



Секция по компютърна лингвистика



Езикови технологии и големи езикови модели за  
приложение в хуманитаристиката

**Стажове за студенти**



# Студентски стажове



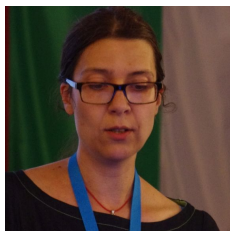
- Продължителност на стажантската програма: стажовете са с продължителност 30 часа, с възможност за удължаване при желание от страна на студента.
- Изисквания за предварителна подготовка: няма.
- Изисквания за студентски статут: студенти в бакалавърски специалности, завършили втори курс, и студенти в магистърски специалности.
- Удостоверение в края на стажа: при успешно завършване на програмата студентите получават сертификат за завършен стаж и придобита квалификация в областта на езиковите технологии.



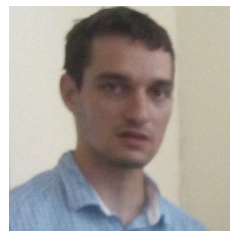
# Ментори в стажантската програма



проф. д-р  
Светла Коева



доц. д-р  
Цветана Димитрова



гл. ас. д-р  
Йордан Кралев



гл. ас. д-р  
Христина Кукова



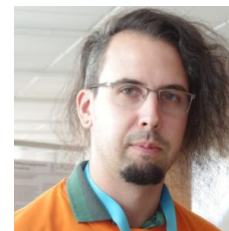
гл. ас. д-р  
Светлозара Лесева



гл. ас. д-р  
Валентина Стефанова



гл. ас. д-р  
Мария Тодорова



Димитър Христов



# Проекти на Секцията по компютърна лингвистика



Многоезиков корпус от изображения (MIC 21) (2022–2023)

[https://dcl.bas.bg/projects\\_list/mic21/](https://dcl.bas.bg/projects_list/mic21/)





# Проекти на Секцията по компютърна лингвистика



Инфраструктура за фина настройка на предварително обучени големи езикови модели (2024–2026)

[https://dcl.bas.bg/projects\\_list/ifGPT/](https://dcl.bas.bg/projects_list/ifGPT/)



Създаване на инфраструктура от алгоритми и софтуер (със свободен достъп) за подбор и предварителна обработка на големи данни за български, както и на специфични за дадена компания или индустрия данни и фина настройка на подходящи свободно достъпни големи езикови модели за решаване на конкретни задачи.



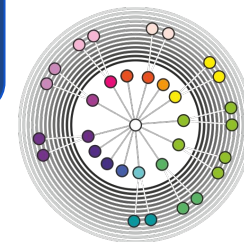
# Проекти на Секцията по компютърна лингвистика



Семантични ресурси и програми за обработка на езика  
(лексикално-семантични мрежи и езикови модели) (2023–2026)

**Българският уърднет:**  
<https://dcl.bas.bg/bulnet/>

**Българският фреймнет:**  
<https://dcl.bas.bg/bulframe/>



Голяма семантична мрежа, представляваща лексикалната система на езика посредством:

- над 120 000 синонимни множества (повече от 250 000 лексикални единици);
- свързани чрез 14 вида семантични релации (над 250 000 семантични връзки);
- съотнесени с еквивалентни синонимни множества в други езици.

Семантична мрежа, представляваща глаголите във вид на:

- система от абстрактни концептуални фреймове, съотнесени със семантичните фреймове на Фреймнет;
- описание на езиково специфичната задължителна съчетаемост между глаголите и техните обкръжения.



Езикови технологии и големи езикови модели за  
приложение в хуманитаристиката

**Задачи за изпълнение**



# Анотация на обекти в изображения



- Сегментиране и класифициране на обекти в изображения според категориите в предварително дефинирана анотационна схема.
- Запознаване с принципите на многомодалната анотация и съотнасянето на текстово съдържание с изображения.
- Запознаване с принципите на онтологичното представяне на знание и възможностите за търсене и извличане на информация от изображения по конкретни и абстрактни критерии: по названието на класовете, по езикови релации, както и чрез използване на релациите в онтологията.







# Анотация на обекти в изображения



- Сегментиране на обекти в изображения от определена тематична област чрез прецизното очертаване на границите им.



