

North Campus  
School of ICT  
Centre for Research in Information and Cyber Security  
 s211144843@mandela.ac.za

Dear participant,

Thank you for taking the time to complete this questionnaire. We truly value the information you have provided.

The rate of information that is being produced in research has rapidly increased in recent years. Researchers are struggling to understand or comprehend the large amount of information being thrown around. Thus, created the need to identify topics in research papers.

The high-level aim of this research is to create a model that helps researchers identify topics within a domain.

The model will consist of employing certain Natural Language Processing (NLP) techniques and algorithms such as; tokenization, lemmatization, n-grams, removal of stop words and, Latent Dirichlet Allocation, to achieve the high-level aim as mentioned in above paragraph.

The contribution of this questionnaire is to see if the mapping between the NLP generated topics and the original content can be made.

Your participation in this research project is completely voluntary. You may decline altogether or leave blank any questions you don’t wish to answer. There are no known risks to participation beyond those encountered in everyday life. Your responses will remain confidential and anonymous. Data from this research will be kept under lock and key and reported only as a collective combined total. No one other than the researchers will know your individual answers to this questionnaire.

Once again, we are extremely grateful for your contributing your valuable time, your honest information, and your thoughtful suggestions.

Many thanks,

Juandre van Heerden

Student number: 211144843

Email: s211144843@mandela.ac.za

You are required to perform 2 tasks:

1. Completing the Title + Abstract coding spreadsheet.
2. Completing the Clusters coding spreadsheet.

**Task 1: Completing the Title + Abstract coding spreadsheet.**

In this spreadsheet, you will be shown academic papers titles and abstracts separated by a “-“. You are required to complete the spreadsheet by filling in the primary topic and, if applicable, the secondary topic column.

Step 1: Open the excel document named “Coding – Title + Abstract”.

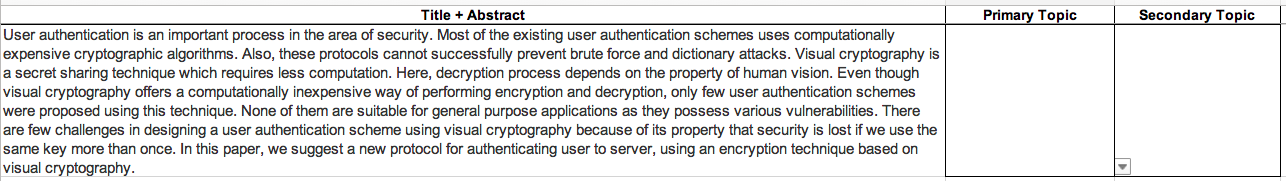
Step 2: Read through the papers title and abstract provided in the Title + Abstract column. As represented by Figure 1.

Figure 1 – Demo example - Title + Abstract example

Please note: Only one option can be selected for each topic.

Step 3: Once completed, select the most accurate primary topic from the drop-down menu.

An accurate primary topic would be the core topic that is being discussed, described, or dealt with. The demo example in Figure 2 covers a variety of cryptography methods used in a network security setting. Therefore, the primary topic will be cryptography, and the secondary topic will be Network & Communication security as illustrated in Figure 3.

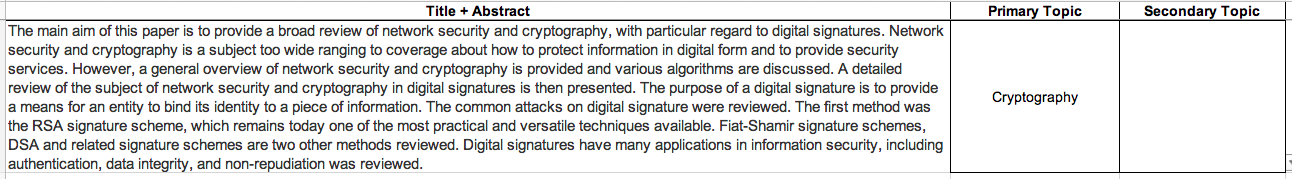


Figure 2 – Demo example - Primary topic

Step 4: Once completed, select the most accurate secondary topic from the drop-down menu.

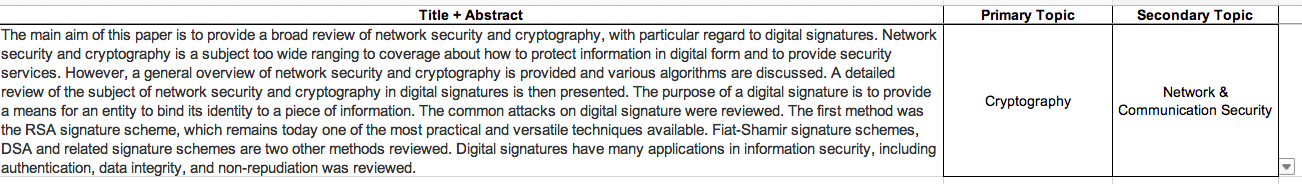
Where papers have secondary topics, there will be a column labeled “Secondary topics” from which selections can be made.

Figure 3 - Demo example - Secondary topic

Step 5: When you have completed the Title + Abstract coding spreadsheet, email me the completed spreadsheet at [s211144843@mandela.ac.za](mailto:s211144843@mandela.ac.za)

**Task 2: Completing the Clusters coding spreadsheet.**

In this spreadsheet you will be shown Clusters. You are required to complete the spreadsheet by filling in the primary topic and, if applicable, the secondary topic column.

Step 1: Open the excel document named “Coding – Clusters”.

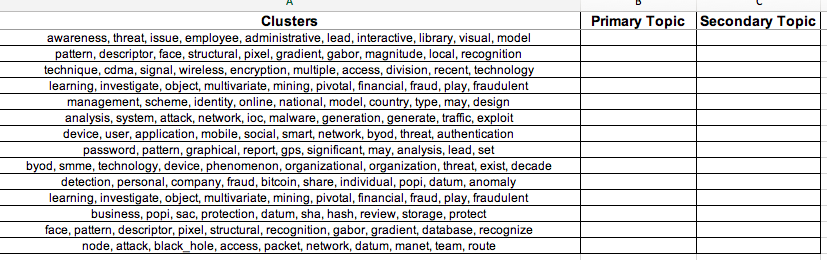
Step 2: Read through the clusters provided in the Clusters column as illustrated in Figure 4.

Figure 4 - Demo example - Clusters

Please note: Only one option can be select for each topic.

Step 3: Once completed, select the most accurate primary topic from the drop-down menu.

An accurate primary topic would be the core topic that is being discussed, described, or dealt with. The demo example in Figure 5 mentions network security terminology in conjunction with the term “access” which can lead to belief that it covers access control. Therefore, the primary topic will be Network and Communication security, and the secondary topic will be Access control as illustrated in Figure 6.

Figure 5 - Demo example – Primary topic



Step 4: Once completed, select the most accurate secondary topic from the drop-down box.

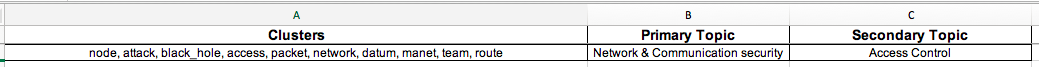
Where papers have secondary topics, there will be a column labeled “Secondary topics” from which selections can be made.

Figure 6 - Demo example - Secondary topic

Step 5: When you have completed the Clusters coding spreadsheet, email me the completed spreadsheet at [s211144843@mandela.ac.za](mailto:s211144843@mandela.ac.za)