




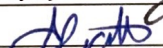
MCAST

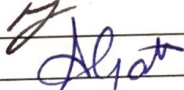
Malta College of Arts, Science & Technology


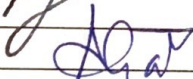
**INSTITUTE OF INFORMATION
AND COMMUNICATION TECHNOLOGY**

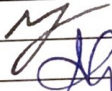
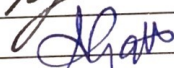
Project Supervision Log Book

Institute	Institute of Information and Communication Technology
Programme	BSc. Honours in Business Analytics
Project Title	Using Natural Language Processing to assure Data Quality in Web Scraping Operations.
Supervisor	Alan Gatt
Student	Matthew De Giorgio
Student ID No	0047796(M)

Meeting	Date: 15/03/2019
Issues discussed at the meeting The direction of the project. Will focus on global suppliers passing through Suez, Gibraltar. Gather recent news (like RSS and Google News)	
Supervisor Recommendations Identify the process of gathering and cleaning data. Mainly: <ul style="list-style-type: none">scraped data > NLP process > cleans data > aggregate data by same news > extract relative keywords Gain knowledge with NLTK on Python. The possible direction of using air data instead of shipping.	
Date of Next Meeting	20/3/19
Student Signature	20/3/19 
Supervisor Signature	

Meeting	Date: 20/3/19
Issues discussed at the meeting Database to store scraped data Air traffic or ship traffic direction will affect the overall focus of the project	
Supervisor Recommendations Start scraping in Python. Set up Git and give me access.	
Date of Next Meeting	29/3/2019
Student Signature	
Supervisor Signature	

Meeting	Date: 28/3/2019
Issues discussed at the meeting	
<p>Matthew decided on using Air Traffic control data.</p> <p>Articles on air traffic will be scraped related to airline delay, weather delay, airport issues, customs issues, etc.... From these articles Location (Country, City, Airport), type of event, duration if applicable.</p>	
Supervisor Recommendations	
Finish SOL.	
Date of Next Meeting	10/5/2019
Student Signature	
Supervisor Signature	

Meeting	Date: 10/5/2019
Issues discussed at the meeting Setting up GIT and uploading files. Initial proposal finished off. Went through existing work and prepared a plan of what needs to be done. Scraping is working and getting results (articles related to air traffic delays)	
Supervisor Recommendations Use NLP to extract: <ul style="list-style-type: none"> - Location - Time - Duration - Affected Airways - Cause 	
Date of Next Meeting	30/5/2019
Student Signature	
Supervisor Signature	

Issues discussed at the meeting

- Ethical Considerations in the SOIs --> Use of Business SOIs.
- Which Research Poster template to use? --> Use New Template
- IEEE --> Research Methodology Section, the research philosophy has to be explained, would it be appropriate to include Artificial Intelligence Philosophies? (can mention)
- IEEE Keywords: NLP, Natural Language Processing, Web Scraping, AI, Artificial Intelligence, Python Programming
- IEEE: After the Abstract, you need to make an introduction section or start with the Literature Review Heading and put an intro under it? --> Move Introduction Section from the literature review section
- IEEE: Evaluation Sub Topics:
 - Small Explainer on Sentiment Analysis
 - For all Tests, comparison of single vs multi-tag runs
 - Data Quality Metric and how the final Data set compares
 - Metric, how many of the designated tags were encountered within each article.
- Current Literature Review Sub Topics - Artificial Intelligence, Web Scraping, Natural Language Processing, Data Quality
- Research MethodologySection, Is a sampling method required?

Accuracy Chart --> Open the articles, manually count the tags and compare to how many the program caught. (Visualise).

Video ---> Walk Through (leave a text file link in the git repository)

Share Repo with Ivan.

Supervisor Recommendations

- The Poster Template was finalised
- Discussion on Research Philosophy, mention it but do go into great detail
- Finalised Tags
- Changes to Literature Review intro and IEEE introduction
- Methodology on determining the accuracy of the program.

Date of Next Meeting	
Student Signature	<i>[Signature]</i>
Supervisor Signature	<i>[Signature]</i>

Project Supervision Meeting Summary Sheet				
Meeting No	Planned Meeting Date	Actual Meeting Date	Attended (Y/N)	Remarks (if any)
1	15/03/2019	15/03/2019	Yes	
2	20/03/2019	20/03/2019	Yes	
3	29/03/2019	28/03/2019	Yes	
4	30/5/2019	30/5/2019	No	
5	04/06/2019	04/06/2019	Yes	
6	10/5/2019	10/5/2019	Yes.	
7				
8				
9				

Total Planned Meetings	5	6
Total Actual Meetings	4	5

Supervisor Signature	Student Signature
	