I051	**Computational Linguistics**	L	S	E	ECTS
		2	0	2	6

Course objectives. Obtaining knowledge of natural language and its computational processing using algorithms and programs for morphology, syntax and semantics, with special attention to extracting information from documents and forming linked LLOD data (*linguistic linked open data*). Getting familiar with *Python* module for text processing and obtaining techniques for pattern recognition. Applying mathematical and statistical knowledge of modeling and computer processing of Croatian language.

Course prerequisites. Introduction to Computer Science, Data structures and algorithms I

Syllabus

1. Word, forms and types of words, phrases, sentences. Grammar. Morphology, syntax, semantics.
2. Introduction to computational linguistics. Natural and artificial language grammar. Language as a subset of free monoids (Kleen closure).
3. Categorial or C-grammar. Generative or PS grammar (phrase structure grammar)
4. Grammatical derivatives; LA (left-associate) grammar.
5. Morpho-syntactic tagging (lemma tagging, syntactic SPO tagging, PoS – part of speech tagging, category, annotation). Algorithms in morphology, XML and visual syntactic tagging.
6. Semantic tagging classes (role & sense tagging). Semantic trees. Parsers.
7. Regular expressions. Information retrieval from sentence. Supervised and non-supervised machine learning.
8. Data mining. Classification and clustering of documents, probabilistic models.
9. Term frequency–inverse document frequency (tf-idf) approach, latent semantic analysis (LSA).
10. Network ontologies and linked data, SparQL queries, triple store storage (s-p-o: subject-object-predicate) triples of information.

EXPECTED LEARNING OUTCOMES

No.	LEARNING OUTCOMES
1.	Demonstrate knowledge and understanding which ensures a foundation for original development and application of mathematical and computer science ideas in linguistics.
2.	Apply knowledge, understanding and skills of problem solving in a broader context of linguistics.
3.	Integrate new knowledge for successful solving of programming problems in designing and modeling linguistic data
4.	Clearly and unequivocally explain own conclusions to peers and laity, based on knowledge and arguments.
5.	Apply acquired learning skills on lifelong education in the area of linguistics.

COUPLING OF THE EXPECTED LEARNING OUTCOMES, TEACHING PROCESS ORGANIZATION AND THE EVALUATION OF THE TEACHING OUTCOMES

TEACHING PROCESS ORGANIZATION	ECTS	LEARNING OUTCOMES **	STUDENT ACTIVITY *	EVALUATION METHOD	SCORE	
					min	max
Lecture attendance	1	1-5	Class attendance, discussion, solving the problems individually and in a team	Lists with signatures, observing the activity during the lectures	0	10
Homework	1	1-4	Solving the problems individually	Grading	18	30
Repeated exam	2	1-4	Revising	Grading, discussion	16	30
Final exam	2	1-4	Student project	Oral exam	16	30
TOTAL	6				50	100

Teaching methods and student assessment. Lectures will be conducted by demonstrating programming examples in Python+NLTK (natural language toolkit) with the help of several Python modules designed for Croatian language (corpus, morphology, semantics). In the practical part of the lectures (exercises) students will become acquainted with Web2Py http://www.web2py.com MVC (*model-view-controller framework*) with the purpose of having all students' homework automatically embedded in the web environment. Lectures and exercises are obligatory. The final exam consists of a successfully done programming project and it is held after the completion of all lectures and exercises and successful participation in homework assignments and mid-term exams. In the projects, students will use a part of programming modules from known NLT repositories: http://www.nltk.org/, http://www.clips.ua.ac.be/ and http://scikit-learn.org/stable/.

Can the course be taught in English: Yes

Basic literature:

1. S. Bird, E. Klein, E. Loper: "Natural Language Processing with Python", http://www.nltk.org/book/, O'Reilly Media, 2009.

Recommended literature:

1. C. Chiarcos, S. Nordhoff, S. Hellmann: "Linked Data in Linguistics", Springer-Verlag, 2012.
2. M. Essert, K. Štrkalj Despot: "Python za jezikoslovce (*u pripremi*)", IHJJ, 2016.

3. R. Hausser: “Foundations of Computational Linguistics: Human-Computer Communication in Natural Language”, 3Ed, Springer, 2014.
4. R. Grishman: “Computational linguistics: an introduction. (Studies in natural language processing)” , Cambridge University Press, third edition, 1994.
5. D. Jurafsky, J. H. Martin: “Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition”, Pearson Education, 2009.
6. R. Delmonte: “Computational linguistic text processing: lexicon, grammar, parsing and anaphora resolution”, Published by Nova Science Publishers, Inc., New York. 2008.
7. S. P. Abney: “Semisupervised learning in computational linguistics”, Taylor & Francis Group/ Chapman & Hall/CRC, 2008.
8. C. J. Fillmore: “Semantics of Natural Language”, Springer Netherlands, 1973.
9. M. W. Berry: “Survey of text mining: clustering, classification, and retrieval”, Springer-Verlag New York, Inc., 2004.
10. C. C. Aggarwal, C. Zhai: “Mining Text Data”, Springer Science+Business Media, 2012.
    P. Harrington: “Machine Learning in Action”, Manning Publications Co., 2012.
11. P. Harrington: “Machine Learning in Action”, Manning Publications Co., 2012.