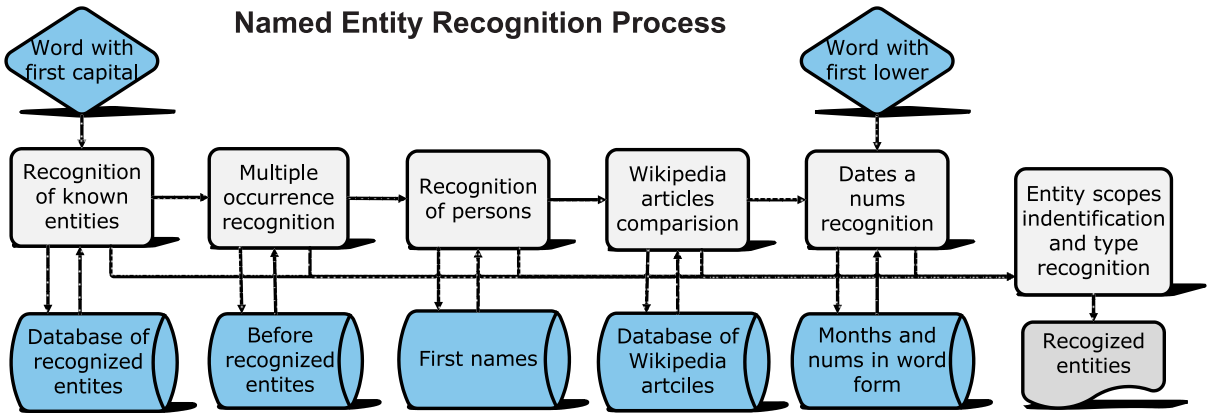
Named Entity Recognition for Slovak and Related Languages

Ondrej Kaššák
ondrej.kassak@gmail.com
supervisor: Michal Kompan

Named Entity Recognition Process



Text Example

Čínske úrady zrušili desiatky internetových stránok

Cenzúra internetu s najväčšou pravdepodobnosťou súvisí s odvolaním vysokého stranického predstaviteľa **Poa Si-laja**.

PEKING. Čínske úrady za posledný mesiac znefunkčnili **42** internetových stránok a odstránili vyše **210-tisíc** "škodlivých" odkazov a správ v rámci boja proti údajnému šíreniu klebiet a fám. **12. 4. 2012** o tom informovala oficiálna tlačová agentúra **Sin-chua**. Zásahy na internete s najväčšou pravdepodobnosťou súvisia s odvolaním vysokého stranického predstaviteľa **Poa Si-laja** a s vraždou, ktorú údajne spáchala jeho manželka. Na internete sa objavili mnohé špekulácie o tomto škandále. Už dva týždne sú odstavené dva hlavné čínske portály **Sina Weibo** a **Tencent QQ**, z ktorých každý má asi **300 miliónov** užívateľov.

```
<ENAMEX TYPE= “PERSON”>Poa Si-laja</ENAMEX>
<ENAMEX TYPE= “LOCATION”>PEKING</ENAMEX>
<NUMEX TYPE= “NUM”>210-tisíc</NUMEX>
<TIMEX>12.4.2012</TIMEX>
<ENAMEX TYPE= “ORGANIZATION”>Sin-chua</ENAMEX>
. . .
```

Recieved Results

Type	Precision	Recall	F-measure
Persons	0.97	0.80	0.88
Organizations	0.94	0.67	0.78
Locations	0.83	0.73	0.78
Dates	0.97	0.76	0.85
Numbers	0.90	0.87	0.88
Percents	0.83	0.68	0.75
Sums of Money	1.00	0.76	0.86
Miscellaneous	0.50	0.66	0.57
Total	**0.84**	**0.74**	**0.79**

- 60 articles
- 1620 entities
- 1204 correctly recognized
- 232 wrong identified

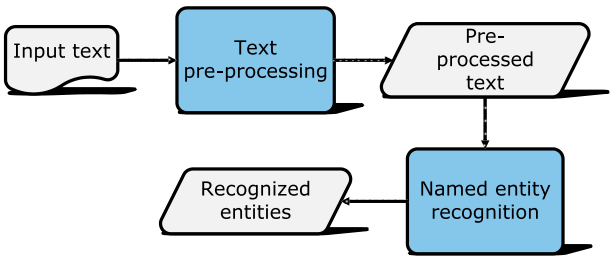
- Named entity recognition in texts
- Tool for recommendations based on content or search
- Focus on Slovak, but language independent

- Linguistic method, self learning
- Recognizes persons, organizations, locations, dates, numbers, percentage, money sums, miscellaneous

- Verification on newspaper articles, corpus
- Tested for Slovak, plan to test on Czech texts

- Recieved 79% F-measure (84% precision, 74% recall)

Proposed Method



STU FIIT
Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava, Slovakia
iit.src
PeWe@FIIT personalized web group