Crawl Web
- Process Documents fr

[Get more operators from the Marketplace](#)

Process

Process

100%

Process design canvas showing a 'Crawl Web' operator connected to an 'ExampleSet' output.

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

Parameters

Crawl Web

url: rencia.gov.br/servidores

crawling rules: Edit List (0)...

max crawl depth: 1

☐ retrieve as html

☐ enable basic auth

☐ add content as attribute

[Hide advanced parameters](#)

Help

Crawl Web

Web Mining

Tags: Crawl, Crawler, Spider, Web Mining

Synopsis

This operator allows to crawl the web and store the retrieved links and pages in an ExampleSet or on disk.



Views: Design Results Turbo Prep Auto Model

Find data, operators...etc. All Studio

Result History

IOObjectCollection (Data to Documents)

ExampleSet (Crawl Web) Document (Create Document) Document (Read Document)

Open in: Turbo Prep Auto Model

Filter (0 / 0 examples): all

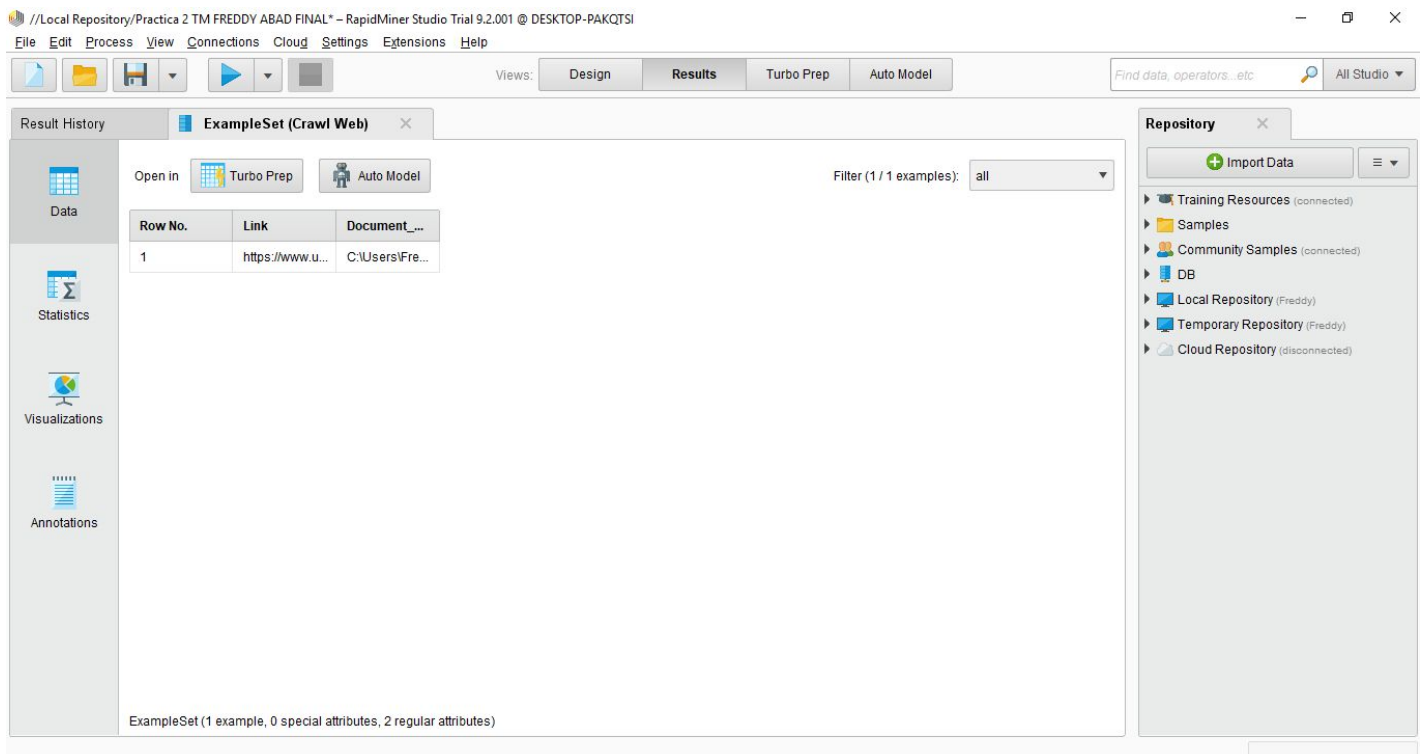
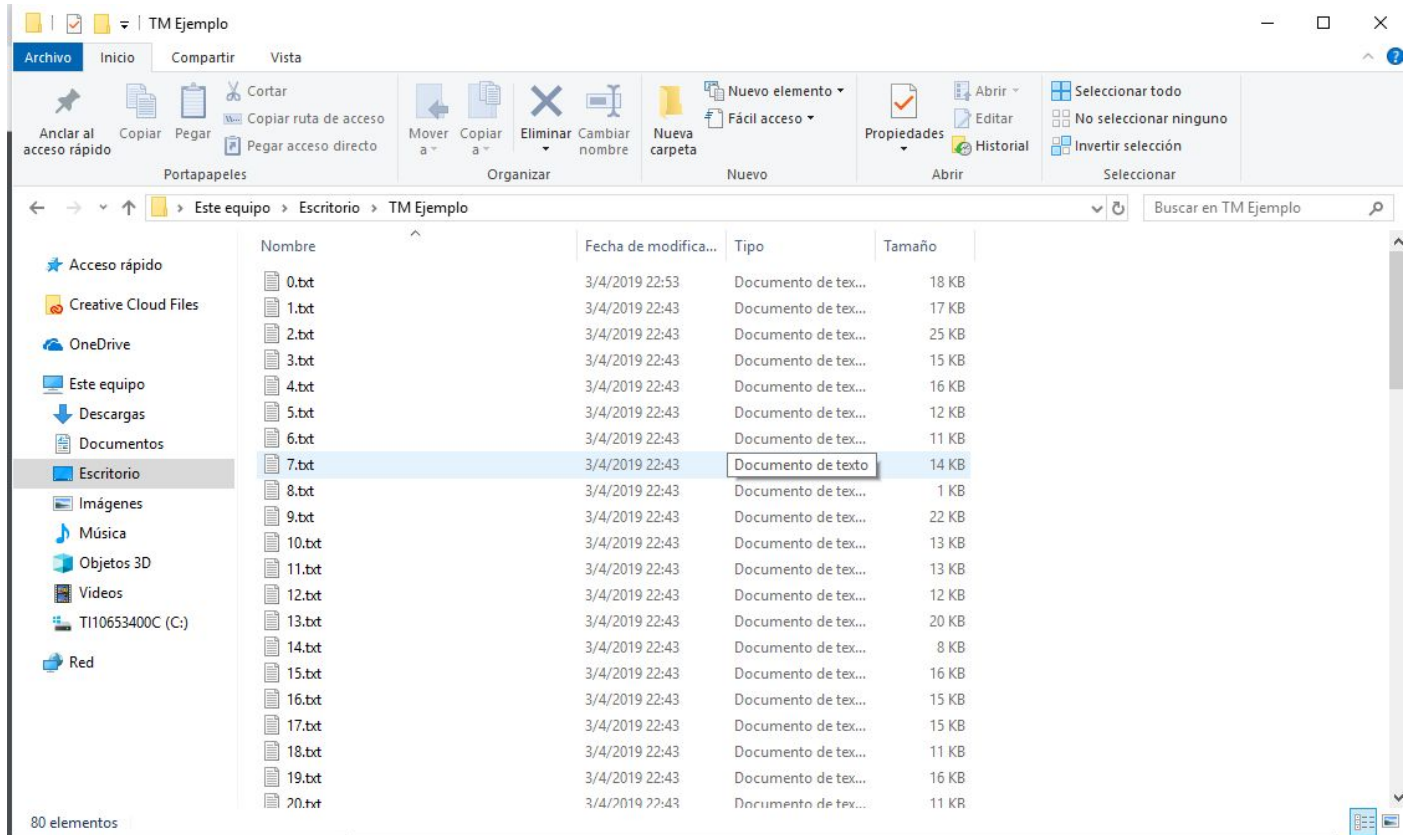
Row No.	Link
---------	------

ExampleSet (0 examples, 0 special attributes, 1 regular attribute)

Repository

Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- DB
- Local Repository (Freddy)
- Cloud Repository (disconnected)



\\Local Repository\Practica 2 TM FREDDY ABAD FINAL* – RapidMiner Studio Trial 9.2.001 @ DESKTOP-PAKQTSI

\\Local Repository\Practica 2 TM FREDDY ABAD FINAL* – RapidMiner Studio Trial 9.2.001 @ DESKTOP-PAKQTSI

