

Metallurgy

Jakub Kříž, Martin Lipták
Tomáš Kramár, Marián Šimko

Metallurgy

- text metadata extraction service
- <http://metallurgy.fiit.stuba.sk>

Current features

- **input**
 - url
 - text
- **preprocessing**
 - main readable part extraction (Readability)
 - strip HTML tags
 - stemming (Porter)
- **services**
 - content extraction
 - token extraction
 - keyword extraction (TF/IDF)
- **simple english corpus**

Planned features

- categorized english corpus
 - categorized slovak corpus
 - lemmatization
 - category detection
 - language detection
-
- **ideas ???**

Metallurgy

Built by nerds for nerds.