

## Evaluating Data Narratives

This evaluation aims to determine the best way to describe a data analysis method in 6 different scenarios. For each scenario, you will get:

- 1) a situation where data analysis is needed,
- 2) a sketch of the data analysis workflow,
- 3) several alternative descriptions of the workflow in text.

You will be asked to rate how appropriate is each description to the given situation. Then you will be asked to choose the descriptions that you think fits the situation best, or provide a new description if you would describe the analysis in some other way.

Your participation in this user study is on a volunteer basis. You may discontinue the evaluation at any time. Your participation will be anonymous, and no personal information will be collected from you. If you have any questions, please contact Yolanda Gil at [gil@isi.edu](mailto:gil@isi.edu).

At the end don't forget to click "submit"!

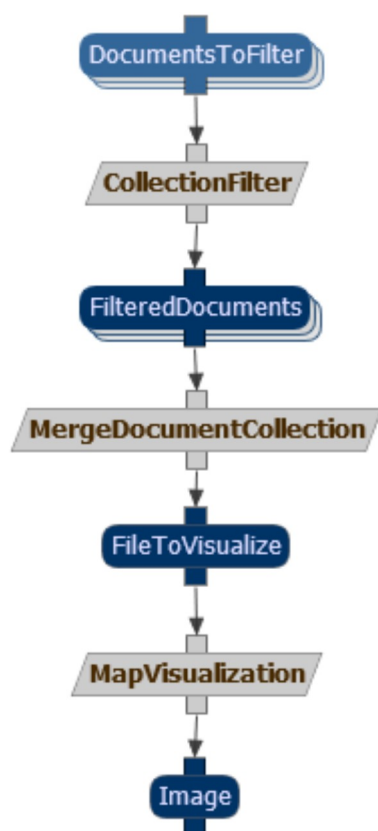
\*Required

### Scenario 1: Web Analytics Explanation for Supervisor

Bob works at Computer Inc. As part of his job, Bob has created a successful web page, which is accessed by users all over the world. The locations of the users who access the page are collected on a daily report. Bob uses this user access data in an analysis to produce a map-based visualization for his boss, Alice. The workflow depicted on Figure 1 illustrates the main steps for an accessibility analytics method that produces the map-based visualization given a set of daily user access reports. One day, Alice asks Bob which specific reports were used to create the latest visualization.

Below are several descriptions, please rate them based on whether you think they are appropriate for Bob to give Alice.

### **Figure 1: Workflow for accessibility analytics used for deriving the map figures**



## Description I

---

The "Accessibility Analytics" method was run on the "UserAccess01-04-2016.csv", "UserAccess01-01-2016.csv", "UserAccess01-05-2016.csv", "UserAccess01-02-2016.csv", "UserAccess01-03-2016.csv" datasets, which parameters "MINIMUMNUMBEROFACCESSES" set to 100, and "MAXIMUMNUMBEROFCOUNTRIES" set to 90.

1. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

1      2      3      4      5

Missing information                        Too detailed

## Description II:

---

The Image results have been derived from the "UserAccess01-02-2016.csv", "UserAccess01-05-2016.csv", "UserAccess01-03-2016.csv", "UserAccess01-01-2016.csv", and "UserAccess01-04-2016.csv" datasets.

**2. Is the level of detail of this description appropriate for the scenario? \****Mark only one oval.*

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

**Description III:**

---

The method "Accessibility Analytics" performs 1 main type of analysis on the input datasets. The Map Visualization step produces the main results, after a Combine step (Merge Document Collection), and a Filter step (Collection Filter).

The Image results are the product of the Map Visualization, a Data Analysis and Data Visualization step.

**3. Is the level of detail of this description appropriate for the scenario? \****Mark only one oval.*

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

**Description IV:**

---

The method "Accessibility Analytics" has three steps: First, the "CollectionFilter" processes the data, then the "MergeDocumentCollection" processes the results and finally the "MapVisualization" step produces the final results.

The Image results are the product of the "MapVisualization" step a Data Analysis and Data Visualization step.

**4. Is the level of detail of this description appropriate for the scenario? \****Mark only one oval.*

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

**Description V:**

---

The Image results were obtained from the execution of the method "Accessibility Analytics". First, the "CollectionFilter" component (implemented with a "RemoveReportData" step) processes the data, then the "MergeDocumentCollection" (implemented with the "JoinReports" step) processes the results and finally the "MapVisualization" (implemented with the "AllCountryVisualization" step) produces the final results.

