EDAN70
Project in Computer Science
http://cs.lth.se/edan70/
Project in Intelligent Systems

Pierre Nugues

Lund University
Pierre.Nugues@cs.lth.se
http://cs.lth.se/pierre_nugues/

March 27, 2019
Projects

● Define a study topic and an application in machine learning or language processing. You may define them yourself or with the help of your instructor.

● Do a quick survey the relevant literature. Start from one or two papers.

● Select datasets, algorithms, and outline an implementation strategy and timeline

● Iterate:
  ● Implement a prototype. This should be started the first week.
  ● Evaluate it

● Write a project report. This should be started the first week and shared through sharlatex (with Latex and Bibtex) using a conference template (ACL 2019)

● Release your code (optional)

● Submit paper to a conference (optional). No funding for conference fees and travel.
The project will take place in the 4th LP. There is no dedicated location for it. The participants will work on the machines in the basement or on their own machines. The duration of time spent on the project should be of about 200 hours. Each participant can work alone or collaborate with one or two other people (preferred). Weekly progress meeting with your instructor
Possible Subjects

1. Projects building on the assignments:
   1. Extend and improve logistic regression. Program a neural network;
   2. Choose a dataset from Kaggle and participate in the competition.

2. Projects connected to research:
   - Knowledge graph in Swedish, possibly English or other languages:
     1. Develop simple extraction tools starting from regexes and dependency parsing to build a knowledge graph from wikipedia.
   - Entity linking:
     1. Apply a transformer architecture to named entity recognition. Use the CoNLL dataset;
   - Translation:
     1. Apply a sequence-to-sequence or transformer algorithm to build a translator
Dataset provided by Sonja Aits:

1. Find chemicals (drugs, toxins) that affect lysosomal damage and cell death positively or negatively (these could have therapeutic applications but also be risk factors for degenerative disorders and cancer)

Projects connected to research in medicine (Kaggle datasets):

1. Cervical Cancer Risk Classification
2. Breast Cancer Wisconsin (Diagnostic)
3. Personalized Medicine: Redefining Cancer Treatment

NIST/TAC dataset on:

1. Drug-Drug Interaction Extraction from Drug Labels (DDI)