Opción 4

Opción 4



Views: Design Results Turbo Prep Auto Model

Find data, operators...etc

All Studio

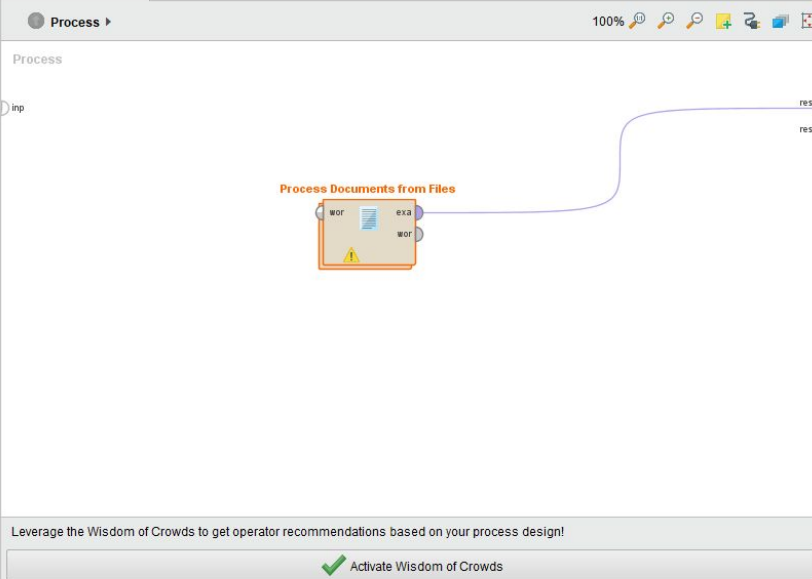
Repository

- [Import Data](#)
- Training Resources (connected)
 - Samples
 - Community Samples (connected)
 - DB
 - Local Repository (Freddy)
 - Temporary Repository (Freddy)
 - Cloud Repository (disconnected)

Operators

- [Process Documents from Files](#)
- Extensions (1)
- Text Processing (1)
 - [Process Documents from Files](#)
- No results were found.

Process



Parameters

Process Documents from Files

text directories [Edit List \(2\)...](#)

file pattern *

☒ extract text only

☐ use file extension as type

content type txt

encoding UTF-8

[Hide advanced parameters](#)

Help

Process Documents from Files

Text Processing

Tags: [Text Processing](#)

Synopsis

Generates word vectors from a text collection stored in multiple files.

Ejercicio Propuesto

FileEditProcessViewConnectionsCloudSettingsExtensionsHelp

Views:DesignResultsTurbo PrepAuto Model

Find data, operators...etc

All Studio

Repository

Import Data

Training Resources (connected)

Samples

Community Samples (connected)

DB

Local Repository (Freddy)

Temporary Repository (Freddy)

Cloud Repository (disconnected)

Operators

Search for Operators

Unity (4)

Get Page

Get Pages

Crawl Web

Process Documents from Web

Read RSS Feed

Clear Cookies

Get more operators from the Marketplace

Process

Process

100%

inp

res

res

Crawl Web

exa

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

Parameters

Crawl Web

url

vatch?v=_KhsQ3nn6Kw

crawling rules

Edit List (0)...

max crawl depth

2

retrieve as html

enable basic auth

add content as attribute

Hide advanced parameters

Help

Crawl Web

Web Mining

Tags: Crawl, Crawler, Spider, Web Mining

Synopsis

This operator allows to crawl the web and store the retrieved links and pages in an ExampleSet or on disk.

FileEditProcessViewConnectionsCloudSettingsExtensionsHelp

Views:DesignResultsTurbo PrepAuto Model

Find data, operators...etc

All Studio

Result History

ExampleSet (Crawl Web)

Open in

Turbo Prep

Auto Model

Filter (230 / 230 examples): all

Data

Statistics

Visualizations

Annotations

Row No.	Link
1	https://www.y...
2	https://www.y...
3	https://www.y...
4	https://www.g...
5	https://www.y...
6	https://www.y...
7	https://www.y...
8	https://www.g...
9	https://www.y...
10	https://www.y...
11	https://www.g...
12	https://youtu...
13	https://policie...
14	https://suppor...
15	https://www.y...

ExampleSet (230 examples, 0 special attributes, 1 regular attribute)

Repository

Import Data

Training Resources (connected)

Samples

Community Samples (connected)

DB

Local Repository (Freddy)

Temporary Repository (Freddy)

Cloud Repository (disconnected)