

СЪЗДАВАНЕ, УПРАВЛЕНИЕ И СПОДЕЛЯНЕ НА ЕЗИКОВИ ДАННИ: СЪЩЕСТВУВАЩИ ПРАКТИКИ И ПРЕДИЗВИКАТЕЛСТВА В БЪЛГАРИЯ ПАНЕЛНА СЕСИЯ



ДАНИ ЗА ЕЗИКОВИ ТЕХНОЛОГИИ

- Езиковите технологии (и ИИ като цяло) се нуждаят от данни, които имплицитно кодират това, което искаме да направим;
- Езиковите данни са специален вид данни: езикът се генерира постоянно, но за да обучаваме системи за изкуствен интелект с езикови данни, се нуждаем от **цифрови и обработваеми данни**:
 - в големи количества - колкото повече, толкова по-добре;
 - на различни езици, области, жанрове... в зависимост от задачите и целите ни.
- Събирането на данни е сложна задача.

ДАНИИ ЗА ЕЗИКОВИ ТЕХНОЛОГИИ

Основните пречки пред събирането и споделянето на езикови данни за превод, посочени в **Бялата книга на ELRC** от 2019 г. за България са идентифицирани на база проведения Втори национален семинар за споделяне на езикови ресурси:

- Трудности при идентифицирането и убеждаването на високопоставени служители да разрешат обмена на данни;
- Липса на установени процедури за превод на документи на административно ниво, което води до:
 - ✓ Потенциални правни проблеми, напр. относно собствеността на данните, въпроси, свързани с личните данни и авторското право;
 - ✓ Технически трудности, свързани с обработката на данни;
 - ✓ Съпротива срещу новите технологии;
 - ✓ Липса на ресурси, които са необходими за подпомагане на техническата и правната подготовка и обмена на езикови данни;
 - ✓ Загриженост относно качеството на преводите, които биха могли да бъдат споделени (усещане, че качеството им може да не е достатъчно високо за споделяне).

ПАНЕЛ: СЪЗДАВАНЕ, УПРАВЛЕНИЕ И СПОДЕЛЯНЕ НА ЕЗИКОВИ ДАННИ

**МОДЕРАТОР
ХРИСТИНА
ДОБРЕВА**

**Единен
информацио-
нен портал за
Ковид-19**

Държавен експерт
Министерство на
транспорта и
съобщенията,
Национално
контактно лице

**СТАНИМИР
МИНКОВ**

**Портал за
обществени
консултации**

Главен експерт
Дирекция
„Модернизация на
администрацията“,
Администрация на
Министерския съвет

**РЕНИ
БОРИСОВА**

**Портал за
отворени
данни**

Държавен експерт
Министерство на
електронното
управление, отдел
"Електронна
идентификация и
данни"

**ТЕМЕНУЖКА
СПАСОВА**

**Портал за
пространст-
вени данни**

Министерство на
електронното
управление, отдел
"Електронна
идентификация и
данни"

**Д-Р ОРЛИН
КУЗОВ**

**Приложение
на технологиите
в администра-
цията**

GATE Institute -
Sofia University

БЛАГОДАРИМ ВИ ЗА ВНИМАНИЕТО!

Уебсайт: www.lr-coordination.eu

Twitter: @LR_Coordination

Facebook:

www.facebook.com/EuropeanLanguageResourceCoordination

www.linkedin.com/in/lrcoordination

Имейл: info@lr-coordination.eu