The Image results are the product of the "AllCountryVisualization" step, a Data Analysis and Data Visualization step.

**5. Is the level of detail of this description appropriate for the scenario? \****Mark only one oval.*

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

**Description VI:**

---

The Image results were obtained from the execution of the method "Accessibility Analytics". First, the "CollectionFilter" component (implemented with a "RemoveReportData" step) processes the data, then the "MergeDocumentCollection" (implemented with the "JoinReports" step) processes the results and finally the "MapVisualization" (implemented with the "AllCountryVisualization" step) produces the final results

The RemoveReportData is a bash script that uses Perl to perform its functionality.

The JoinReports is a bash script that invokes a Java program for merging documents.

The AllCountryVisualization is a bash script that invokes a Java program for generating a simple map based visualization.

**6. Is the level of detail of this description appropriate for the scenario? \****Mark only one oval.*

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

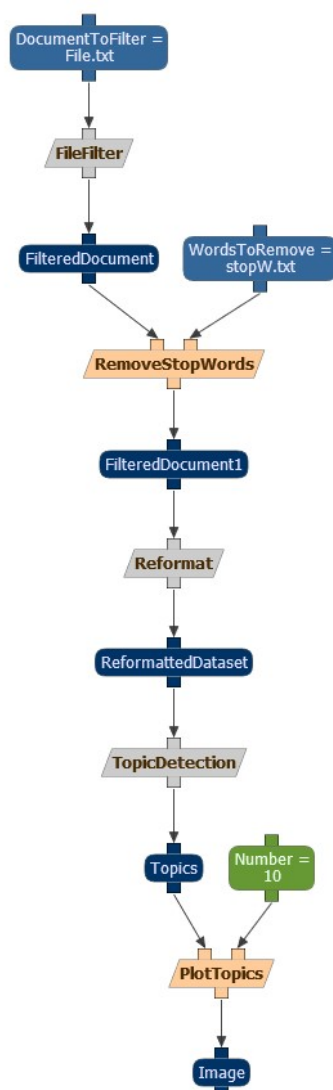
**7. Which description was most appropriate for the scenario? If none, choose "other" and write the description that you would provide \****Mark only one oval.*

- Description I
- Description II
- Description III
- Description IV
- Description V
- Description VI
- Other: .....

**Scenario 2: Detecting popular topics**

Clarence is an experienced programmer that works at Computer Inc. Alice asks him to implement a topic detection and visualization analysis that she has seen that her colleague Steve uses. Since he is very familiar with topic modeling as he has implemented several versions for the company in the past, Clarence then asks Alice for the main steps of the analysis. Alice asks her colleague Steve, who uses the workflow shown in Figure 2 and gives a description.

Below are several descriptions, please rate them based on whether you think they are appropriate to give Clarence.

**Figure 2: Topic detection workflow**

## Description I

---

The Detect Topics method was run on the "ProductReviews-CNET.txt", "ProductReviews-Amazon.txt" datasets, with "NUMBEROfTopics" set to 10, and "ITERATIONS" set to 5.

8. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description II

---

The Detect Topics method was derived from the "ProductReviews-CNET.txt", "ProductReviews-Amazon.txt" datasets.

9. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description III

---

The method "Detect Topics" performs 1 main type of analysis on the input datasets. The Topic Detection step produces the main results of the workflow, after a FormatTransformation step (Reformat step), a Filter step (Remove Stop Words step), and a Filter step (File Filter step).

The Image results are the product of the Plot Topics Node, a Data Visualization step.

10. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description IV:

---

The method "Detect Topics" has five steps. First, the FileFilter step processes the data, followed by the RemoveStopWords, Reformat and TopicDetection steps. Finally, the "PlotTopics" step produces the final results.

The Image results are the product of the "PlotTopics" step, a Data Visualization step.

11. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description V:

---

The Image results have been obtained from the execution of the method "Detect Topics". First, the FileFilter (implemented with a "RemoveMarkup" step) step processes the data, followed by the RemoveStopWords, Reformat (implemented with the "FormatMallet" step) and TopicDetection (implemented with the "LDA" step) steps. Finally, the "PlotTopics" step produces the final results.

The Image results are the product of the "PlotTopics" step, a Data Visualization step.

