



Семинар на проекта *Координиране на езиковите ресурси в Европа*

Езикови ресурси за автоматичен превод от и на български

Стоян Михов

- проф. Сия Колковска, Институт за български език
- Рамона Стоянова, Агенция „Софита“
- Ани Русинова, Национална агенция за приходите
- Калина Цакова, Национален институт на правосъдието
- Веселин Петров, Сиела Норма АД

- Да се запознаят представителите на публичните институции със съществуващите ресурси
- Да се даде първоначална идея какви ресурси се търсят и къде могат да се намерят
- Да се обсъди с какви данни разполагат публичната администрация, министерствата и правителствените организации
- Да се информират ELRC и CEF.AT

- С какви езикови ресурси за автоматичен превод от и на български разполага Вашата институция? Свободен ли е достъпът до тях? Ако не, какви са лицензите?
- По какъв начин във Вашата организация е организирана работата за превод на документи от и на български: посредством външни изпълнители или имате специализиран отдел?
- Кои отдели във Вашата организация използват и съхраняват преводни документи? Кой в организацията се нуждае от тези данни?

- Как се създават, обработват и разпространяват преводните текстове във Вашата организация?
- Какви ресурси използвате, за да създадете преводни текстове: документи, преводна памет, речници, терминологични бази (едно- и многоезикови)?
- В какъв формат се запазват данните? Къде се съхраняват, архивират данните?

- Разработва се в *Института за български език* към БАН.
- Представява многоезиков корпус, който включва:
 - едноезичен корпус на български език – 1.2 милиарда думи, 240 000 текста
 - 47 чуждоезични паралелни корпуса с различна големина, наброяващи общо 4.2 милиарда думи. Включват текстове на английски, немски, френски, повечето славянски и балкански езици и др.
- Свободен достъп в интернет: <http://dcl.bas.bg/bulnc/>

Обем на най-големите паралелни корпуси в БНК

