

Subject code	Credits
INF5017	6

Course title in Lithuanian

INFORMACIJOS PAIEŠKA

Course title in English

INFORMATION RETRIEVAL

Short course annotation in Lithuanian (up to 500 characters)

Kurse aptariama informacijos modeliavimo ir paieškos teorija ir taikymai. Pristatomos klasikinės ir modernios informacijos modeliavimo teorijos, bei jų taikymas paieškai. Pradedama nuo bazinių žinių, ir baigiama naujausiais iššūkiais.
--

Short course annotation in English (up to 500 characters)

The course aims to study the theories and techniques used in modelling of information and text based retrieval. It discusses classical and modern techniques of knowledge modelling and their application for information retrieval. Students start from the basic notions and concepts, and finish with the newest challenges in the field.
--

Prerequisites for entering the course

Mathematical Statistics, Algebra, Software Development, Information Retrieval.
--

Course aim

Learn to build IR system

Links between study programme outcomes, course outcomes and criteria of learning achievement evaluation, study methods and methods of learning achievement

Course outcomes	Criteria of learning achievement evaluation	Study methods	Methods of learning achievement assessment
1. Knowledge and understanding of documents, their collection, preprocessing and representation.	Text and other data preprocessing methods, web crawling, data storage, collection of running corpora	Lectures, practical works, individual work	Practical work and its presentation Mid-term exam
2. Knowledge and understanding of information retrieval models (Boolean, Vector, Probabilistic and their extensions)	Implementation of simplified models and experiments with running corpora	Lectures, practical works, individual work	Practical work and its presentation Mid-term exam
3. Evaluate information retrieval models	Evaluation of IR models for running corpora	Lectures, practical works, individual work	Practical work and its presentation Mid-term exam
4. Knowledge and understanding ranking of linked documents (PageRank, HITS)	Implementation and ranking of test corpora	Lectures, practical works, individual work	Practical work and its presentation Exam
5. Knowledge and understanding of application of clustering in IR	Applying selected clustering techniques for running corpora	Lectures, practical works, individual work	Practical work and its presentation Exam
6. Knowledge and understanding of application of classification in IR	Applying selected classification techniques for running corpora	Lectures, practical works, individual work	Practical work and its presentation Exam
7. Choosing best models and tools for specialized IR systems	Implementing a simple IR systems using selected tools	Lectures, practical works, individual work	Practical work and its presentation Exam
8. Presenting results	Formulate problem, present solution	Individual work and studies, discussions, consultations	Practical work and its presentation Mid-term exam Exam

Links between study programme outcomes and course outcomes

Study programme outcomes	Running number of course outcome							
	1	2	3	4	5	6	7	8
5. Develop mathematical models integrating the knowledge from various fields and different mathematical modelling techniques, and analyse the modelling results assessing the model adequacy and accuracy		+	+	+	+	+	+	
7. Analyse, understand and use mathematical methods	+	+	+	+	+	+	+	
11. Convey mathematical information to specialists of different fields orally and/or in written form, critically evaluate it								+
13. Take moral responsibility for the results of work								+

Content

No	Content (topics)
1.	Introduction to Advanced Information Retrieval
2.	Advanced NLP in IR.
3.	Indexing of big amounts of data, distributed processing.
4.	Extensions of boolean and vector IR model.
5.	Probabilistic IR model and its extensions.
6.	Latent semantic and other advanced IR models.
7.	Advanced evaluation, behavioural analysis.
8.	Advanced search in WWW: tools for scanning, processing; HITS, PageRank.
9.	Clustering of results: flat clustering, hierarchical clustering, clustering of big amounts of data, feature construction and selection.
10.	Classification in IR: Naïve Bayes, Vector Space, SVM
11.	Focused search.

Distribution of workload for students (contact and independent work hours)

Lectures	45 hours
Group work	5
Laboratory work	10
Individual students work	100
Total:	160

Structure of cumulative score and value of its constituent parts

Final written exam (50%), mid-term written exam (17%), and assessments of laboratory (practical) work (33%).
--

Recommended reference materials

No.	Publication year	Authors of publication and title	Publishing house	Number of copies in		
				University library	Self-study rooms	Other libraries
Basic materials						
1.	2008	C. D. Manning, P. Raghavan, H. Schütze. Introduction to Information Retrieval.	Cambridge University Press			On-line book: http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html .
2.	1979	C. J. Van Rijsberge. Information Retrieval.	Butterworths			http://www.dcs.gla.ac.uk/Keith/Preface.html
3	2015	W. Bruce Croft Donald Metzler Trevor Strohman. Search Engines Information Retrieval in Practice	Pearson			http://ciir.cs.umass.edu/downloads/SEIRiP.pdf

<i>Supplementary materials</i>			
		<i>Mi Islita</i>	http://www.miislita.com/
		<i>www.google.com</i>	<i>www.google.com</i>

Course programme designed by

Prof. dr. Tomas Krilavičius
