

ИКОНОМИЧЕСКИ УНИВЕРСИТЕТ – ВАРНА
ФАКУЛТЕТ „ИНФОРМАТИКА“
КАТЕДРА „ИНФОРМАТИКА“

Приета от ФС (протокол № 27/ 26.04.2022 г.)

Приета от КС (протокол № 10/12.04.2022 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: 180 ч.; в т.ч. аудиторна 60 ч.

КРЕДИТИ: 6

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	30	2
• УПРАЖНЕНИЯ (семинарни занятия/ лабораторни упражнения)	30	2
ИЗВЪНАУДИТОРНА ЗАЕТОСТ	120	-

Изготвили програмата:

1.
(доц. 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.

- **Математическа компетентност и компетентност в областта на точните науки, технологиите и инженерството – група 3.** Способност за развиване и прилагане на математическо мислене и поглед с цел създаване на правилни подходи и алгоритми за обработка на различни видове текст.

- **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. Стремение за непрекъснато усъвършенстване и прилагане на новите технологии.

II. ТЕМАТИЧНО СЪДЪРЖАНИЕ

No. по ред	НАИМЕНОВАНИЕ НА ТЕМИТЕ И ПОДТЕМИТЕ	БРОЙ ЧАСОВЕ		
		Л	СЗ	ЛУ
	Тема 1. Основи на компютърната лингвистика	4	2	
1.1.	Същност на компютърната лингвистика	2	-	
1.2.	Развитие на технологиите за обработка на естествен език	2	2	
	Тема 2. Основи на обработката на текст	8	8	
2.1.	Задачи на интелигентната обработка на текст	2	2	
2.2.	Етапи на процеса на обработка на текст	2	2	
2.3.	Основни техники в обработката на текст	2	2	
2.4.	Методи за обработка и анализ на текст	2	2	
	Тема 3. Извличане на ключови понятия в текст	4	4	
3.1.	Процес на извличане на ключови понятия	2	2	

3.2.	Подходи за извличане на ключови понятия	2	2	
Тема 4. Клъстерен анализ на текстови данни		7	8	
4.1.	Основни алгоритми за клъстеризация на текст	4	4	
4.2.	Анализ и оценка на прилагането на алгоритми за клъстерен анализ	3	4	
Тема 5. Класификация на текстови данни		7	8	
5.1.	Основни алгоритми за класификацията на текст	4	4	
5.2.	Сравнителен анализ от прилагането на различни класификатори за текст	3	4	
Общо:		30	30	

III. ФОРМИ НА КОНТРОЛ

№ по ред	ВИД И ФОРМА НА КОНТРОЛА	Брой	ИАЗ ч.
1.	Семестриално оценяване		
1.1.	Контролна работа	1	30
1.2.	Тест	1	30
1.3.	Курсов проект	1	30
Общо за семестриалното оценяване:		3	90
2.	Сесийно оценяване		
2.1.	Изпит (тест)	1	30
Общо за сесийното оценяване:		1	30
Общо за всички форми на контрол:		4	120

IV. ЛИТЕРАТУРА

ЗАДЪЛЖИТЕЛНА (ОСНОВНА) ЛИТЕРАТУРА:

1. Jurafsky, D. and Martin, J. Speech and Language Processing. <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>, 2020.
2. Vajjala, S., Majumder, B., Gupta, A., Surana, H. Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems 1st Edition, O'Reilly Media, 2020.
3. Учебни материали по дисциплината.

ПРЕПОРЪЧИТЕЛНА (ДОПЪЛНИТЕЛНА) ЛИТЕРАТУРА:

1. Сълова, С. Извличане на знания и анализ на данни от интернет източници. Варна: Наука и икономика, 2021, (Библиография. Проф. Цани Калянджиев; Кн. 77).
2. Bankov, B. 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. 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, 2021, 93-98.
4. Sulova, S., Todoranova, L., Penchev, B., Nacheva, R. Using text mining to classify research papers. 17th International Multidisciplinary Scientific Geoconference SGEM 2017, Albena, Vol. 17, 2017, Issue 21, 647-654.