

Minor enhancements in the user interface

Vít Suchomel, Vojtěch Kovář

Lexical Computing Ltd.

`{vit.suchomel, vojtech.kovar}@sketchengine.co.uk`

SKEW 3

March 21, 2012

Outline

- 1 Corpus Architect
- 2 Bonito

- new icons for configuration templates in the 2nd step of the corpus creation form which indicate availability of associated functions and features
- generate queries without spaces when using BootCaT for languages which do not put spaces between words
- uploading archive files with multiple documents
- pagination for comfortable corpus files listing
- boilerplate removal using jusText (replaced BTE)
- option of de-duplicating the corpus with onion (before compilation)

Language specific tools

- added preloaded sketch grammar for Japanese tokenizer ChaSen
- corrected matching patterns in Arabic tokenizer

- enabled sharing customized preloaded corpora (CPC) with other users
- improved user lookup by pattern (when creating user groups or sharing stuff with other users) so that better matches are positioned higher in the list
- added support for SSO authentication (generalized with IP-based authentication as "anonymous authentication")
- displaying Configuration template in My corpora table

Bonito enhancements

- all presented currently installed on beta
- (<http://beta.sketchengine.co.uk>)
- user interface language switch
- relative frequency figures
- wrapping lines in wide context according to structure markup
- displaying structures in wide context
- extended options of the frequency form
- proper filenames for results saved from the interface
- jump to a letter if the concordance is sorted

- keyword as a part of gramrel name – %w
- unstructured word sketch
- filtering selected gramrels
- collocation sorting switch
- structured/unstructured switch
- thesaurus clustering switch

For discussion

- new concordance query form proposal
- http://nlp.fi.muni.cz/~xbaisa/skeint/first_form.html