**Arabic Tagging Project**

3rd meeting – 25.02.2020

**Dictionary**

Didn’t find API for لسان العرب for now we found a downloadable dictionary الوسيط which is saved in MySQL it has around 30k terms.

Ali is checking out a different , better (bigger) dictionary and should get back to us with it.

**User interface**

for the next iteration the website should include 2 main pages which are

1. Tags page – display all saved tags in their tree (in order for the user to be able to see the hierarchical pattern) allow the user to add new children to current nodes, we didn’t talk about deleting or editing current nodes but that should also be an option (might be able to fix a typo or a previous mistake).

Don’t forget to implement checks when adding nodes, the tree must remain a tree and and not become a graph at some point (don’t allow loops... )

1. Tagging page – display the entered document (the whole text or poem), when the use hovers over a word the system should display it’s pre tagged tags (the it’s definition from dictionary), using right-click the user will be able to choose a tag or tags for the current word.

Interaction -> during tagging if the user doesn’t find a suitable tag he’ll have the option to “Add new tag” which takes him to the tags page where he can choose where to add it in the tree and after saving the user should be brought back to the tagging page exactly where he left and to select the new tag for the work he was on.

**Ideas for future iterations**

* **Roots** – words in Arabic are strongly connected to their roots (שורש של המילה); i.e. words with same root are likely to have same tags, we can expand our database to save a root and all the words that are it’s, this way when tagging a word that another work with the same root already has a tag we’ll be able to suggest it (or even tag it) 🡪 מילון מבוסס על שורשים

Note - there are a few ways of getting the root of a word online and fast so if we build the dictionary now without roots when we want to them it will be an easy job.

* **Sentiment Analysis** – when analyzing the document we can run sentiment analysis on each sentence or phrase and save the results, Heyam probably has a dictionary or system for sentiment analysis that we can use.