

XV UNL School, 21-25 July 2014, Geneva

Exercise #3 – Word list

- Goal: To prepare the word list from the training corpus
- Deliverable:
 - *wordlist_<ID>.txt*
- Activities:
 1. Download the Default Dictionary from the UNLARIUM
(UNLWEB>UNLARIUM>DICTIONARY>[YOUR LANGUAGE]>EXPORT>DEFAULT DICTIONARY)
 2. Upload the Default Dictionary to the UTK
(UNLWEB>UNLDEV>UTK>CORPUS>DICTIONARY>ADD)
 3. Extract the word list from your corpus:
 - a. UNLWEB>UNLDEV>UTK>CORPUS>CORPUS>LOAD <corpus_org>
 - b. UNLWEB>UNLDEV>UTK>CORPUS>DICTIONARY>LOAD <DEFAULT DICTIONARY>
 - c. UNLWEB>UNLDEV>UTK>CORPUS>PROCESS>EXTRACT LISTS (TYPE =TOKEN,
ORDER=FREQUENCY DESC, DETERMINISTIC=TRUE)
 4. Export the word list