Sketch Engine after Adam Kilgarriff

Miloš Jakubíček



milos.jakubicek@sketchengine.co.uk

6th Sketch Engine Workshop Herstmonceux, August 10, 2015

Sketch Engine as of 2015

$$+12$$
 -1

Sketch Engine as of 2015

- (2003) two-man show: Adam Kilgarriff and Pavel Rychlý
- (2015) still a very small company 10 people growing steadily
- since 2011: two companies
 - UK Ltd. (Brighton)
 - CZ subsidiary (Brno)
- rather B2B than B2C

Sketch Engine Data

By August 2015 more than 400 corpora for 82 languages:

- 100+ corpora having more than 100 million tokens
- 30+ corpora having more than 1 billion tokens
 - In 2010 a series of TenTen (10¹⁰) corpora started
- 60+ languages with a PoS-tagged corpus
- 42 languages with word sketches
- 26 languages with integrated tagger for tagging user corpora
- parallel corpora: EUROPARL, DGT, OPUS, ...

Sketch Engine Users

- Lexicographers
- Researchers
- Teachers
- Language Learners
- Translators
- Terminologists
- Copywriters

Sketch Engine Marketing

- Very many trials
- Very many failures
- I do not understand why it's existence is not generally known by linguists ???

I personnally know it by a very young Spanish colleague . . . and i was immediately convinced.

None of my belgian colleagues from my and other universities I presented results restreived with it did ever hear about it but they were all amazed by its potential. Was mention of it made in Academia, Research Gate, Linguists' list?

(An e-mail from a French university professor, Feb 2015)

Sketch Engine Marketing

- Selling Sketch Engine is simply different
- We do not need a salesman
- We do need ambassadors and advocates in the Sketch Engine community
- ⇒ (a past, present and future) Thank you!

Sketch Engine Ambassadors Programme

Spreading the word about Sketch Engine

- mainly via workshops
- supported by us (content, workshop licences, flyers, ...)

Sketch Engine Future: Legal

- $\bullet \ \, \mathsf{status} \,\, \mathsf{quo} \colon \, \mathsf{AK}/\mathsf{GL} \, \leftarrow \, \mathsf{LCL} \, \leftarrow \, \mathsf{LCC}$
- future status: $MJ \leftarrow LCC \leftarrow LCL$
- no immediate changes to any customers

Sketch Engine Future: Research agenda

https://www.sketchengine.co.uk/research-agenda/

- Parallel and Distributed Processing of Very Large Text Corpora
- Building Very Large Text Corpora from the Web
- Corpus Heterogeneity and Homogeneity
- Corpus Evaluation
- Terminology Extraction
- Corpora and Language Teaching
- Language change over time

Conclusions

- Sketch Engine: up and running even in difficult conditions
- Some administrative changes ahead
- Apart of that ...

Conclusions



Conclusions

