ЦЕНТЪР ПО ИНФОРМАТИКА И ТЕХНИЧЕСКИ НАУКИ

УЧЕБНА ПРОГРАМА

Утвърждавам:
Декан:

**CS611**

# ОБРАБОТКА НА ЕСТЕСТВЕНИ ЕЗИЦИ

Приета: Прот. №29 от 29.06.2023 г.

Преподаватели доц. д-р П. Георгиева

## АНОТАЦИЯ

Дисциплината Обработка на естествени езици е въведение в областта на машинното обработване на текст, разпознаване и генериране на реч.. Включени са основни концепции и подходи от лингвистиката, математиката, вероятностите и компютърните науки, използвани при обработката на естествените езици. Разглеждат се теми като текстова прилика, част от етикетиране на реч, анализ, семантика, извличане на информация, отговаряне на въпроси и обобщаване на текст.

## ОСНОВНИ ЦЕЛИ И ИЗХОДНИ УЧЕБНИ РЕЗУЛТАТИ

Основна цел на „Обработка на естествени езици“ е да формира знания и умения за прилагане на теоретичните понятия за решаване на реални практически проблеми.

В края на обучението си по „Обработка на естествени езици“ студентът трябва да:

• притежава знания за основните понятия (гласово разпознаване, OCR, машинна обработка на текст);

• умее да моделира практически задачи, в които се използват основните концепции на разглежданите в курса теми;

• може да тества създадените алгоритми за гласово разпознаване, машинна обработка на текст и/или OCR, като използва уменията придобити по време на курса.

## ПРЕДПОСТАВКИ

Дисциплината „Обработка на естествени езици“ е избираема за специалността Изкуствен интелект и изисква предварително изучаване на материала по дисциплините: Теория на алгоритмите, Компютърни архитектури, Програмиране, Дискретна математика, Висша математика.

Друга важна предпоставка за успешното постигане на целите на курса е наличието на програмно оборудвани компютърни системи.

## ФОРМА НА ОБУЧЕНИЕ

Формите на обучение са редовна и задочна.

**СТАТУТ И СТРУКТУРА**

| специалност | статут | Кредити | редовно обучение | | | | задочно обучение | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | л | с | у | общ | л | с | у | общ |
| ИИ | избираема | 6 | 30 | 30 | | 60 | 15 | 15 | | 30 |

**СЪДЪРЖАНИЕ**

**А. СЕМИНАРНИ УПРАЖНЕНИЯ**

**Тема 1.** Естествени езици – морфология, синтаксис, семантика. Класификатори.

**Тема 2.** Обработка на текст. Машинен превод.

**Тема 3.** Извличане на информация от текст.

**Тема 4.** Машинно отговаряне на въпроси.

**Тема 5.** Обобщаване.

**Тема 6.** Анализ на чувства.

**Тема 7.** Гласова обработка(VoiceNet).

**Тема 8.** OCR.

**Тема 9.** Морфологичен анализ.

**Тема 10.** Генериране на текст (LSTM, GAN, RNN, Markov Chains).

**Б. КУРСОВ ПРОЕКТ**

Всеки студент получава индивидуално задание, състоящо се от три модула, като всеки модул съдържа:

- теоретична част;
- практическа част;
- проведени тестове.

Студентите разработват (погрупово или самостоятелно) курсовите проекти и се провежда защита.

**ПЛАНИРАНИ УЧЕБНИ ДЕЙНОСТИ И МЕТОДИ НА ОБУЧЕНИЕ**

1. На студентите се предоставя пълна информация за съдържанието на програмата по дисциплината, както конспекта и изискванията за оценяване.
2. Изпълнението на курсовата работа, както и успешната подготовка, изисква аудиторна работа, самостоятелна работа през семестъра и личен контакт с преподавателите.
3. Електронните материали, които се предоставят на обучаемите, подпомагат учебния процес и така се разширяват възможностите на преподавателите при осъществяване на задълбочено и систематизирано обучение.

**МЕТОДИ И КРИТЕРИИ ЗА ОЦЕНЯВАНЕ**

Крайната оценка по „Обработка на естествени езици“ се основава на точкова система, утвърдена в ЦИТН на БСУ, като максималният брой точки е 100, които се получават от следните дейности:

**1. Аудиторна заетост до 20 точки**

- посещение на лекции до 4 точки
- участие до 16 точки

**2. Извънаудиторна заетост (курсова работа, реферат, доклад и др.) до 36 точки**

**3. Изпитна процедура до 54 точки**

Скала за оценяване

- от 36 до 50 точки - среден 3
- от 51 до 65 точки - добър 4
- от 66 до 80 точки - много добър 5
- от 81 до 100 точки - отличен 6.

**ПРЕПОРЪЧИТЕЛНА ЛИТЕРАТУРА И ИНТЕРНЕТ РЕСУРСИ**


1. Christopher D. Manning, Hinrich Shutze, Foundations of Statistical Natural Language Processing, MIT Press, 28.05.1999 г.-720 pages
2. Dante Jurafsky, James H. Martin, Speech and Language Processing, 2nd Edition,
3. https://www.youtube.com/playlist?list=PL3FW7Lu3i5Jsnh1rnUwq_TcylNr7EkRe6
4. https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processing.htm
5. https://www.scm.tees.ac.uk/isg/aia/nlp/NLP-overview.pdf

# ИЗПИТЕН ВЪПРОСНИК

**CS611**

## ОБРАБОТКА НА ЕСТЕСТВЕНИ ЕЗИЦИ

Приета: Прот. №29 от 29.06.2023 г.

1. Естествени езици – морфология, синтаксис, семантика. Класификатори.
2. Обработка на текст. Машинен превод.
3. Извличане на информация от текст.
4. Машинно отговаряне на въпроси.
5. Обобщаване.
6. Анализ на чувства.
7. Гласова обработка(VoiceNet).
8. OCR.
9. Морфологичен анализ.
10. Генериране на текст (LSTM, GAN, RNN, Markov Chains).

**ПРЕПОРЪЧИТЕЛНА ЛИТЕРАТУРА И ИНТЕРНЕТ РЕСУРСИ**


1. Christopher D. Manning, Hinrich Shutze, Foundations of Statistical Natural Language Processing, MIT Press, 28.05.1999 г.-720 pages
2. Dante Jurafsky, James H. Martin, Speech and Language Processing, 2nd Edition,
3. https://www.youtube.com/playlist?list=PL3FW7Lu3i5Jsnh1rnUwq_TcylNr7EkRe6
4. https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processing.htm
5. https://www.scm.tees.ac.uk/isg/aia/nlp/NLP-overview.pdf