

# Management of long-running jobs

Miloš Jakubíček



`milos.jakubicek@sketchengine.co.uk`

5<sup>th</sup> Sketch Engine Workshop  
Bolzano, July 14, 2014

# Corpus querying

- should be fast
- trade-off between compile-time and run-time computations
- sometimes compile-time computations not possible (e.g. user subcorpora)
- $\Rightarrow$  there will always be long-running jobs

# Current status quo

- processes spawn by Bonito to run in background
- no option to run the job on another server
- no management options for user or administrators
- often cumbersome to track down related issues

# New approach

- a separate jobserver (or cluster of these) responsible for executing and managing long-running jobs providing:
  - progress reporting
  - completion notification
  - jobs overview for users
  - jobs management for administrators

Sketch Engine server



Job servers cluster



job spawning  
job reporting

user request and response



user

# User experience & Conclusions

Long-running jobs that

- provide more reliable information about estimated time of completion
- provide notification
- are executed on multiple servers and hence are faster