# Advanced features of Sketch Engine

Vojtěch Kovář, Vít Suchomel



name.surname@sketchengine.co.uk

6<sup>th</sup> Sketch Engine Workshop Herstmonceux, August 10, 2015

## Advanced features of Sketch Engine

 $Underused\ functionality. \ . \ .$ 

you might not be aware of

### Word Sketch definition tricks

### Global attributes in CQL

- 1.number = 2.number ... tokens #1 and #2 have the same number
- 1.id = 2.ref ...token #2 is dependent on #1

m4 macros to make the word sketch definition better readable

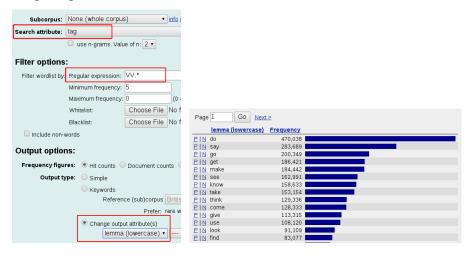
- A noun define('noun', '[tag="N.\*"]')
- An adjective define('adj', '[tag="JJ.?"]')
- Agreement of words in gender, number and case (Slavic languages) define('agree', '\$1.gender=\$2.gender &
  - \$1.number=\$2.number & \$1.case=\$2.case')

#### Combined

- Two adjectives and a noun in agreement
  - 3:adj 2:adj 1:noun & agree(1,2) & agree(1,3)
- Better readable than the same expanded
  - 3:[tag="JJ.?"] 2:[tag="JJ.?"] 1:[tag="N.\*"] &
  - \$1.gender=\$2.gender & \$1.number=\$2.number &
  - \$1.case=\$2.case & \$1.gender=\$3.gender &
  - \$1.number=\$3.number & \$1.case=\$3.case

## Search attribute and output attribute in the wordlist form

### To get e.g. a list of lemmata of lexical verbs:



### WebBootCaT from Wikipedia seeds

Need a domain corpus but don't know the seed words? WebBootCaT can be initialised using keywords from Wikipedia articles:

Wikipedia URL	carbon di		
Additional Wikipedia URL	Carbon dioxide		
	Carbon disulfide		
	Carbon dioxide removal		
	Carbon dioxide poisoning		
Additional Wikipedia URL	Carbon dioxide in Earth's atmosphere		
	Carbon dioxide laser Carbon diet		
Additional Wikipedia URL	Carbon diet  Carbon dioxide equivalent		
	Carbon dioxide scrubber		
Out both dioxide both both			
effect, (9)	surface. (7)	glass, (5)	Cumulative (6)
years. (9)	Mitigation. (7)	Meyer. (5)	Retrieved (7)
Change. (9)	acidification (9)	(at (5)	Yamaguchi (5)
effect. (8)	Contribution (14)	air, (5)	pre-industrial (5)
countries. (8)	ppm (19)	Rogner, (5)	Climate (42)
(see (8)	greenhouse (6)	pollution, (5)	temperature, (4)
emissions, (8)	Metz, (6)	pollution. (5)	effect" (4)
< Back Use WebBootCaT with selected words			

Term extraction of seedwrods from Wikipedia articles will be added in the future.