Assignment 4

Due: 11/18/2015 (5 extra credits before 11/8/2015)

Go to “Annotation tool” in

http://blender02.cs.rpi.edu:3300/elisa_ie

1. For the following students who know foreign languages, please select news documents about natural disaster, in total about 20,000 word tokens.

   Marathi: Shivani Ginde
   Hindi: Akshay Matta
   German: Kevin Stoner
   French: Austin Ralls
   Hebrew: Jonathan Beekman
   Spanish: Brian Sheedy
   Spanish: Stephen Gross
   Spanish: Zachary Vanderzee
   Korean: Yi Lu
   Swedish: Chris Marsh
   Japanese: Frank Giacoio III

Other students please select tweet messages or facebook messages about natural disaster, in total about 20,000 word tokens.

2. Tag entity names, provide translations and English Wikipedia titles for each document.

3. List the possible ideas for name translation (in English social media case, it’s ‘name normalization’).

4. (5 extra credits) Implement a name translation or name normalization system.