

Corpus manager Manatee / Bonito

- Searching in corpora according to all attributes or their combinations (regular expressions)
- Sorting and filtering results
- Basic statistics
- Graphical overview of occurrences within the whole corpus
- Display of more attributes (word, lemma, different types of tags, ...)
- Creating subcorpora

Searching

- most common attributes:
 - word
 - lemma
 - morphological tag
 - (analytic functors)
- Regular expressions
- It is possible to combine the attributes

```
[tag="R.+"][lemma="Praha.*"]
```

Displaying

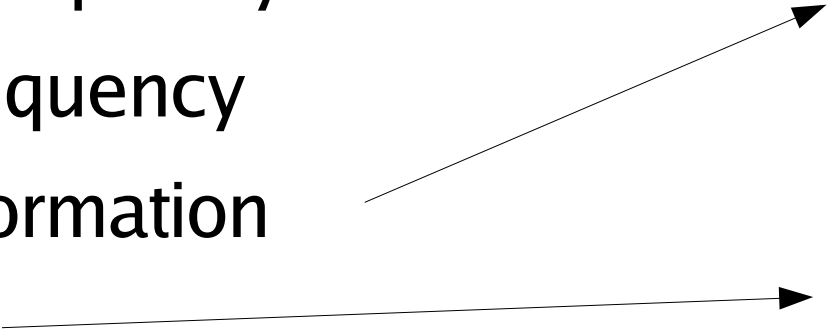
- Selection of attributes
 - for KWIC
 - for all positions
- Changing range - number of concordances
 - from beginning / end / middle
 - randomly
- Information about the source of the text
- Changing context
- Wider context in extra region

Sorting and Filtering

- Sorting according to
 - kwic (key word in context)
 - left or right context
 - combination
- Filters
 - positive
 - negative

Statistics

- Frequency distribution
 - according to selected attributes
- Collocations
 - absolute frequency
 - relative frequency
 - Mutual information
 - t- score
 - distribution overview + Average Reduced Frequency


$$mi(x, y) = \log_2 \frac{N \cdot f(x, y)}{f(x) \cdot f(y)}$$

$$T = \frac{\left(f(x, y) - \frac{f(x) \cdot f(y)}{N} \right)}{\sqrt{f(x, y)}}$$

$$ARF = 1/v \sum_{i=1}^f \min\{d_i, v\}$$