баскетболист, баскетболен играч

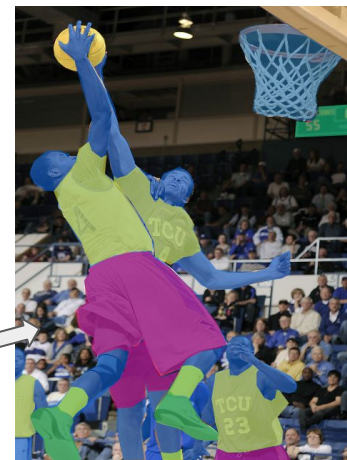
баскетболно табло

баскетболен кош, кош

баскетболна топка

фланелка, тениска

шорти



- Класифициране на сегментираните обекти чрез еднозначното приписване на подходящ клас измежду дефинираните в специално създадената *Онтология на визуалните обекти*.

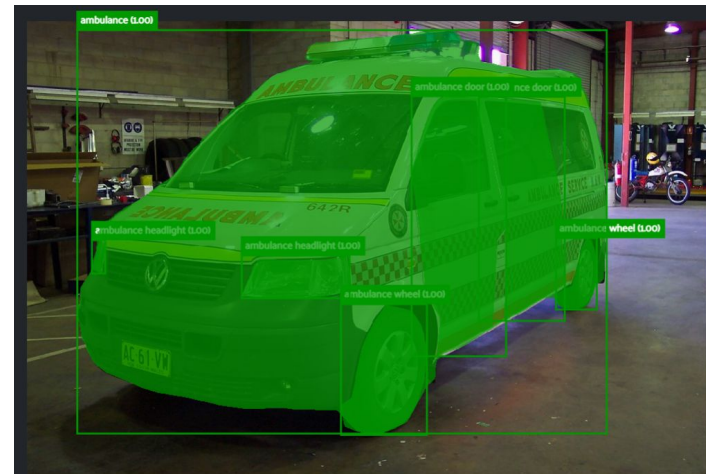


# Анотация на обекти в изображения



Придобиване на практически познания и умения за:

- анотация на многомодално съдържание чрез сегментиране на обекти в изображения и съотнасянето им с езиково описание;
- приложение на анотацията за търсене и извличане на абстрактно съдържание въз основа на анотираните обекти в изображенията;



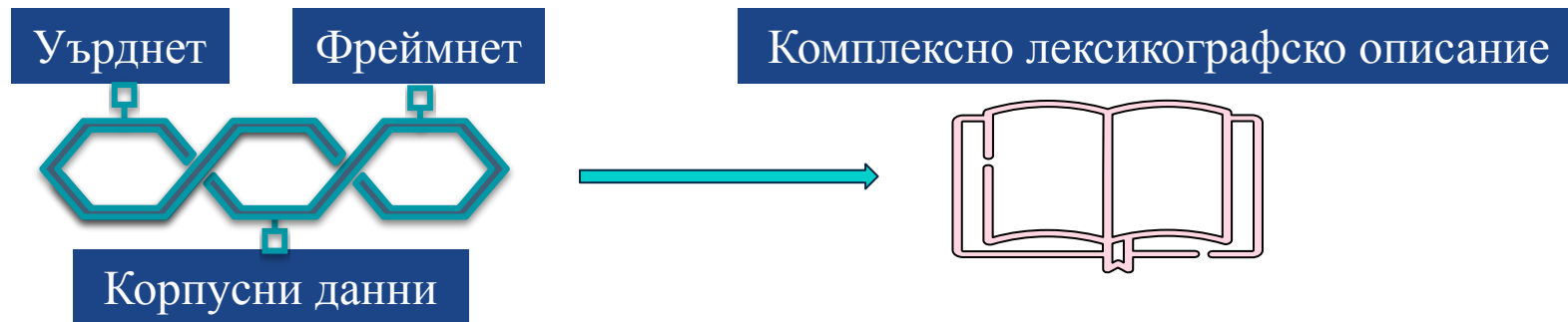
- приложение на многомодалните данни в системи за автоматично разпознаване, сегментиране и класифициране на обекти и описание на ситуации в изображения и видео.



# Лексикално-семантични мрежи и онтологии



- Представяне на лингвистично знание в структурирани лексикално-семантични ресурси.
- Запознаване с принципите на онтологичното представяне на знание в семантичните мрежи (Уърднет, Фреймнет и др.).
- Демонстриране на спецификите на лингвистичното описание в Уърднет и Фреймнет с оглед на приноса на всеки от ресурсите към обогатяването на семантичното представяне на лексикалните единици.
- Лингвистични обобщения за организацията на лексикалните единици в семантични класове.





