

**Doctoral School of Information and Biomedical Technologies of the Polish Academy of Sciences**

**Subject:** The reverse clustering paradigm in the partially supervised classification of Big Data

**Supervisor:** Jan W. Owsinski, Ph.D., D. Sc., Eng.; +48 22 3810 213; [owsinski@ibspan.waw.pl](mailto:owsinski@ibspan.waw.pl)

**Co-supervisor:** Jarosław Stańczak, Ph.D., Eng.; [stanczak@ibspan.waw.pl](mailto:stanczak@ibspan.waw.pl)

**Place of research:** Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warszawa, Poland

**Scientific discipline:** Technical Computer Science

**Recruitment form:** Interview

**Available positions:** 1 or 2

**Project description:**

One of essential keys to Artificial Intelligence (AI) algorithms resides with the techniques of classification, based on very widely differentiated methodologies and theories. Among the basic scientific issues in this area is the effective identification and adaptation of the rules or models for classifying large sets of data, having dynamic character, for which we dispose of only a limited number of (and limited capacity of assigning) labels. The paradigm of reverse clustering, coined in and then described, e.g., in [1, 2, 3], constitutes a good tool for identifying the principles of classifying for this class of tasks, with the classification principles established belonging to the broadly conceived domain of cluster analysis. The very task of the reverse clustering consists in finding, for some given data set (of objects, observations) and a partition of this set, of a clustering-based partition that is possibly similar to the given one. The intention is to construct the principles of partition that could be applied to a much wider class of data sets, having, in general, similar nature to the original one. The purpose of the project is to build and verify, on the empirical material, in particular – concerning web traffic – the effective algorithms, especially regarding optimization, implementing the concept of reverse clustering. The starting point is provided by the already existing methodological solutions, described, e.g., in [1, 2, 3], in particular – the evolutionary algorithms, while the yardstick of success is significant improvement of the computational effectiveness through appropriate adjustment to the character of the problem being solved and the broadening / flexibilisation of the considered class of parameters of the clustering methods accounted for.

**Bibliography:**

1. Owsinski, J. W., Stańczak, J., Opara, K., Zadrożny, Sł., Kacprzyk, J.: *Reverse Clustering: Formulation, Interpretation and Case Studies*. Studies in Computation Intelligence 957. Springer International Publishing, 2021.
2. Owsinski, J. W., Stańczak, J. and Zadrożny Sł.: Designing the Municipality Typology for Planning Purposes: The Use of Reverse Clustering and Evolutionary Algorithms. W: P. Daniele and L. Scrimali, eds., *New Trends in Emerging Complex Real Life Problems. ODS, Taormina, Italy, September 10-13, 2018*. AIRO Springer Series, vol. 1. Springer, Cham, 2018. ISBN/EAN 9783030004736 [https://doi.org/10.1007/978-3-030-00473-6\\_33](https://doi.org/10.1007/978-3-030-00473-6_33)
3. Owsinski, J. W., Kacprzyk, J., Opara, K., Stańczak, J., Zadrożny, Sł.: Using a reverse engineering type paradigm in clustering: An evolutionary programming based approach. W: V. Torra, A. Dalbom and Y. Narukawa, eds., *Fuzzy Sets, Rough Sets, Multisets and Clustering. Dedicated to Prof. Sadaaki Miyamoto. Studies in Computational Intelligence* 671, Springer, 2017, ISBN 978-3-319-47556-1; DOI 10.1007/978-3-319-47557-8.