

**ИКОНОМИЧЕСКИ УНИВЕРСИТЕТ – В А Р Н А**  
**ФАКУЛТЕТ „ИНФОРМАТИКА“**  
**КАТЕДРА „ИНФОРМАТИКА“**

---

Приета от ФС (протокол № 9/24.04.2024 г.)

Приета от КС (протокол № 10/16.04.2024 г.)

**УТВЪРЖДАВАМ:**

**Декан:**

(проф. д-р Владимир Сълов)

**У Ч Е Б Н А   П Р О Г Р А М А**

ПО ДИСЦИПЛИНАТА: „ОБРАБОТКА НА ЕСТЕСТВЕН ЕЗИК“

ЗА СПЕЦ: „Data Science“; ОКС „бакалавър“ – редовно обучение

КУРС НА ОБУЧЕНИЕ: 4; СЕМЕСТЪР: 7

ОБЩА СТУДЕНТСКА ЗАЕТОСТ: 240 ч.; в т.ч. аудиторна 75 ч.

КРЕДИТИ: 8

**РАЗПРЕДЕЛЕНИЕ НА СТУДЕНТСКАТА ЗАЕТОСТ СЪГЛАСНО УЧЕБНИЯ ПЛАН**

<i>ВИД УЧЕБНИ ЗАНЯТИЯ</i>	<i>ОБЩО (часове)</i>	<i>СЕДМИЧНА НАТОВАРЕНОСТ (часове)</i>
АУДИТОРНА ЗАЕТОСТ:		
Т.ч.		
• ЛЕКЦИИ	45	3
• УПРАЖНЕНИЯ (семинарни занятия/ лабораторни упражнения)	30	2
ИЗВЪНАУДИТОРНА ЗАЕТОСТ	165	-

Изготвили програмата:

1. ....  
(проф. д-р Снежана Сълова)

2. ....  
(гл. ас. д-р Борис Банков)

Ръководител катедра: .....  
„Информатика“ (проф. д-р Юлиан Василев)

## I. АНОТАЦИЯ

Неструктурираните текстови данни понастоящем представляват голям дял от съществуващите информационни ресурси. За тяхната обработка се използват автоматизирани алгоритми, базирани на обработката на естествен език (Natural Language Processing, NLP), които позволяват бързо и прецизно извличане на знания. Компютърната обработка на естествен език се занимава с автоматично разпознаване на човешки езици. Технологиите, които се използват при NLP по своята същност представляват синтез между достиженията на компютърните науки и лингвистиката.

Дисциплината „Обработка на естествен език“ дава на студентите знания за основните термини и подходи за преобразуване на текста и формалното му представяне във вид, подходящ за обработка от компютърни програми. Разглеждат се спецификите при използване на методи за машинно обучение в компютърната обработка на естествени езици.

В резултат на изучаване на дисциплината студентите ще познават същността на системите за автоматично разбиране на текст, ще могат да ползват програмни библиотеки и софтуерни системи, да реализират програмни решения за обработка на естествен език и да използват съществуващи езикови корпуси.

В хода на обучение се прилагат и развиват следните ключови компетентности, съгласно препоръката на Съвета на Европейския съюз от 22 май 2018 г, а именно:

- **Цифрова компетентност – група 4.** Способност за ползване на технологиите за машинно обучение и създаване на програмни решения и обработка на естествени езици.
- **Личностна компетентност, социална компетентности и компетентност за придобиване на умения за учене – група 5.** Способност за справяне в сложни ситуации, при които от съществено значение е анализаторът и неговите умения. Стремех за непрекъснато усъвършенстване и прилагане на новите технологии.
- **Предприемаческа компетентност – група 7.** Способност за планиране и управление на проекти, свързани с извличане на знания от неструктурирани данни. Сферата на NLP предоставя много варианти за реализация на проектите, което налага анализ и оценка на силните и слаби страни на всеки един възможен подход. Знанията, получени по дисциплината помагат на студентите лесно да трансформират идеите си в реални приложения.

