

PLIN064 Úvod do *digital humanities*

Zuzana Nevěřilová
xpopelk@fi.muni.cz

Centrum zpracování přirozeného jazyka, B203
Fakulta informatiky, Masarykova univerzita

4. prosince 2019

WebLicht

webová služba CLARIN-D, která nabízí mnoho nástrojů (pro mnoho jazyků)

uživatel zadá vstup (soubor, copy&paste, CQL, ...)

uživatel sestaví řetězec zpracování (processing chain)

výhody: je možné uložit řetězec zpracování i vstupní data → reproducibilita výsledku

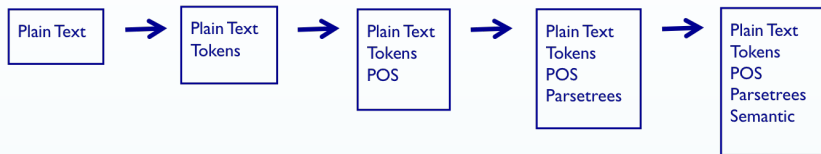
(související téma: jak citovat korpus? jak zajistit, aby byl korpus stabilní, a tudíž výsledek reprodučibilní?)

WebLicht: Jak používat?

Jednotlivé nástroje vyžadují vstupní jazyk a formát, poskytují výstup (a mohou být vstupy jiných nástrojů).

Používané vstupní formáty: plain text, CQL, TCF

WebLicht: Jak vytvářet řetězce zpracování



Text Corpus Format (TCF):

- XML speciálně navržené pro WebLicht
- interoperabilita nástrojů
- obsahuje lingvistické anotace (na libovolné úrovni)

příklad: `<https://weblight.sfs.uni-tuebingen.de/weblightwiki/index.php/The_TCF_Format>`

Vyzkoušejte sami na libovolný vstup.

Tipy:

- nebojte se používat i testovací nástroje
- zkuste různé vstupy
- zkuste různá pořadí komponent řetězce

```
<https://weblight.sfs.uni-tuebingen.de/weblight/>
```

