**British National Corpus Full Word Count**

**CMIST**

**University of Manchester**

**John Mcloughlin**

**September 2016**

Extract from corpus via xsltproc

Split into 50 files for loading into NVIVO. NVIVO can’t load full corpus. Have to split.

Load fifty files into NVIVO.

NVIVO cannot export more than 65536 rows to an Excel file. Need to run word count query on three files at a time to get export under this limit.

Save query output as 15 csv files from Excel.

Load CSV files into SQLite database table.

Run word count query on table to generate output. Save as CSV file.

Problem: Running the word count query on only three files at a time means that the full corpus cannot be compared all at once. This is a problem caused by NVIVO’s inability to export more that 65536 rows and inability to run the query on all of the corpus in one go.

To create bar charts, use grep ( Linux/ Mac ) or FINDSTR ( Windows ) to extract the required word from the CSV file. EG

grep teach BNC\_Full\_word\_count.csv > teach.csv

This will extract values for all words related to the stem ‘teach’ and create a new CSV file, teach.csv. Load this into Excel and create a bar chart. Save as BNC\_Full\_teach\_bar.xsls