|  |
| --- |
| Discipline „Internet Information Retrieval”, code: 44-a |
| Annotation:  The course aims to introduce the students to the principles of information retrieval in the Internet. The basic concepts and methods of document retrieval are explored, focusing on contemporary approaches and search algorithms in Web space. Attention is paid to questions related to indexing information, information retrieval models, ratings and queries. The principles of Web search engine building, as well as the features of modern information retrieval systems on the Internet, are explored |
| Main issues of the syllabus content:   * Basic principles of information retrieval. * Search Engine Architecture. Main components. Functioning. * Extracting the web pages. Crawling Web. RSS feeds. Storing retrieved documents. * Text processing. Evaluation of resultant set. * Parsing a document. Link analysis. * Ratings and indexing. Index building. Investment indices. Compression. * Queries. Transformation and refinement of inquiries. Delivering results. * Data retrieval models. * Assessment of search engines. * Classification and clustering. Spam detection. * Social demand. Tags. Filtering documents. * Extracting XML documents. Features. * Information retrieval systems. LEXIS / NEXIS, SMART, Dialogue, Dow Jones News / Retrieval, INQUERY. * Google search engine architecture. * Retrieving multi-media information |