

Digging in the Dirt: Extracting Keyphrases from Texts with KD

Giovanni Moretti, Rachele Sprugnoli, Sara Tonelli

CENTER FOR INFORMATION TECHNOLOGY
DIGITAL HUMANITIES



KD = KEYPHRASE DIGGER

Keyphrase Digger is a rule-based system for keyphrase extraction. It is a **Java** re-implementation of KX tool (Pianta and Tonelli, 2010) with a **new architecture** and **new features**.

It combines **statistical measures** with **linguistic information** given by PoS patterns to identify and extract weighted keyphrases from texts.

MAIN FEATURES:

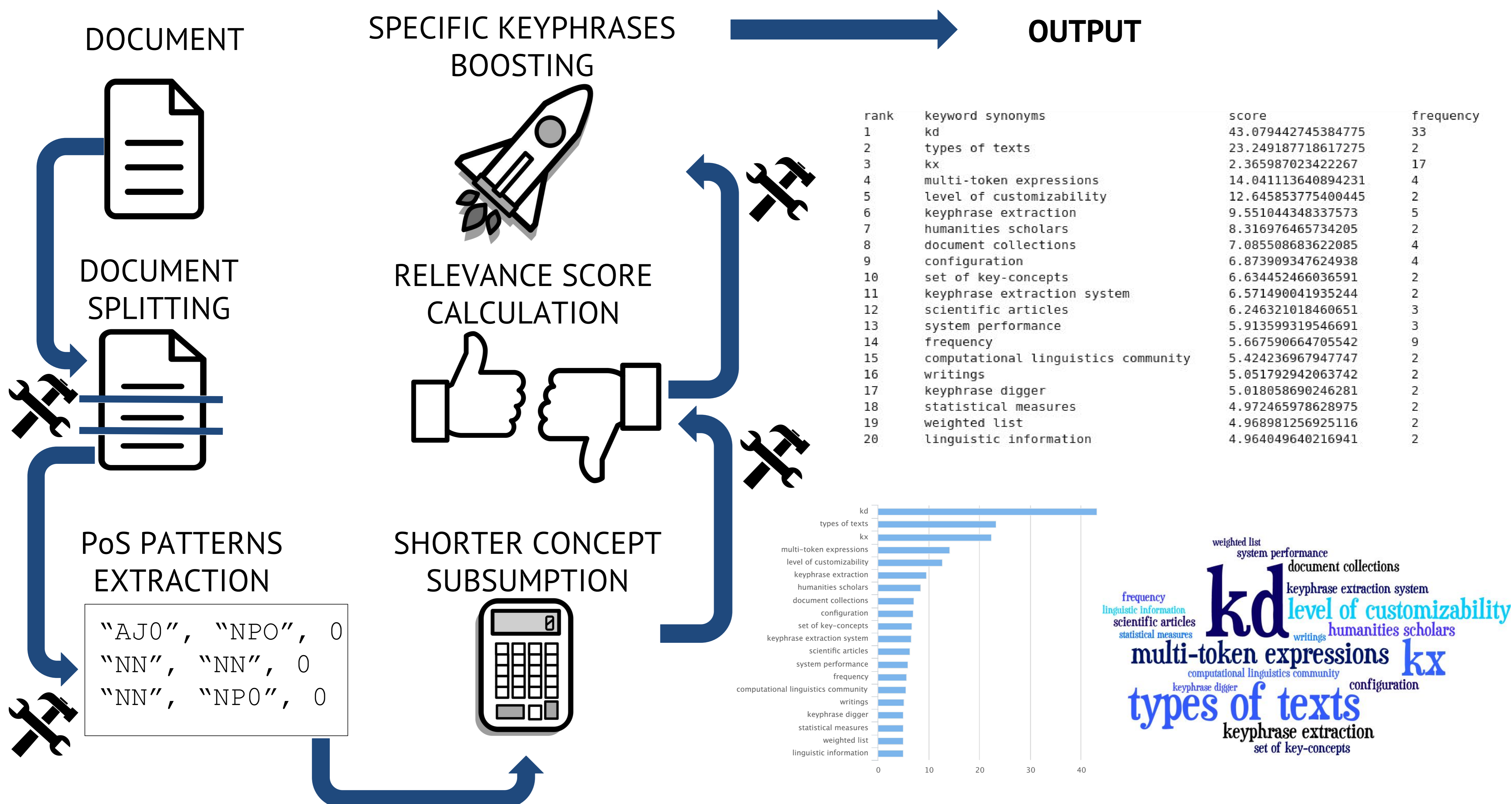
- Extraction of multi-words
- Multilinguality (EN and IT)
- Easily extendible to other languages
- Higher customizability than KX
- High processing speed
- Clustering of keyphrases under the same lemma
- Various accepted formats and PoS tagsets
- Boost of specific PoS patterns
- Integration of Apache Lucene Library



NEW!

KD is about 25 times faster than KX:
very apt for web applications!

KD ARCHITECTURE



KD AVAILABILITY

ONLINE DEMO

http://celct.fbk.eu:8080/KD_KeyDigger/

JAVA RUNNABLE JAR and LIBRARY

<https://dh.fbk.eu/technologies/kd>



<http://dh.fbk.eu>
https://twitter.com/DH_FBK



moretti@fbk.eu
sprugnoli@fbk.eu
satonelli@fbk.eu

