**Arabic Tagging Project**

9th meeting – 19.05.2020

**What we did**

What we did, we worked on preprocessing terms before we even add them to our database, the processes we make are:

* Tokenization
* Stemming and lemmatization
* Root
* Form – בניין
* Proclitic
* Enclitics

Ali – Since we take a term, put all حركات on it we sometimes end up with “legal” terms when it comes to grammar but illegal when it comes to Arabic dictionary… maybe instead of building a new pre processing method that might give us wrong terms we use a pre built one even if it’ll cost some money.

Ilai – here we have some terms are legal when it comes to grammar but not in the Arabic dictionary, and poets tend to “come up” with new terms in their poems to make it sound good, so trying different terms that might not be legal may actually end up collecting these terms.

Also + there’s רמת דיוק for each term we come up with so terms that are not in the dictionary we can give them lower grade and have them at the end.

About where to put this analysis:

In UI we have the poem, whenever we press on a term as we said before we’ll open the option to tag it – if there are previous tags of this term suggest them + the term’s definition next to that add a button whenever pressed will open this analysis.

**For next time**

Show the UI of how to tag if in PowerPoint or in the actual project, to make sure its easy to use for the user. And continue the preprocessing of terms.