MI002	**Data Clustering and Applications**	L	P	S	ECTS
		2	1	1	5

Course objectives. Students will be introduced to fundamental facts and results in the field of data clustering, as well as possible applications.

Prerequisites. Undergraduate mathematics or computer science study programme.

Course content.

1. Introduction and motivation. Problem statement and basic properties. Various examples from the application.
2. Representative of the set of data. Basic quasi-metric functions (LS and LAD quasi-metric function). Representative of the finite set from Rn. Mahalanobis quasi-metric function. Interpretation of periodical data on the unit circle.
3. Hard clustering. Clustering of the data that have one, two or more features. Primary and dual problem based on LS quasi-metric function. Criteria functions and minimum distance principle.
4. K-means algorithm. Incremental algorithm.
5. Partition with the most appropriate number of clusters – indices (Calinski-Harabasz, Davies-Bouldin, Silhouette Width Criterion). Partition comparison (Rand index, Hausdorf distance).
6. Analysis and text recognition
7. Recognition of other geometric objects in plane. Line, circle and ellipsis as representative of the set of data in a plane. Recognition of more aforementioned geometrics objects of the same type.
8. Soft clustering. Fuzzy c-means algorithm. Gustafson-Kessel k-means algorithm.

LEARNING OUTCOMES

No.	LEARNING OUTCOMES
1.	Recognize problems with corresponding databases where they can apply the acquired knowledge;
2.	Choose appropriate methods for data clustering on its own;
3.	Analyze complexity of optimization problems of data clustering and learn how to apply the basic k-means algorithm, as well as some other methods;
4.	Apply the idea of various geometrics objects as the representative of the data;
5.	Use mathematical correctness in derivations of formulas and procedures that are used in this course

RELATING THE LEARNING OUTCOMES, ORGANIZATION OF THE EDUCATIONAL PROCESS AND ASSESSMENT OF THE LEARNING OUTCOMES

TEACHING ACTIVITY	ECTS	LEARNING OUTCOME **	STUDENT ACTIVITY*	EVALUATION METHOD	POINTS	
					min	max
Attending lectures and exercises	1	1-5	The presence at lectures, discussions, teamwork and	Attendance lists, tracking activities	0	4

			independent work on assignments			
Written exam (Mid-terms)	2	1-5	Preparing for the written exam	Verification of correct answers (evaluation)	25	48
Final exam	2	1-5	Revising	Oral exam	25	48
TOTAL	5				50	100

Teaching methods and knowledge assessment. Lectures and exercises are illustrated by ready-made software packages. Exercises are partially presented in computer laboratory. Participation in lectures, exercises and seminars hours is compulsory. Exam contains oral and written exam which can be taken when lectures are finished. Acceptable mid-term written exams replace written part of the exam. Student can influence of the final exam by working on their homeworks and seminars during semester. Homeworks contain additional topics that require independence and creativity for their solutions. Seminars are considered as an extension of homework.

Can the course be taught in English: Yes.

Basic literature:

1. R.Scitovski, K.Sabo, Grupiranje podataka i primjene, recenzirani nastavni materijal na web stranici
2. S. Theodoridis, K. Koutroumbas, Pattern Recognition, Academic Press, Burlington, 2009, 4th edition.
3. J. Kogan, Introduction to Clustering Large and High-Dimensional Data, Cambridge University Press, 2007.

Recommended literature:

1. J. C. Bezdek, J. Keller, R. Krisnapuram, N. R. Pal, Fuzzy models and algorithms for pattern recognition and image processing, Springer, 2005.
2. S. Butenko, W. A. Chaovalitwongse, P. M. Pardalos, Clustering Challenges in Biological Networks, World Scientific, 2009.
3. I. S. Dhillon, S. Mallela, R. Kumar, A divisive information theoretic feature clustering algorithm for text classification, Journal of Machine Learning Research, 3(2003) 1265–1287.
4. L. Kaufman, P. J. Rousseeuw, Finding groups in data: An introduction to cluster analysis, John Wiley & Sons, Chichester, UK, 2005.
5. B. Mirkin, Data clustering for Data Mining, Chapman & Hall/CRC, 2005.
6. R. Scitovski, M. B. Alić, Grupiranje podataka, Odjel za matematiku, Sveučilište u Osijeku, 2016.
7. H. Späth, Cluster–Formation und- Analyse, R. Oldenburg Verlag, München, 1983.
8. P. N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Wesley, 2006.