12. Is the level of detail of this description appropriate for this scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description VI:

---

The Image results were obtained from the execution of the method "Detect Topics". First, the FileFilter (implemented with a "RemoveMarkup" step) step processes the data, followed by the RemoveStopWords, Reformat (implemented with the "FormatMallet" step) and TopicDetection (implemented with the "LDA" step) steps. Finally, the "PlotTopics" step produces the final results.

The RemoveMarkup step is a bash script that uses Java to perform its functionality

The RemoveStopWords step is a bash script that uses a Java script to perform its functionality

The FormatMallet step is a bash script that uses Java to prepare the dataset

The LDA step is a bash script that uses a Weka implementation of the LDA algorithm

The PlotTopics is a bash script that uses Java to perform its functionality.

13. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

14. Which description was most appropriate for the scenario? If none, choose "other" and write the description that you would provide \*

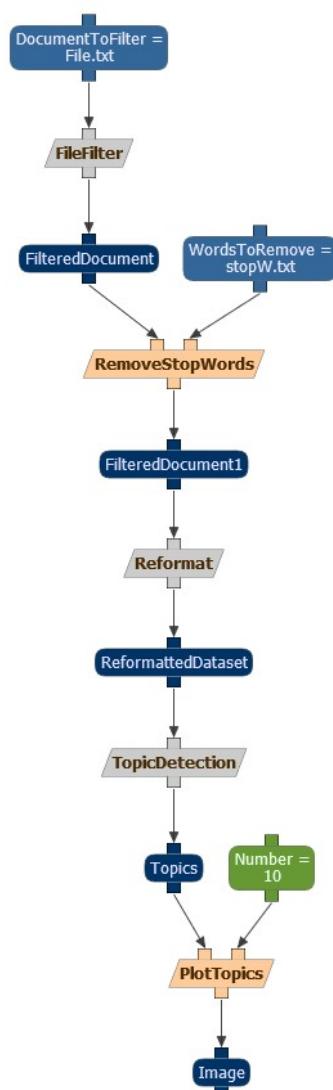
Mark only one oval.

- Description I
- Description II
- Description III
- Description IV
- Description V
- Description VI
- Other: .....

## Scenario 3: Detecting popular topics: junior programmer

David is a new programmer that just started working at Computer Inc. Alice asks him to implement a topic detection and visualization analysis that she has seen that her colleague Steve uses. David tells Alice he is not familiar with topic detection, and asks her for a description of the analysis. Alice asks her colleague Steve, who uses the workflow shown in Figure 2 and gives her a description.

Below are several descriptions, please rate them based on whether you think they are appropriate to give David.

**Figure 3: Topic detection workflow**

## Description I

---

The Detect Topics method was run on the "ProductReviews-CNET.txt", "ProductReviews-Amazon.txt" datasets, with "NUMBEROfTopics" set to 10, and "ITERATIONS" set to 5.

15. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description II:

---



The Image results have been derived from the "ProductReviews-CNET.txt", "ProductReviews-Amazon.txt" datasets.

16. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description III

---

The method "Detect Topics" performs 1 main type of analysis on the input datasets. The Topic Detection step produces the main results of the workflow, after a FormatTransformation step (Reformat step), a Filter step (Remove Stop Words step), and a Filter step (File Filter step).

The Image results are the product of the Plot Topics Node, a Data Visualization step.

17. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description IV:

---

The method "Detect Topics" has five steps. First, the FileFilter step processes the data, followed by the RemoveStopWords, Reformat and TopicDetection steps. Finally, the "PlotTopics" step produces the final results.

The Image results are the product of the "PlotTopics" step, a Data Visualization step.

18. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description V:

---

The Image results were obtained from the execution of the method "Detect Topics". First, the FileFilter (implemented with a "RemoveMarkup" step) step processes the data, followed by the RemoveStopWords, Reformat (implemented with the "FormatMallet" step) and TopicDetection (implemented with the "LDA" step) steps. Finally, the "PlotTopics" step produces the final results.

The Image results are the product of the "PlotTopics" step, a Data Visualization step.

19. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description VI:

---

The Image results were obtained from the execution of the method "Detect Topics". First, the FileFilter (implemented with a "RemoveMarkup" step) step processes the data, followed by the RemoveStopWords, Reformat (implemented with the "FormatMallet" step) and TopicDetection (implemented with the "LDA" step) steps. Finally, the "PlotTopics" step produces the final results.

The RemoveMarkup step is a bash script that uses Java to perform its functionality

The RemoveStopWords is a bash script that uses a Java script to perform its functionality

The FormatMallet step is a bash script that uses Java to prepare the dataset

The LDA step is a bash script that uses a Weka implementation of the LDA algorithm

The PlotTopics is a bash script that uses Java to perform its functionality.

20. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

21. Which description was most appropriate for the scenario? If none, choose "other" and write the description that you would provide \*

Mark only one oval.

- Description I
- Description II
- Description III
- Description IV
- Description V
- Description VI
- Other: .....

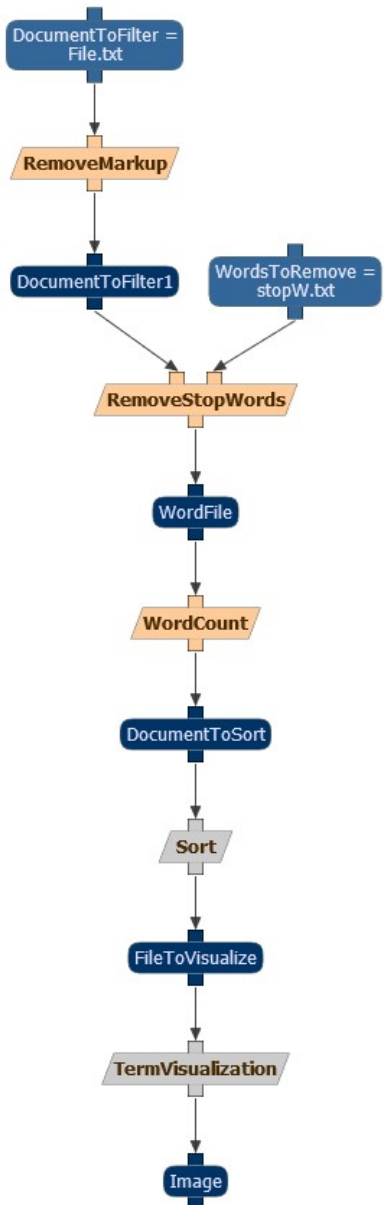
## Scenario 4: Popular terms across the years

Alice likes to analyze the most popular terms on Computer Inc's main web page. She runs an analysis with today's data. The analysis uses the workflow shown in Figure 4 to detect and illustrate the most popular terms in the web page, although Alice has never seen the workflow. Alice is asked by her boss if the popular terms today are different from the terms on the same day last year. Alice run the analysis last year and has the results, and she would like to make sure that the analysis was done in the same way as she did it today so that the results are comparable. She asks Clarence for a description of the analysis that was done on today's data.

Below are several descriptions, please rate them based on whether you think they are appropriate to

give Alice.

**Figure 4: Workflow for visualizing popular terms**



## Description I

---

The "Trending Words Visualization" method was run on the "MainPage-10Oct2016.txt" dataset, with the "SORT APLPHABETICALLY" parameter set to "false".

22. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

1      2      3      4      5

Missing information                        Too detailed

## Description II:

---

The Image results were derived from the "MainPage-10Oct2016.txt" dataset.

23. Is the level of detail of this description appropriate for the scenario? \*

*Mark only one oval.*

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description III:

---

The method "Trending Words Visualization" performs 1 main type of analysis on the input datasets. The Term Visualization step produces the main results of the workflow, after Filter step (RemoveMarkup step), a Filter step (Remove Stop Words step), and a Sorting step (Quicksort step).

24. Is the level of detail of this description appropriate for the scenario? \*

*Mark only one oval.*

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description IV:

---

The method "Trending Words Visualization" has five steps. First, the RemoveMarkup step processes the data, followed by the RemoveStopWords, WordCount and Sort steps. Finally, the "TermVisualization" step produces the final results.

The Image results are the product of the "TermVisualization" step, a Data Visualization step.

25. Is the level of detail of this description appropriate for the scenario? \*

*Mark only one oval.*

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description V:

---

The Image results were obtained from the execution of the method "Trending Words Visualization". First, the RemoveMarkup step processes the data, followed by the RemoveStopWords, WordCount and Sort (implemented with the "Quicksort" step) steps. Finally, "TermVisualization" (implemented with the TagCloud step) produces the final results.

The Image results are the product of the "TermVisualization" step, a Data Visualization step.

26. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description VI:

---

The Image results were obtained from the execution of the method "Trending Words Visualization". First, the RemoveMarkup step processes the data, followed by the RemoveStopWords, WordCount and Sort (implemented with the "Quicksort" step ) steps. Finally, the "TermVisualization" step produces the final results.