# Лексикално-семантични мрежи и онтологии



- Създаване на синонимни множества за даден език и представянето им в структурата на Уърднет.

Подбор на лексикалните единици, обозначаващи дадено понятие.

Създаване на дефиниция, представяща лексикалното значение на синонимното множество изчерпателно и точно.

Подбор на примери за употребата на синонимите в реален контекст.

Проверка за наличие на релации с други синонимни множества:

хиперонимия	подобие	деривационни релации
меронимия	каузация	морфосемантични релации
антонимия	подсъбитие	...

✓ Synset: **bg** - **n**: лек автомобил:1; лека кола:1; кола:3

дефиниция: пътническо четириколесно моторно превозно средство с от две до пет-шест места

➤ **литерал**: лек автомобил:1

➤ **литерал**: лека кола:1

➤ **литерал**: кола:3

➤ **пример**: Той има нужда от лека кола, с която да ходи на работа.

част на речта: **n** ili: eng-30-02958343-n + 0  
- 0 семантичен клас: noun.artifact

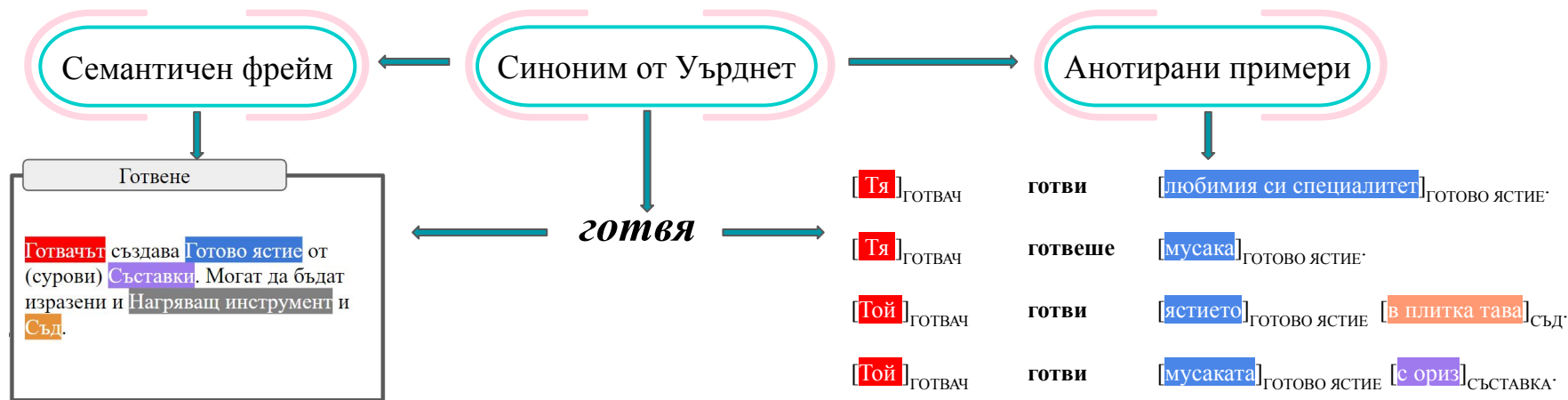
➤ **хипероним**: **bg** - **n**: автомобил:1; кола:5; моторно превозно средство:1



# Лексикално-семантични мрежи и онтологии II



- Семантико-синтактична анотация на изречения, илюстриращи избрани синоними от Уърднет, с помощта на семантичните фреймове в лексикално-семантичната мрежа Фреймнет:
- анотация на участниците в ситуацията, описвана от дадено изречение, чрез специален набор от етикети, назоваващи елементите на семантичното описание (фреймови елементи);
- определяне на синтактичната реализация на фреймовите елементи.

