Argument Mining of Ruggero Lemma

In this laboratory in the first moment I download the three datasets inserted in the pdf of this work. Then I start a first phase of pre-processing of data in which using a python code, I have generated some json file that a web app can use to show the different type of data visualization. The three parts that we must develop are the visualization of:

1. Argument pairs from debate platforms (Debatepedia and ProCon)
2. Argument components and relations from persuasive essays:
3. Argument components from political speeches

PRE-PROCESSING OF DATA

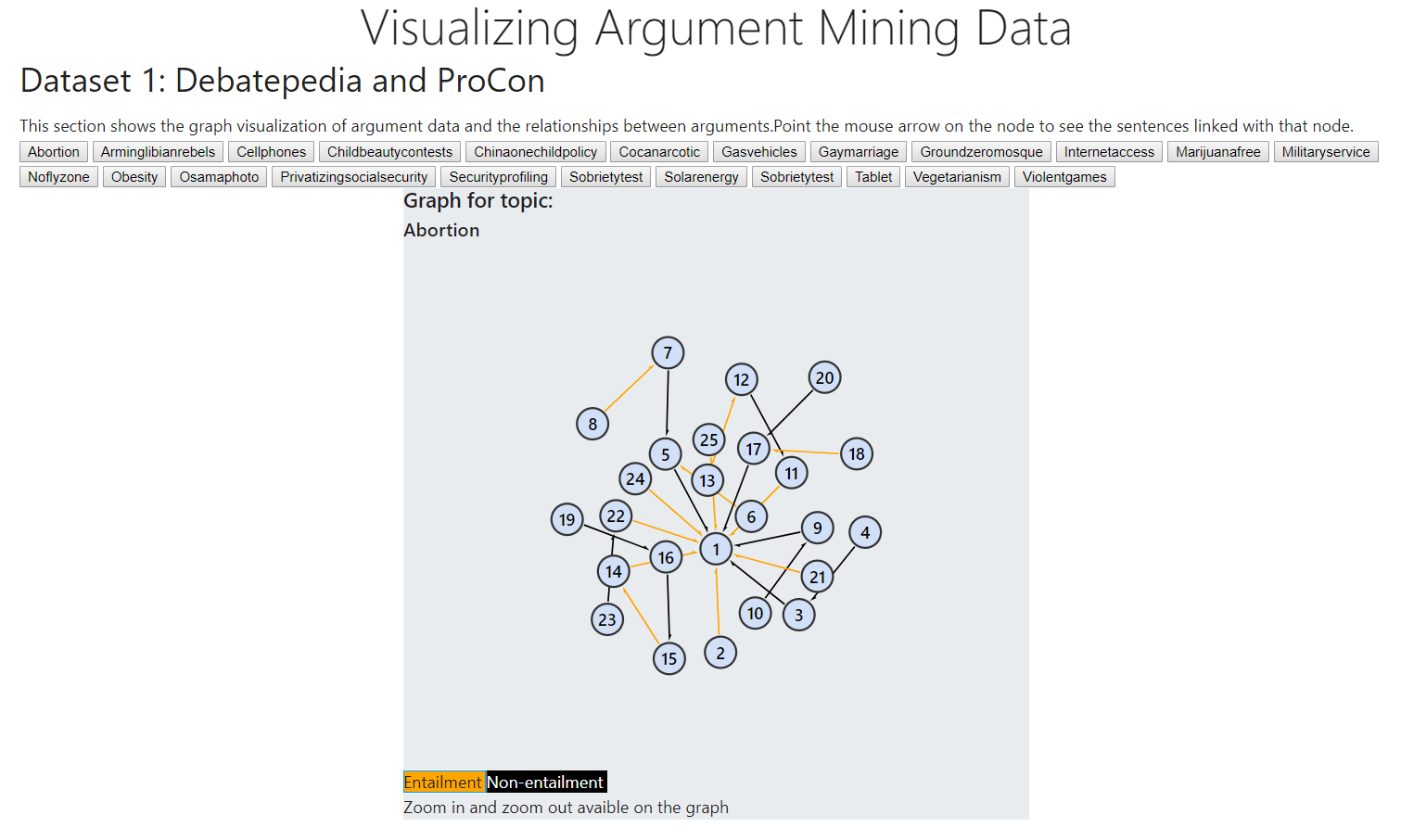
In the first part of the python file, I import the “xml.etree.ElementTree” library to parse the file datasetXML.xml that contain all the data necessary to create the graph of the first point. For that file I have joined the three xml file because they have the same structure, I have correct some mistake and I have change the value of the attribute “entailment” with the value 1 if there is one and 0 if there isn’t, to make easier the use of this. Then in the python code, I import the “networkx” library to create the graph and here there will be one graph for each topic. Then I work on every pair to create a directed graph and adding the nodes and edges. At the end I import the “json\_graph” and “json” library to be able to create for every topic a json file with a graph structure and I will use that files in my web app.

