Biuletyn
Polonistyczny

06.01.2016
Data wydarzenia: 11.01.2016 g.13:00

Referat "Ekstrakcja terminologii z tekstów w języku polskim – program TermoPL" (Małgorzata Marciniak, Agnieszka Mykowiecka, Piotr Rychlik)

Typ wydarzenia: Zebranie
Organizatorzy: Instytut Podstaw Informatyki Polskiej Akademii Nauk PAN

Referat zostanie wygłoszony w ramach połączonego Seminarium Instytutu Podstaw Informatyki PAN i seminarium „Przetwarzanie języka naturalnego” 11.01.2016 r. o godz. 13.00 w siedzibie Instytutu przy ul. Jana Kazimierza 5 w Warszawie.

Podczas seminarium omówione zostaną zagadnienia ekstrakcji terminologii z korpusów tekstów w języku polskim oraz przedstawią program TermoPL, w którym zaimplementowane zostały zaproponowane rozwiązania. W omawianych pracach stosują oni metodę C-value, która ustala istotność terminów na podstawie ich częstości i liczby kontekstów. Metoda ta uwzględnia terminy zagnieżdżone – czyli takie, które są zawarte w dłuższych terminach. Niekiedy prowadzi to do ekstrakcji niepoprawnych semantycznie fraz (np. 'giełda papierów', 'USG jamy'). W celu eliminacji tego typu terminów autorzy referatu zaproponowali metodę, która w procesie wyodrębniania zagnieżdżonych fraz kieruje się siłą powiązań między słowami.

Linki:

- http://zil.ipipan.waw.pl/seminarium

Słowa kluczowe: terminy zagnieżdżone, informatyka, korpusy tekstów, terminologia