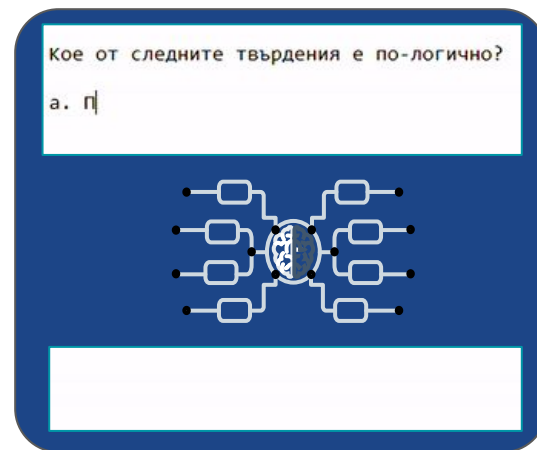




# Фина настройка на големи езикови модели



- Запознаване с големите езикови модели, основните технологии за създаването им и принципите на тяхната работа.
- Представяне на принципите, технологиите и моделите за програмиране на инструкции към големите езикови модели и демонстрация на приложението им в различни задачи.
- Демонстриране на възможностите на различни инструменти за профилиране на големите езикови модели във връзка с начина им на работа и конкретните задачи за изпълнение.
- Запознаване с инструменти за оценка и сравнение на работата на големите езикови модели.





# Фина настройка на големи езикови модели



- Създаване на набори от данни за фина настройка на големи езикови модели с цел специализация в конкретна тематична област или решаването на конкретна практическа задача.
- Профилиране на големи езикови модели с помощта на различни технологии.
- Създаване на набори от данни за тестване на резултатите от работата на големи езикови модели.
- Сравнение и оценка на работата на големите езикови модели с помощта на специализирани средства.





# Фина настройка на големи езикови модели



- Практически умения за използване на големи езикови модели за коректно извличане на информация и в нови приложения.
- Усвояване на умения за създаване на ефективни инструкции в зависимост от поставените цели и практически опит при работа с конкретни примери.
- Възможности за тестване на големите езикови модели чрез оценяване със стандартизирани набори от данни за оценка.

Кой от филмите може да се представи със следните емотикони?



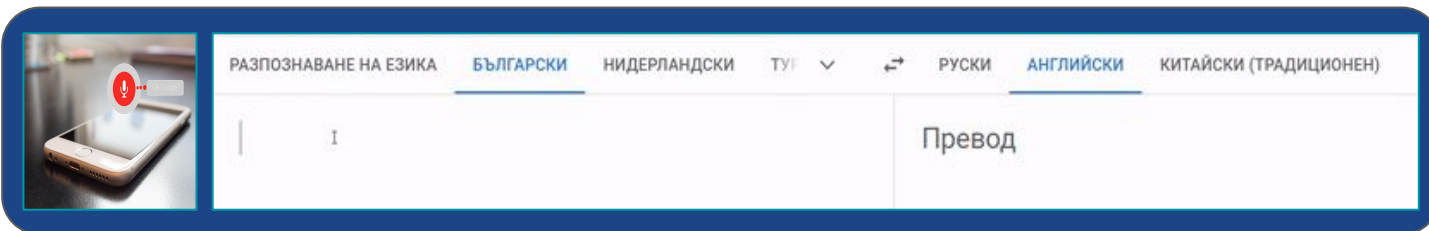




# Компютърната лингвистика



- Интелигентни приложения, способни да реагират адекватно на инструкции и да генерират коректна информация или реакция:
  - автоматичен превод;
  - езикови или многомодални диалогови системи с различно предназначение;
  - получаване на медицинска информация или консултация;
  - управление на свързани устройства в дома или офиса;
  - създаване и управление на интелигентни градове и много други.





# Специалистите по компютърна ЛИНГВИСТИКА



- Създават компютърни модели, методи и технологии за представянето на езикови структури, така че да бъдат успешно интерпретирани от компютърните програми и системи.
- Разработват езикови ресурси в подходящи формати с цел профилиране, фина настройка и тестване на езикови модели.





# Нови професии



- Изследователи в сферата на обработката и анализа на многоезиково и многомодално съдържание (потоци от текст, реч, видео в реално време) за нуждите на различни сектори: медии, финанси, търговия, политика, медицина, образование и много други.
- Специалисти по създаване на езикови ресурси и моделиране на езиковото знание за целите на езиковите технологии: автоматичен превод, разбиране и генериране на реч, автоматично отговаряне на въпроси и други.
- Специалисти по създаване на езикови модули и моделиране на езиковите компоненти при взаимодействието между човек и компютър.
- Специалисти по създаване на инструкции и еталонни данни за профилиране, фина настройка и тестване на големи езикови модели.



# Секция по компютърна лингвистика



<https://dcl.bas.bg/>



<https://www.linkedin.com/company/department-of-computational-linguistics>



<https://www.instagram.com/compling2/>



<https://x.com/CLinBulgaria>



<https://www.facebook.com/CompLingInBulgaria>

