Zgłoszenie tematu **INŻYNIERSKIEJ** pracy dyplomowej

STUDIA I STOPNIA rok akademicki 2021/22

Promotor:	**dr hab. Jozef Kapusta, prof. UP**
Temat pracy dyplomowej (j. polski, j. angielski):	Algorithms for Sequence Rule Analysis in Web Usage Mining *Algorytmy analizy reguł sekwencji w Web Usage Mining*
Zakres pracy i oczekiwane rezultaty praktyczne:	Web usage mining is focused on the analysis of the behaviour of users while surfing the net. The most frequent sources of data are the ones automatically stored in the log files. In such data, we follow series – sequences of visiting individual pages by the user. In the sequences, we can look for users behaviour patterns. For this purpose, it is the best way to use sequence rule analysis, the aim of which is to extract sequence rules. The most commonly used algorithm for the sequence is "Apriori All" algorithm. There are other algorithms for sequence patterns analysis like SPADE or prefixScan which could be interesting for the sequence patterns identification. The aim of the thesis is summarization information about algorithms for sequence rule analysis and creation of practical examples of these algorithms in the selected data mining application.
Aspekt inżynierski*:	Application selected methods of sequence rule analysis (Apriori All, SPADE, PrefixScan, etc.), implementation selected methods for web portal analysis, understanding web-visitors behavior using web usage mining.
Wymagane oprogramowanie/języki programowania**:	Jupyter Notebook (Python)
Środowisko uruchomieniowe**:	
Dodatkowe wymagania i uwagi:	English Language
Literatura**:	1. Bing, L. 2011. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer Heidelberg. p. 637. ISBN 978-3-642-19459-7. 2. Gouda, K., Hassaan, M., Zaki, M.J. 2010. Prism: An effective approach for frequent sequence mining via prime-block encoding, Journal of Computer and System Sciences, Volume 76, Issue 1, 2010, Pages 88-102 3. Pei, J., Han, J., et al. 2004. Mining sequential patterns by pattern-growth: the prefixspan approach, IEEE Transactions on Knowledge and Data Engineering, 16 (2004), pp. 1424-1440 4. Zaki, M.J. 2001. SPADE: an efficient algorithm for mining frequent sequences. Machine Learning, 42 (2001), pp. 31-60

	5. AGRAWAL, R., SRIKANT, R. 1994. Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the 20th International Conference on Very Large Databases, pp 487-499.

***należy uzasadnić/wskazać, czy praca spełnia wymagania inżynierskie**

**pola opcjonalne