



# World cloud procedure

- Find a text
- Do lemmatization

# lemmatization

- <https://lindat.mff.cuni.cz/services/morphodita/>

comments and reactions are welcome.

**Model:**  Czech MorfFlex2+PDT-C  Czech MorfFlex+PDT  English  Slovak

english-morphium-wsj-140407

**Task:**  Tag  Lemmatize  Analyze  Generate  Tokenize

Use morphological guesser for unknown words

**Input:**  Plain text  Vertical (word per line)

**Tag set:**  Original  Strip negative prefix

**Output:**  Formatted  XML (format description)  Vertical (format description)

Input Text

Input File

we have contributed policy ideas on issues such as Global Britain, healthcare, education, and the digital age.

The Conservatives are the only party on the ballot paper in Sunderland this May to oppose a so-called People's Vote (i.e. a second referendum on our membership of the European Union). A majority of Sunderland, Washington and Houghton residents voted to leave, and Labour and the Lib Dems should (but don't) respect that.

Local Conservatives continue to be proud of much of what our party has achieved nationally:

- Proscribing the terrorist organisation Hezbollah.

← Insert text here

Process Input

# Count and clean in excel

- Copy table „output“ into excel
- Use subtotal to count the words
- Delete stop words
  - Words wchi do not make difference between clouds
- The, a, an, it, I, we, they, their, have, for, in,...
  
- Save as csv with names:
- weight      word   color   url
  - Weight is count, color and url may be empty





- The cloud can be made from text without processing in excel
- Low control over stopwords
- Do not use crazy colors shapes and text orientations

# Manifesto orientation





