

Table S2.6.4. Form for the preparation of the course information sheets				
<b>Name of the subject:</b> Mining unstructured data and natural language processing				
Code of the subject	Status of the subject	Semester	Number of ECTS credits	Class load
	Optional		10	
<b>Study programme for which it is organized:</b> PhD Program "Sustainable Development" (Computer Science and Informatics)				
<b>Dependency by other subjects:</b> none				
<b>Objectives of studying this subject:</b> To acquire theoretical and practical knowledge and skills for processing unstructured data.				
<b>Contents of the subject (teaching units, forms of students' individual work, forms of testing) presented per working weeks in the academic calendar:</b>				
Preparatory week				
I week	Use of information retrieval techniques, data mining and machine learning on web, source code, text and other unstructured data.			
II week	Usage of recurrent neural networks for classification and creation of text and other unstructured data.			
III week	Unstructured data processing with convolutional neural networks, Hopfield networks, restricted Boltzmann machine, and recursive networks with long-short term memory (LSTM), gated recurrent units (GRU) and other neuron types in deep neural networks.			
IV week	Unstructured data representation models (bag-of-words, LDA, topic models, vector representation with word2vec, doc2vec and others).			
V week	Classification, clustering and concept mining from unstructured data based on their meaning of the content.			
VI week	Sentiment analysis aiming to determine in the text expressed feelings and attitude.			
VII week	Document summarization for shortening of text to only the major points of original document.			
VIII week	Optimization of unstructured data mining with nature inspired algorithms.			
IX week	Named-entity recognition and entity relation modeling			
X week				
XI week				
XII week				
XIII week				
XIV week				
XV week				
<b>Methods of education:</b>				
<ul style="list-style-type: none"> <li>lectures,</li> <li>project assignment.</li> </ul>				
<b>Students' load</b>				
<u>Weekly</u>		<u>In Semester</u>		
		<ul style="list-style-type: none"> <li>Lectures: 60</li> <li>Individual work: 240</li> </ul>		
<b>Students' obligations during the teaching:</b>				

**Literature:**

- Sarkar, D., 2016. Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from Your Data. Apress.
- Bird, S., Klein, E. and Loper, E., 2009. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc."
- Reese, R.M., 2015. Natural language processing with Java. Packt Publishing Ltd.
- Hearty, J., 2016. Advanced Machine Learning with Python. Packt Publishing Ltd.
- Goodfellow, I., Bengio, Y. and Courville, A., 2016. Deep learning. MIT press.
- Géron, A., 2017. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media.

**Learning outcomes (complied with the outcomes for the study programme):**Knowledge and understanding:

On completion of this course the student will be able to

- demonstrate knowledge and understanding of the process of preparation and building of the mining system for unstructured data
- demonstrate knowledge and usage of methods of unstructured data mining on real cases
- demonstrate the ability to evaluate and interpret the results of the mining process for unstructured data

Transferable/Key skills and other attributes:

Use of tools and technologies: use of software tools and technologies for data mining of unstructured data in one of the programming languages.

**Forms of tests and evaluation:**

- completed project – 50%
- oral examination – 50%

**Name and surname of teacher and associate:**

Sašo Karakatič

**Particularities needed to be emphasized for the subject:**

*Note (if needed):*