The RemoveMarkup step is a bash script that uses Java to perform its functionality

The RemoveStopWords step is a bash script that uses a Java script to perform its functionality

The WordCount step is a bash script that uses a Java script to perform its functionality

The Quicksort step is a bash script that invokes a C program with an implementation of the quicksort algorithm

The TagCloud step is a bash script that invokes a Java program to perform its functionality

27. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

28. Which description was most appropriate for the scenario? If none, choose "other" and write the description that you would provide \*

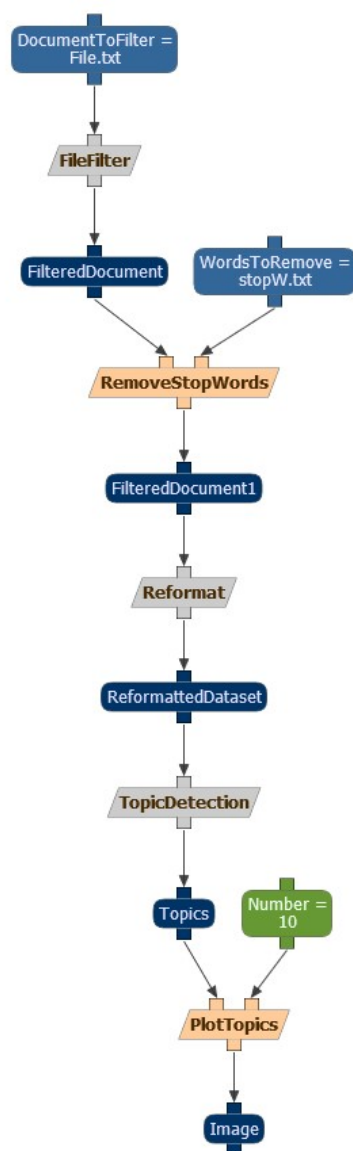
Mark only one oval.

- Description I
- Description II
- Description III
- Description IV
- Description V
- Description VI
- Other: .....

## Scenario 5: Topic detection for a board of directors

After Clarence successfully implemented the topic detection workflow shown in Figure 5, he runs an analysis and shows Alice the results. Alice needs to make a presentation of the results to the board of directors of the company. The members of the board of directors are mostly investors and lawyers. Alice asks Clarence for a description of the analysis so she can explain the results to the board of directors.

Below are several descriptions, please rate them based on whether you think they are appropriate to give the board of directors.

**Figure 5: Workflow for visualizing popular topics**

## Description I

---

The "Topic Detection" method was run on the "MainPage-10Oct2016.txt" dataset, with the "ITERATIONS" parameter set to "5".

29. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description II:

---

The Image results have been derived from the "MainPage-10Oct2016.txt" dataset.

30. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

### Description III:

---

The method "Topic Detection" performs 1 main type of analysis on the input datasets. The Topic Detection step produces the main results of the workflow, after a FormatTransformation step (Reformat step), a Filter step (Remove Stop Words step), and a Filter step (File Filter step).

The Image results are the product of the Plot Topics Node, a Data Visualization step.

31. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

### Description IV:

---

The method "Detect Topics" has five steps. First, the FileFilter step processes the data, followed by the RemoveStopWords, Reformat and TopicDetection steps. Finally, the "PlotTopics" step produces the final results.

The Image results are the product of the "PlotTopics" step, a Data Visualization step.

32. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

### Description V:

---

The Image results have been obtained from the execution of the method "Detect Topics". First, the FileFilter (implemented with a "RemoveMarkup" step) step processes the data, followed by the RemoveStopWords, Reformat (implemented with the "FormatMallet" step) and TopicDetection (implemented with the "LDA" step) steps. Finally, the "PlotTopics" step produces the final results.

The Image results are the product of the "PlotTopics" step, a Data Visualization step.

33. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description VI:

---

The Image results were obtained from the execution of the method "Detect Topics". First, the FileFilter (implemented with a "RemoveMarkup" step) step processes the data, followed by the RemoveStopWords, Reformat (implemented with the "FormatMallet" step) and TopicDetection (implemented with the "LDA" step) steps. Finally, the "PlotTopics" step produces the final results.

The RemoveMarkup step is a bash script that uses Java to perform its functionality

The RemoveStopWords step is a bash script that uses a Java script to perform its functionality

The FormatMallet step is a bash script that uses Java to prepare the dataset

The LDA step is a bash script that uses a Weka implementation of the LDA algorithm

The PlotTopics is a bash script that uses Java to perform its functionality.

34. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

35. Which description was most appropriate for the scenario? If none, choose "other" and write the description that you would provide \*

Mark only one oval.

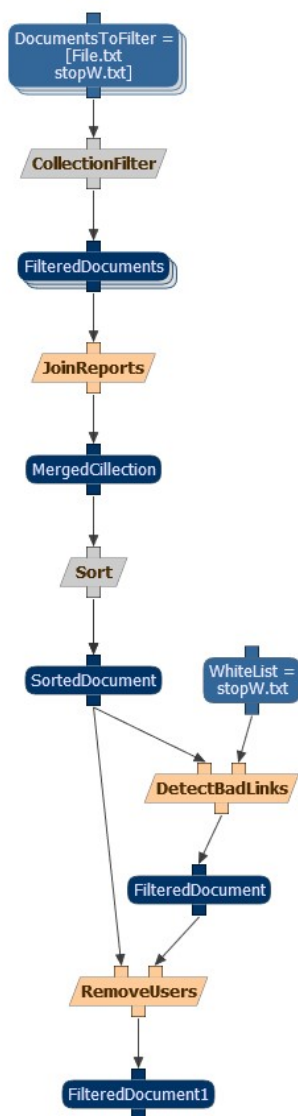
- Description I
- Description II
- Description III
- Description IV
- Description V
- Description VI
- Other: .....

## Scenario 6: Spam detection for a system administrator

The web site of Computer Inc goes under a cyber attack, and a series of users posted spam links on some of the web pages. Clarence knows of a method for detecting attackers (shown in Figure 6) and needs to send specific details to Emily, the system administrator who is very rigorous in terms of the software being installed on the server particularly when undergoing a cyber attack. He sends a description to Emily.

Below are several descriptions, please rate them based on whether you think they are appropriate to give Emily.



**Figure 6: Workflow for detecting attackers**

## Description I

---

The "Address problematic users" method was run on the "UsersAndContributions.txt" and "WhitelistLinks" datasets, with the "SORT APLHABETICALLY" parameter set to "false".

36. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description II:

---

The UsersToRemove results have been derived from the "UsersAndContributions.txt" dataset.

37. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

### Description III:

---

The method "Address problematic users" performs 1 main type of analysis on the input datasets. The Remove Users step produces the main results of the workflow, after a Filter step (Extract Links), a Merge step (Join Reports), a Sort step (MergeSort) and a Filter step (Detect Bad Links).

The "UsersToRemove" results are the product of the RemoveUsers, a Filtering step.

38. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

### Description IV:

---

The method "Address problematic users" has five steps. First, the CollectionFilter step processes the data, followed by the JoinReports, Sort and DetectBadLinks steps. Finally, the "RemoveUsers" step produces the final results.

The Image results are the product of the "RemoveUsers" step, a Data Visualization step.

39. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

### Description V:

---

The UsersToRemove results have been obtained from the execution of the method "RemoveUsers". First, the CollectionFilter step (implemented with "ExtractLinks" ) processes the data, followed by the JoinReports step, Sort step (implemented with "MergeSort" ) and DetectBadLinks step. Finally, the "RemoveUsers" step produces the final results.

The UsersToRemove results are the product of the "RemoveUsers" step, a Filtering step.

40. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

## Description VI:

---

The UsersToRemove results have been obtained from the execution of the method "RemoveUsers". First, the CollectionFilter step (implemented with "ExtractLinks" ) processes the data, followed by the JoinReports step, Sort step (implemented with "MergeSort" ) and DetectBadLinks step. Finally, the "RemoveUsers" step produces the final results.

The ExtractLinks step is a bash script that uses a Java program to perform its functionality. The JoinReports step is a bash script that invokes Python to merge the input datasets. The MergeSort step is a bash script that uses a Java program to sort the input datasets. The DetectBadLinks step is a bash script that uses Python to perform its functionality. The RemoveUsers step is a bash script that uses Python to perform its functionality

41. Is the level of detail of this description appropriate for the scenario? \*

Mark only one oval.

	1	2	3	4	5	
Missing information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too detailed

42. Which description was most appropriate for the scenario? If none, choose "other" and write the description that you would provide \*

Mark only one oval.

- Description I
- Description II
- Description III
- Description IV
- Description V
- Description VI
- Other: .....

43. Please let us know if the answers were difficult, or if you have any other feedback

.....

.....

.....

.....

.....

