# ODESSA I.I. MECHNYKOV NATIONAL UNIVERSITY
## FACULTY OF MATHEMATICS, PHYSICS AND INFORMATION TECHNOLOGIES
### DEPARTMENT OF MATHEMATICAL SUPPORT OF COMPUTER SYSTEMS

### Syllabus of the course *"Methods of Natural Language Text Processing"*

| | |
|---|---|
| **Amount** | the total number of:    credits – 4; hours – 120; content modules - 2 |
| **Semester** | spring |
| **Days, Time, Place** | according to the class schedule |
| **Teacher(s)** | Valery Penko, PhD (Tech.), Associate Professor of the Department of Mathematical Support of Computer Systems |
| **Contact phone number** | (067)7441326 |
| **E-mail** | vpenko@onu.edu.ua |
| **Workplace** | department of mathematical support of computer systems |
| **Consultations** | face-to-face consultations: according to the class schedule<br>online consultations: ZOOM (link is generated at the beginning of classes) |

**COMMUNICATION**

Communication with students will be carried out by e-mail, in the classroom or via ZOOM.

**COURSE ABSTRACT**

**The subject** of the course is the main types of tasks that arise in the context of processing natural language texts, theoretical methods and software tools for solving them.

*Course prerequisites*

The course material is based on previously acquired knowledge, practical skills and skills of topics and areas related to algorithms, data structures, discrete mathematics and probability theory. The corresponding courses are taught within the educational program of the first (bachelor) level of higher education in specialty 126 "Information systems and technologies".

*Post-requisites of the course*

According to the used theoretical approaches and models, this course complements the discipline "Image processing methods and algorithms and computer vision" in the field of data analysis and processing and is the basis for mastering the following disciplines of the educational and professional master's training program in the specialty 126 "Information systems and technologies": "Professional research practice", "Execution of master's qualification work".

**Purpose** of the course is the study of the main modern approaches to solving the basic tasks of text processing in natural language and the practical application of the Python language and libraries to solve these tasks.

*Course content*

Considered:
- Computing over language is simple statistics.
- Classification of natural language text processing tasks.
- Obtaining access to corpora of texts and lexical resources.
- Use of basic lexical resources.
- WordNet Corpus.
- Access to text from the Web and to local text.
- NLP Pipeline. Implementation of individual stages.

**EXPECTED RESULTS**

As a result of studying the course, the student must

**know**: *the main ones opportunities Python languages for developing text processing software; the main capabilities of specialized packages for processing texts in natural language; the main types of tasks related to the processing of texts in natural language; features of the corpus-oriented approach to the processing of natural language texts.*

**be able**: *elaborate software provision that performs basic operations with texts; use specialized Python packages to improve the efficiency of basic text processing tasks in natural language; apply several varieties of language corpora as a resource for solving text processing tasks in natural language.*

***Competencies that the student receives as a result of studying the course:***

− *the ability to develop mathematical, information and computer models of objects and informatization processes.*
− *the ability to use modern data analysis technologies to optimize processes in information systems.*
− *the ability to mathematically model digital data and use effective algorithms for the analysis and transformation of multimedia data in modern information systems.*
− *the ability to conduct information analysis and create multidimensional models of subject areas.*

***Learning outcomes: upon completion of the course, the student will have skills***

− *develop and use data repositories, perform data analysis to support decision-making.*
− *to present research results, conduct discussions, and publish research results.*
− *develop mathematical models and software and information systems for solving current problems of analysis and processing of multimedia information.*
− *create optimized pipelines for preparing data for further storage and processing.*

**FORMS AND METHODS OF TEACHING**

The course will be taught in the form of lectures (16 hours) and laboratory classes (18 hours), organization of students' independent work (86 hours).

The basic training of students is carried out in lectures and laboratory classes.

During the teaching of the course, the following teaching methods are used: verbal (lecture, explanation); face-to-face (Power Point presentation); practical (laboratory works); work with literary sources (independent work of students).