

Terminology

domain: database type,

– e.g. nante, classifieds, pc110 rfr

token: original white space separated token from text file

split token: tokens split by splitter, may not be white space separated

tag: NSW tag

– NUM, LSEQ EXPN.

word: alphabetic character string that is in lexicon or can be pronounced by LTS system.

Architecture

