Seminarium Zespołu Podstaw Sztucznej Inteligencji
Warszawa, 18 kwietnia 2013 r.

Instytut Podstaw Informatyki Polskiej Akademii Nauk

Niezależny językowo system ekstrakcji informacji z pół-ustrukturalizowanych danych tekstowych (SEI BigGramy)

Marcin Michał Mirończuk[(1)]
[(1)]IPI PAN, m.marcinmichal@gmail.com

Streszczenie

Podczas prezentacji zostanie przedstawiony konstruowany w IPI PAN w ramach projektu NEKST system do ekstrakcji informacji z pół-ustrukturalizowanych danych tekstowych w postaci dokumentów internetowych (stron html). Prezentacja będzie miała charakter poglądowy, wprowadzający do konstruowanego systemu ekstrakcji informacji (SEI BigGramy). Zostanie w niej zaprezentowany przeglądowy zarys metod i podejść do ekstrakcji informacji. W dalszej kolejności zostaną wprowadzone bazowe pojęcia związane z konstruowanym systemem SEI BigGramy. Następnie zostanie opisany poglądowy proces ekstrakcji informacji, który jest aktualnie zaimplementowany na klastrze. Na końcu zostaną przedstawione otrzymane dane eksperymentalne z procesu ekstrakcji informacji oraz opisany zostanie zarys problemów jakie powstają przy proponowanej metodzie ekstrakcji informacji wraz z ich rozwiązaniami. Na końcu wystąpienia autor podsumuje aktualnie przeprowadzone eksperymenty związane z SEI BigGramy.

.