**Arabic Tagging Project**

15th meeting **– 08.**0**9.**2020

**Users options**

1. Admin – can do everything there is to do and add new users and guests.
2. User – can do everything there is to do in poems and analysis.
3. Guest – can view the data and nothing more.

**Notes to fix**

When deleting a tag in the manage tags page – turn into two options (1) delete it but keep the children (2) delete it and all it’s children. Instead of having the delete and add in the bottom of the hierarchy.

**Backup mechanism –** make sure we have some sort of backup once every two weeks – months, before checking your own thing ask Yaroon maybe the university server already has one.

**Matrix**

Concurrence matrix, get all terms you have in all poems, put them in a matrix, now go over each line, when two terms are together in a line add +1 in the matrix. This way we have number of times each two terms are together in the same row.

Two terms that occur with the same terms 🡪 they’re similar.

Process terms before adding them to the matrix, (1) remove tashkeel (2) keep stop words (they is a small number of stop words they wont affect the size of the matrix & stops words can change the whole meaning of the metaphor) (3) bi-grams tri-grams and so on… ?

Lucin – apache 🡪 free search engine, take it, load all poems, it builds and indexes and allows us to search within the data.

**What we need to do:**

* Connect login.
* Create 3 users options.
* Small matrix to test.
* Add d3 graph of the tags in manage tags page.
* Add pre-processing step (before tagging stage)
* Connect a dictionary.
* Choose a name for the website.
* Add menu to manage tags (delete, edit, add, delete all, add parent…)

For next meeting we will have

1. Demo (or all) of matrix.
2. Menu to manage tags.
3. D3 graph of the tags in manage tags page.
4. Connect login and create 3 user options.