Goal ⇒ What kind of knowledge is needed for the long tail disambiguation task?

• Questions (30 mins):
  ○ How different are the typical NLP lexical resources from Semantic Web’s knowledge bases? (e.g. WordNet vs DBpedia?)
  ○ How can we define the long tail in knowledge bases?
  ○ For text disambiguation, do we need knowledge beyond what is available in customary sources (e.g. DBpedia and WordNet)? And should it be different or the same for each challenge?
  ○ How can we locate the right knowledge to use for a given context, e.g. topic or time?
  ○ Is there sufficient knowledge? List of entities and what we know about them. How to get it?
  ○ What is the place of the temporal dimension in knowledge bases?
    ■ If, when and how to deprecate out-dated knowledge?
    ■ How to represent historical changes?
    ■ How to deal with concept drift and what is it?
      • Concept drift in NLP is defined through distributional semantics, in SemWeb is defined through: label + key properties
  ○ How to enrich knowledge bases with more long tail knowledge?
  ○ Linked Open Data versus Linked Closed Data (BabelNet, Cornetto).
  ○ Should we address the quality of resources?

• Hands-on data (90 mins):
  ○ Analyze WordNet and DBpedia WRT the long tail
    ■ DBpedia:
      • Online SPARQL
      • TSV with stats: degree/rank, #types, num attributes, ...
      • fragments.dbpedia.org to analyze the temporal change in knowledge bases (e.g. how much has the information about Abraham Lincoln/Cristiano Ronaldo been updated in across the DBpedia versions?)
    ■ BabelNet
      ○ Analyze documents from the laundromat that are likely to represent the long tail (e.g. documents with low connectedness degree)
      ○ Locate domain-specific knowledge (e.g. through LOTUS)

• Concluding the session+recommendations (30 mins):
  ○ How to create knowledge bases that represent the long tail?
  ○ How to create contextual knowledge bases (e.g. WRT time, space, topic, community)?
  ○ How to locate the most appropriate knowledge to use?

• Preparing presentations (15 mins)