
Rok akademicki 2004/2005

Seminarium magisterskie

Narzędzia i metody przetwarzania tekstów

dr hab. **Janusz S. Bień**, prof. UW
dr **Krzysztof Szafran**

czwartki, godz. 12¹⁵–13⁴⁵

Instytut Informatyki UW, ul. Banacha 2 (wejście od ul. Pasteura)
sala 4030 [ZMIANA!] (trzecie piętro)

Seminarium jest przeznaczone dla studentów informatyki zainteresowanych różnymi aspektami przetwarzania języka naturalnego, a także globalizacją i lokalizacją oprogramowania. Na indywidualnie ustalonych zasadach mogą w nim również uczestniczyć zainteresowani humaniści.

Tematyka seminarium jest bardzo szeroka, o czym świadczą już ukończone prace magisterskie (patrz np. <http://www.mimuw.edu.pl/~jsbien/> i <http://www.mimuw.edu.pl/~kszafran/>). Niepełny wykaz proponowanych tematów znajduje się pod adresem <http://www.mimuw.edu.pl/~jsbien/nmtp/>. Tam też znajdują się informacje o innych materiałach związanych z problematyką seminarium.

Obejmuje ona m.in. metody ścisłego opisu różnych aspektów języka naturalnego — w szczególności morfologii i składni języka polskiego — oraz wykorzystanie w praktyce dostępnych zasobów lingwistycznych np. do różnych form automatycznej analizy tekstów. Interesują nas również problemy tłumaczenia dokumentacji technicznej (na przykładzie *Linux Documentation Project*), narzędzia do wspomagania takiego tłumaczenia przez automatyczną analizę tekstów oryginalnych oraz porównywanie istniejących tłumaczeń (tzw. korpusy równoległe), tworzenie i uaktualnianie słowników terminologicznych itp.

W naturalny sposób do zakresu seminarium należy również problematyka reprezentacji tekstów, w tym wielojęzycznych (**XML**, **SGML**, **TEI**, **DocBook**) oraz narzędzi do ich edycji i składu, w szczególności konwersji między różnymi formatami (np. **WVware**, **LaTeX2html**).

Preferowane są tematy prac magisterskich prowadzące do powstania oprogramowania dostępnego na licencji GNU i nadającego się do włączenia do oficjalnych dystrybucji. Najbardziej pożądaną formą wyników programistycznych jest pakiet dystrybucyjny systemu Debian GNU/Linux, dostarczony na samoboootującej się płycie z odpowiednio zmodyfikowanym systemem **Knoppix** (<http://www.knoppix.org>) lub podobnym (np. **Morphix**).

Nie stawiamy uczestnikom żadnych konkretnych warunków wstępnych, ale w razie nadmiaru zgłoszeń preferowani są słuchacze **naszych własnych wykładów monograficznych** a także słuchacze wykładów *Inżynieria lingwistyczna*, *XML i nowoczesne technologie zarządzania treścią* oraz *Programowanie WWW*. Pożyteczna może być też znajomość języków **perl** i **Prolog**.

Uczestnicy seminarium są zobowiązani do subskrypcji zamkniętej listy pocztowej *nmtp-l* (<http://duch.mimuw.edu.pl/mailman/listinfo/nmtp-l>) oraz do przedstawiania streszczeń swoich referatów na ogólnodostępnej moderowanej liście *nmtp-ann* (<http://duch.mimuw.edu.pl/mailman/listinfo/nmtp-ann>).