**Final Project**

1st meeting – 10.02.2020

**מערכת לתיוג מילים**

* **לייצג אונטולוגיה של מילים בערבית**
* **לזהות סתירה במשפטים למשל שתי מילות עוקבות שתיוג שלהם סותרים אחד את השני.**
* **זיהוי מטפורה.**

**מטרת על 🡨 מילון מתויג ("תזאורוס") שיכיל**

* **היררכיה מושגית (מיון/סיווג)**
* **קשר של מילה נרדפת.**
* **קשר של מילה נגדית.**
* **קשר של** association **, למשל שחור מסמן חושך.**

Flow of how it's supposed to work for now

1. Read the added document (word, pdf...)
2. Tokenization.
3. Show a line from the text with option of selecting a specific term/word.
4. Suggest tags for selected word based on
   1. Same word was tagged before - suggest the saved tag (& dictionary if it's different?)
   2. New word - suggest based on dictionary.
5. Save in Database.

Needed data stores

1. All documents.
2. All terms and all taggs with the connections. (Graph)
3. All documents where in each document we connect the term to the choosen tag (because a term can have different tags based on context which can only be understood from the document, we don’t want to lose the connection between tag and context.)

Choosen dictionary - <https://www.almaany.com/>

Example of an ontology (can even take the exact one translate and add/remove things to make it good enough for arabic) - <https://ids.clld.org/>