**Arabic Tagging Project**

17th meeting – 14**.**10**.**2020

**What we did**

Talked to Yaroon, we choose a machine for the server, it’ll be ready next week +-

Added the option of login and logout, the website now includes 3 types of users: super user, staff, and regular user. Only the super user can add new usres and needs specify which type of user he is staff or not staff (regular user = guest).

The initial website only shows the introduction to the user, only users that log in can view the poems and the terms and everything.

**Database**

In Fozi’s website, almost everyday a new poem is added to the database, we need to add a sync every week and with a button in setting which does the load.

**Ideas**

Ilai got a book: a guide to classification, it helps the reader to classify arabic poems with examples, moreover it teaches how to infer what the writer meant in each line. The book/s:

* Shamy, Hasan M. El-,. (1995). Folk Traditions of the Arab World: A Guide to Motif Classification. Bloomington: Indiana University Press.
* Shamy, Hasan M. El-. (2004). Types of the Folktale in the Arab World: A Demographically Oriented Tale-Type Index. Indiana University Press.
* Shamy, Hasan M. El-. (2013/2016). Motific Constituents of Arab-Islamic Folk Traditions. A Cognitive Systemic Approach. Volume I, (A-K). Indiana University Press.

Ali: another book, which collects a number of metaphores in classic arabic poetry, but it’s a book and not a pdf or word so wont really help us.

**Word embedding**

Amjad’s helps us to understand for pre selected terms how much they “weight” in emotions, we want to see frequency of words that point to an emotion that occure with out pre selected,

להבין את המטען הרגשי על מילים נבחנות במונח הרגשע.

What Osama did 🡪 looking for similarity this way seems good, got a poem and lowered each term’s dimention using PCA from 300 to 2, where

מילים צמודות מבחינם הקטסט אבל רחוקות מבחינת ההקשר.

The terms we have from the all the poems in our system is not enough to “train” our corpose, we need more data from places like wikipidia or other website or collection of terms which has similar arabic terms meaning written not spoken arabic, the bigger number of different terms the better.

Better than wikipedia we may be able to user the text from المكتبة الشاملة which has a collection of books – machine readable – mainly from the same time period as the poems we have which means the same writing methods.

Another option is a website called الوراق which also has a collection of books in machine readable format.

**Opposites**

When the system now finds opposites we don’t fully trust it, theres another step where a human goes over the opposites and checks which are right and wrong.

In spesific poems there is a different kind of opposite – تروية – where the machine will probably find it as an opposite but really it should be opposite + تروية here the human selection adds this extra feature.

**Name entity recognition -** will it help? Its an automatic recignition which gives us יישויות for a term. Its not critical, doesn’t help much.

**Emotional dictionary** with 90 languages, might help us in the automatic part - <http://sentiment.nrc.ca/lexicons-for-research/>

**Expanding the tags ontology.**