## II. ТЕМАТИЧНО СЪДЪРЖАНИЕ

No. по ред	НАИМЕНОВАНИЕ НА ТЕМИТЕ И ПОДТЕМИТЕ	БРОЙ ЧАСОВЕ		
		Л	СЗ	ЛУ
<b>Тема 1. Основи на компютърната лингвистика</b>		<b>3</b>	<b>-</b>	
1.1.	Същност на компютърната лингвистика	1	-	
1.2.	Развитие на технологиите за обработка на естествен език	2	-	
<b>Тема 2. Основи на обработката на текст</b>		<b>8</b>	<b>6</b>	
2.1.	Задачи на интелигентната обработка на текст	2	1	
2.2.	Етапи на процеса на обработка на текст	2	1	
2.3.	Основни техники в обработката на текст	2	2	
2.4.	Методи за обработка и анализ на текст	2	2	

<b>Тема 3. Клъстерен анализ на текстови данни</b>		<b>8</b>	<b>6</b>	
3.1.	Основни алгоритми за клъстеризация на текст	4	2	
3.2.	Анализ и оценка на прилагането на алгоритми за клъстерен анализ	4	4	
<b>Тема 4. Класификация на текстови данни</b>		<b>8</b>	<b>6</b>	
4.1.	Основни алгоритми за класификацията на текст	4	2	
4.2.	Сравнителен анализ от прилагането на различни класификатори за текст	4	4	
<b>Тема 5. Извличане на ключови понятия в текст</b>		<b>8</b>	<b>6</b>	
5.1.	Процес на извличане на ключови понятия	4	2	
5.2.	Подходи за извличане на ключови понятия	4	4	
<b>Тема 6. Анотация на текст</b>		<b>10</b>	<b>6</b>	
6.1.	Техники за резюмиране на текст	5	2	
6.2.	Подходи за осъществяване на процеса	5	4	
<b>Общо:</b>		<b>45</b>	<b>30</b>	

### **III. ФОРМИ НА КОНТРОЛ**

<b>№ по ред</b>	<b>ВИД И ФОРМА НА КОНТРОЛА</b>	<b>Брой</b>	<b>ИАЗ ч.</b>
<b>1.</b>	<b>Семестриално оценяване</b>		
1.1.	Контролна работа	1	40
1.2.	Тест	1	40
1.3.	Курсов проект	1	40
<b>Общо за семестриалното оценяване:</b>		<b>3</b>	<b>120</b>
<b>2.</b>	<b>Сесийно оценяване</b>		
2.1.	Изпит (тест)	1	45
<b>Общо за сесийното оценяване:</b>		<b>1</b>	<b>45</b>
<b>Общо за всички форми на контрол:</b>		<b>4</b>	<b>165</b>

### **IV. ЛИТЕРАТУРА**

#### **ЗАДЪЛЖИТЕЛНА (ОСНОВНА) ЛИТЕРАТУРА:**

1. Jurafsky, D. and Martin, J. (2023) Speech and Language Processing. <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>.
2. Учебни материали по дисциплината.

#### **ПРЕПОРЪЧИТЕЛНА (ДОПЪЛНИТЕЛНА) ЛИТЕРАТУРА:**

1. Сълова, С. (2012) Извличане на знания и анализ на данни от интернет източници. Варна: Наука и икономика, (Библ. Проф. Цани Калянджиев; Кн. 77).
2. Bankov, B. (2018) An Approach for Clustering Social Media Text Messages, Retrieved from Continuous Data Streams. Science. Business. Society: International Scientific Journal, Sofia: Scientific Technical Union of Mechanical Engineering INDUSTRY 4.0 et. al., 3, 2018, 1, 6-9.

3. Sulova, S. (2021) Text Mining Approach for Identifying Research Trends. CompSysTech '21: International Conference on Computer Systems and Technologies '21, June 18 - 19, 2021, Ruse, Bulgaria, New York: Association for Computing Machinery, 93-98.
4. Sulova, S., Todoranova, L., Penchev, B., Nacheva, R. (2017) Using text mining to classify research papers. 17th International Multidisciplinary Scientific Geoconference SGEM 2017, Albena, Vol. 17, Issue 21, 647-654.
5. Vajjala, S., Majumder, B., Gupta, A., Surana, H. (2020) Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems 1st Edition, O'Reilly Media.