The second part of the python code is to create json file for each essay. I start importing the “csv” library to be able to use the topics.csv file. In this part for every essay is created a graph with the distinction among major claim, claim and premise that in the web app will be showed with distinct colours.

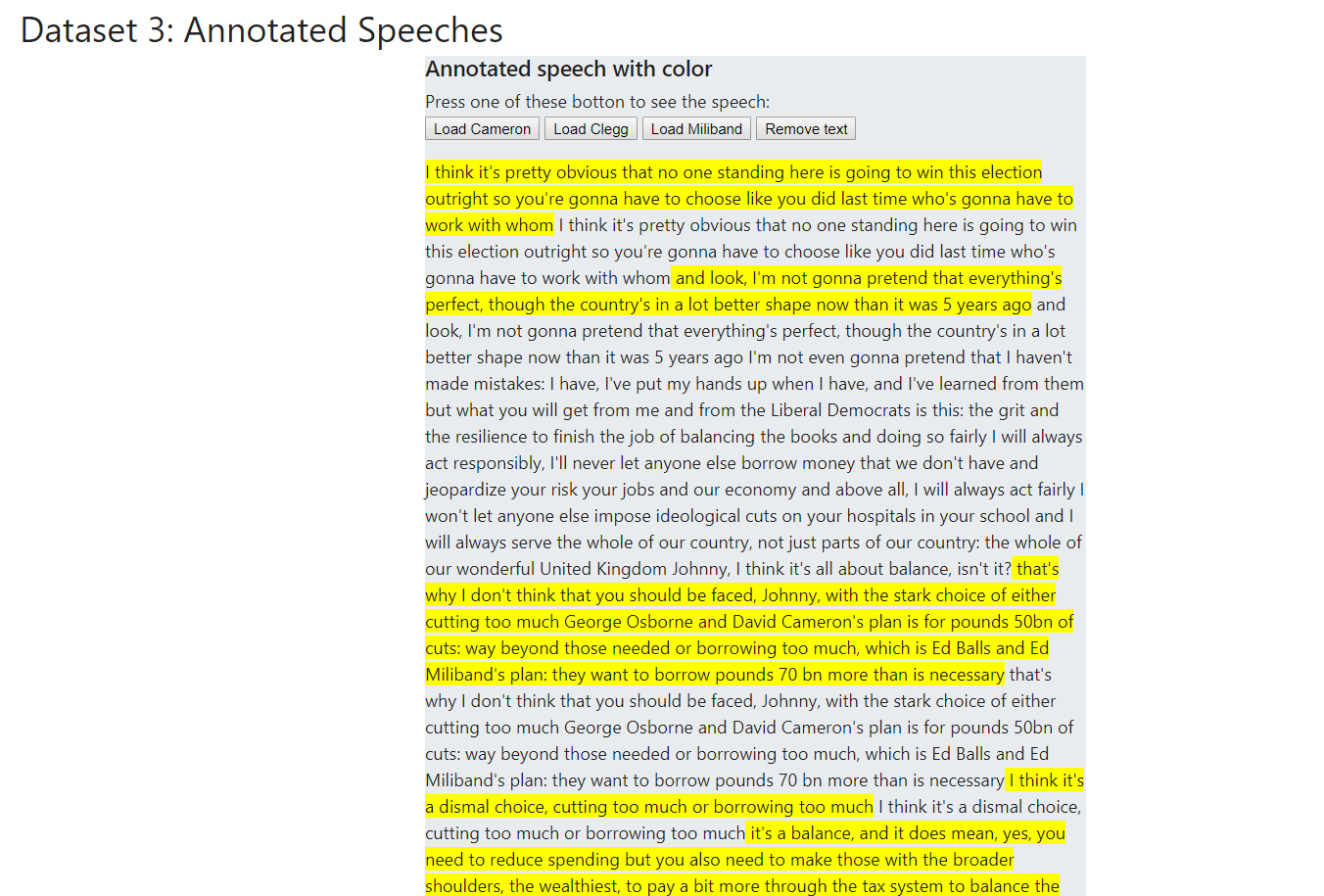
WEB APP

In this part, I follow the advice of the professor that told me to do before the first and third point and then reuse some stuff of that point to do the second point.

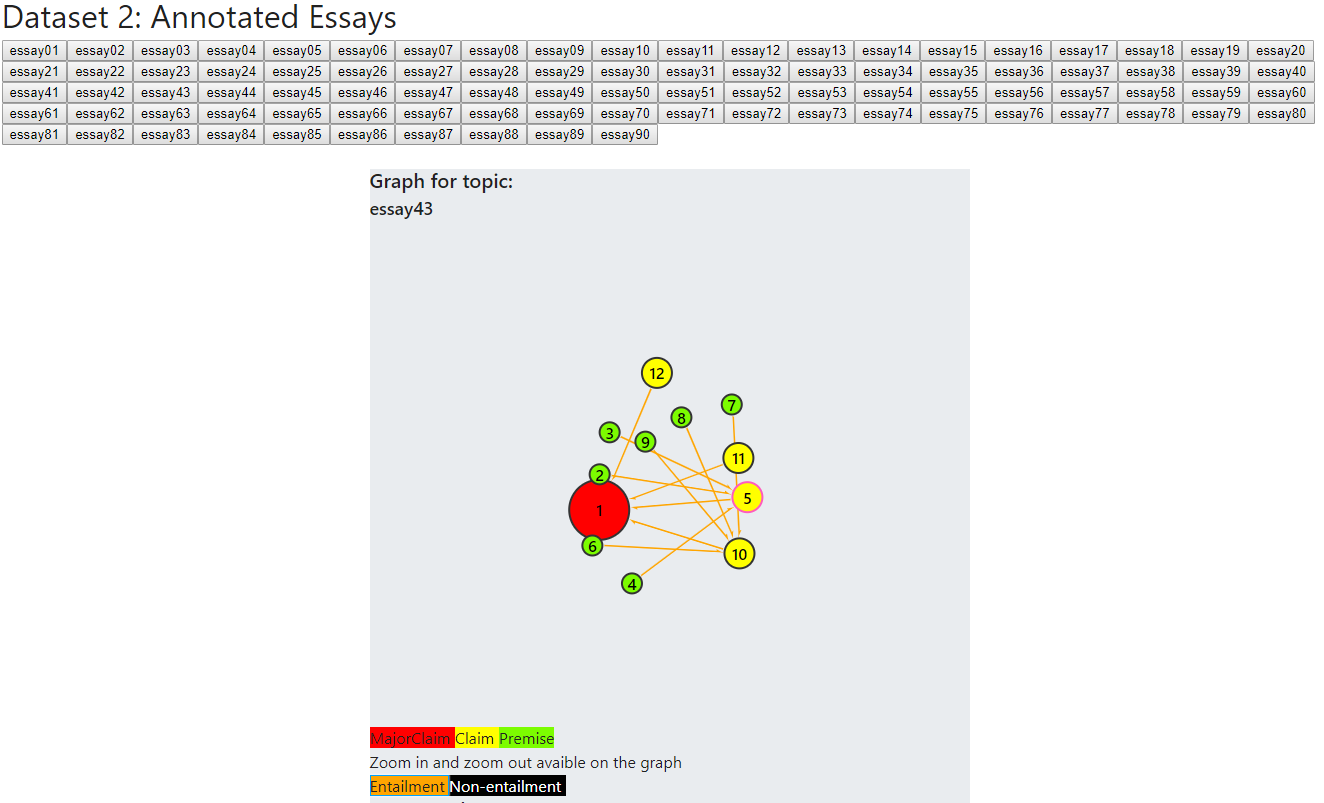
In the first part(Debatepedia), the main function is called “caricaArgomento” and with that I load the json file that are in the directory called “json\_files” and I take an argument, I create the node, link that node with the different part of the argument, using the id of a certain part, and at the end if between the different part there is an entailment, the oriented line is orange otherwise the colour of the edge is black. The function “caricaArgomento” is called when you click on an argument and the previous graph is removed and the new one selected is shown. In the web page the default topic loaded is “Abortion”.



