Learning to generate text from knowledge graphs with seq2seq models
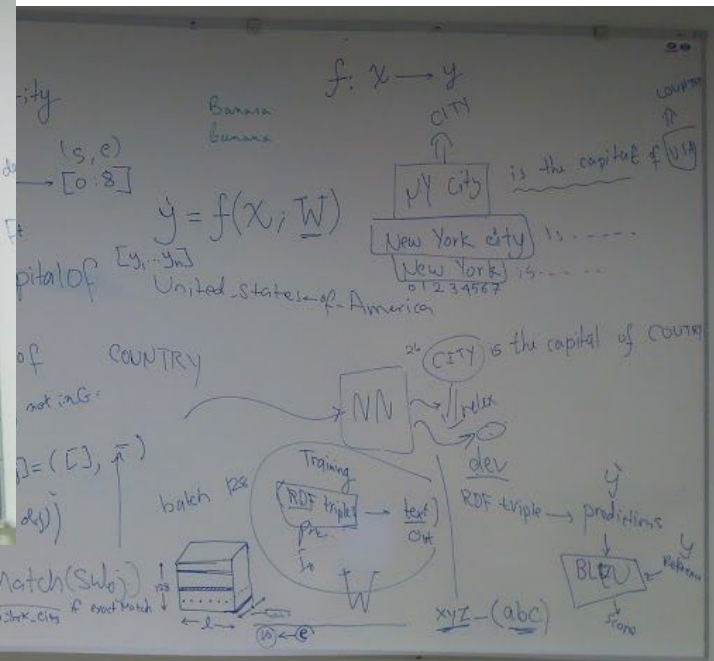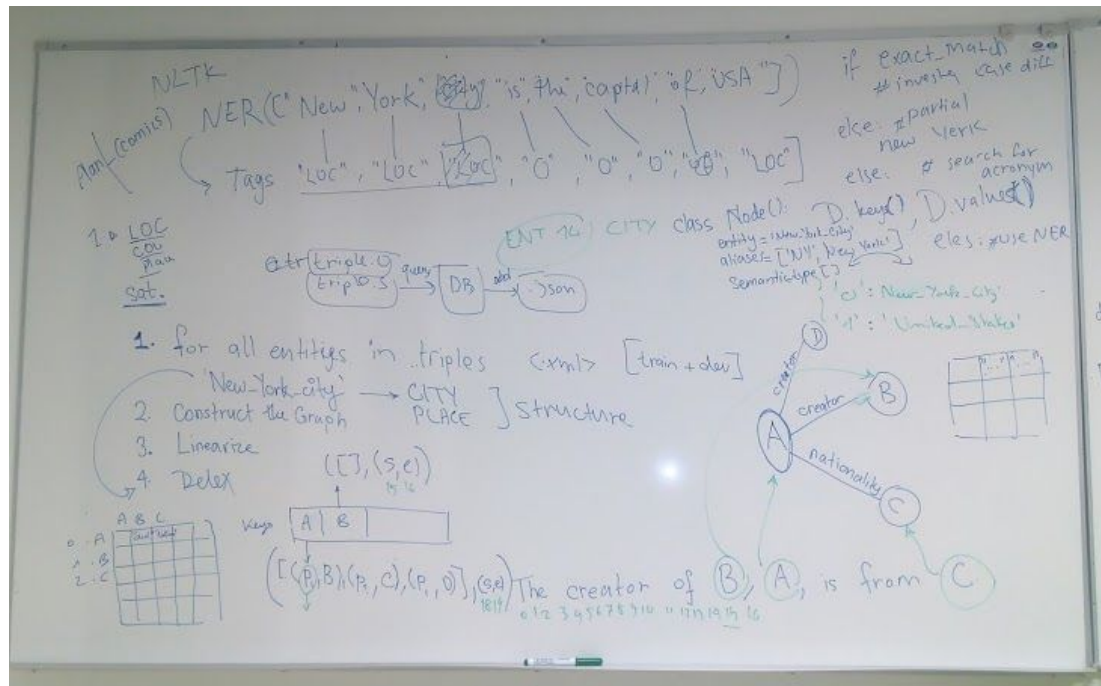
Reem, François, Badr

Back to the Blackboard

Progress

- We built our own preprocessing pipeline (almost from scratch)
  - Modular design
  - XML parsing, RDF instances representation, and entity/graph representation
  - Different ways of graph linearization (flat vs structured)
- Work to be done
  - SPARQL module to communicate with DBpedia
  - Implement sentence de-lexicalization (with text matching)

Example

- Input seq

$^1($
    $^2($ ENTITY-1 AGENT
        $^3($ leaderName ^(ENTITY-2 PATIENT)^ $)^3$ $)^2$
    $^2($ ENTITY-3 AGENT
        $^3($ country ^(ENTITY-1 PATIENT)^ $)^3$
        $^3($ location ^(ENTITY-4 PATIENT)^ $)^3$ $)^2$
  $)^1$

- Output seq

Ethiopia is lead by Hailemariam Desalegn and is the location of the Addis Ababa City Hall in Addis Ababa.

Experiments Plan

- Investigate neural architectures choices (smaller models may perform better)
- Investigate the impact of specific vs generic semantic types
- Investigate the impact of graph representation (flat vs structured)
- For each experiment, we are planning to run evaluation with different evaluation metrics + error analysis

Questions

- Why are there 8 different reference files?