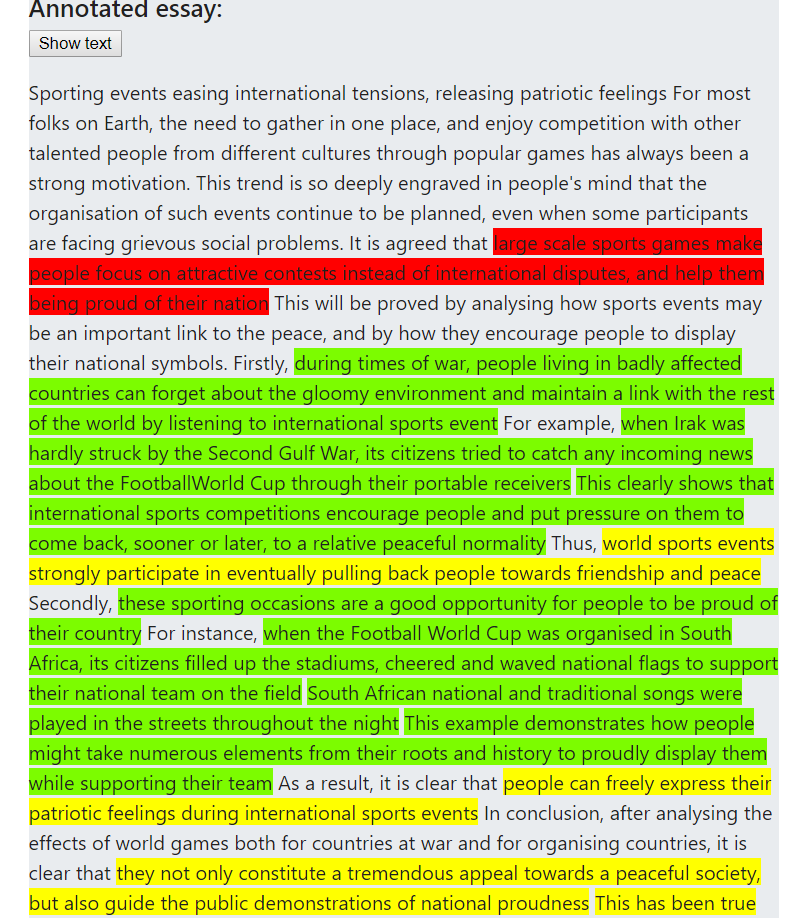
In the third part called “Annotated Speeches”, in the main.js , I take the file of the three speeches and I load them, then with the most important function of this part, called “discorso”, I create the speech\_html that then the html page will show. In the next image there is an example of this part.



At the end I combine the first part and the third to do the second exercise in which I have to show the graph of different essays in which there is the entailment part as well as the first exercise nut now the nodes represents different type of sentences i.e. there are the premise that are the green node, the claims are yellow and the major claim is red. In that part the main function is called “caricaTema” at which I pass the number of the essay and it takes its json and text file to build respectively the graph and the text with the annotated essays. The default graph showed is that of “essay01”. An example:



Then as well as the part of the “annotated speeches”, with the text loaded in the function “caricaTema” and using the function called “tema\_sottolineato”, I have created the text in which the red sentence is the major claim, the yellow sentences are the claims and the green ones are the premises. In the following image an example of this last part:



CONCLUSION

The example of this little application about the visualization of data mining, simply shows the significant role that a system can have in showing different type of information in a clear and simple way. An important aspect is to get right data and in a standard format to allow and simplify the analysis of data and their processing. In the future with the help of artificial intelligence and the machine learning all this data can be also used to reason about some topic and to get some suggest about a solution of some problem or only an opinion about the topic based on the large amount of web information. As seen in class, already there are systems that work in this way an example is “The Debater” of IBM that analyses a lot of documents from Wikipedia and reason using these documents to show interesting information about that topic.