Comparative Overview of Rough Set Toolkit Systems for Data Analysis

Piotr Pięta[1,*], *Tomasz* Szmuc[1], and *Krzysztof* Kluza[1]

[1]AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Krakow, Poland

Abstract. Inconsistency, lacking values of attributes or parameters, as well as discrepancies between records caused by insufficient precision cannot always be managed in the initial phases of knowledge discovery, *i.e.*, data preparation and refinement. The theory of rough sets aims to overcome problems that are caused by uncertainty and lack of precision within the gathered data sets. This approach is a useful tool that operates on a formal model using relational algebra, elementary operations on finite sets and first-order logic. In this paper, we present an analysis of existing rough set tools, namely: Rough Set Exploration System, Rough Sets Data Explorer, Rough Set Data Analysis Framework, Waikato Environment for Knowledge Analysis and Rough Set Toolkit for Analysis of Data. Our comparison is performed only theoretically and covers the available algorithms, preparation of input data, licensing, as well as installation requirements.

1 Introduction

Data mining constitutes one of the rapidly growing areas of information engineering. It can also be considered as a significant part of the process which aims to manage gathered information in a more efficient way. After being initially processed and converted into an appropriate representation, data can serve as a valuable source of knowledge. On the other hand, problems such as type inconsistency, lacking values of attributes or parameters, as well as discrepancies between records caused by insufficient precision, cannot always be managed in the initial phases of knowledge discovery, i.e. data preparation and refinement.

Rough Sets Theory (RST) [1], developed by Prof. Zdzislaw Pawlak in the early 1980s, aims to overcome problems caused by uncertainty and lack of precision within the gathered data sets. Uncertainty and vague concepts almost always appear in large data sets. Uncertainty can be a result of the lack of knowledge about a phenomenon, a human error, or sometimes just the effect of insufficient precision. Rough Set Theory is not the only developed mathematical formalism to deal with such problems. Starting from 1960, numerous attempts were made to construct a model that would uncertainty and eliminate its effects so as to maintain the appropriate proportions. One of them was fuzzy sets, proposed by Lotfi Zadeh, which bases on the concept of membership function presenting the degree of being a part of a set. Both fuzzy sets theory and rough sets theory use three-valued logic formulated in the 19th century by Łukasiewicz and constitute the possible generalisations of the classical set theory expressed by George Cantor [2]. In the rough set theory, the key issue is the concept of approximation of non-crisp set by a pair of crisp sets. This expresses the fundamental fact that using available knowledge about some sets in current knowledge base we cannot define explicitly other sets in the same knowledge base (this depends on the current knowledge that we have about objects). The Pawlak's approach [1] focuses on partitioning of space (universe of discourse) by means of indiscernibility relation (equivalent to classification term) into smaller units, called elementary categories (concepts), which group elements with the same values of their attributes (available knowledge). The smaller units can build larger parts called basic categories. Numerous extensions and generalisations of the classical RST theory were developed as well as hybrid approaches combined from them such as Probabilistic Rough Sets [3], Rough-Fuzzy Sets [4], Fuzzy-Rough Sets [4,5], or Dominance-based Rough Sets Approach (DRSA) [6,7,8], Variable-Consistency Dominance-based Rough Sets (VC-DRSA) [9].

In the following section, we present a comparative overview of the five selected tools, specifically RSES, ROSE2, jMAF, WEKA, and ROSETTA. Our theoretical comparison covers the available algorithms, the implementation language, supported operating systems, licensing, installation requirement, as well as input and output formats available in each of the presented tools.

2 Comparative overview

In the following subsections, we present a brief description of five tools (RSES, ROSE2, jMAF, WEKA, and ROSETTA) and overview algorithms available for each tool. These algorithms support tasks related to data analysis at various stages of the data analysis process,

---

[*] Corresponding author: pipieta@agh.edu.pl

such as pre-processing, discretization, reduction of attributes, classification, etc. To discover the correct and valuable patterns hidden in the data sets, there is a strong need not only to understand the data being processed appropriately, but also to adequately pre-process it. However, if done carelessly, pre-processing can reduce the quality of the result or cause unexpected results in the future. The data mining process using systems based on the rough sets theory consists of a series of successive stages. Most of them [19] are the following: discretization which allows transforming continuous values of attributes to discrete ones, reduction of attributes (called equivalently: feature selection) which decrease the amount of processing data without any impact on the equivalence relation. Additionally, the toolkit systems compared in this paper provide useful features that support knowledge discovery like classification based on induced rules (some algorithms like exhaustive, genetic approach, LEM2 etc) from data, or decision trees, neural network.

2.1 Rough Set Exploration System (RSES)

RSES [20, 21, 22] is a tool that is based only on the classical rough set theory. It employs RSES-lib library for computations. Both library and GUI are designed and implemented at the Group of Logic, Institute of Mathematics, Warsaw University and the Group of Computer Science, Institute of Mathematics, University of Rzeszów, Poland.

RSES allows the user to perform complex experiments on decision tables while providing a simple GUI interface. According to the authors, the data to be processed by the tool should not exceed 30,000 objects. However, this, in fact, is not an absolute value, as it depends on the hardware and computing capabilities of a personal computer.

The algorithms available in RSES and supporting the tasks related to data analysis are as described in the following subsections.

2.1.1 Discretization

Discretization of continuous (real) attribute values of decision table – generation of cuts:
- local method (including the processing of nominal values)
- global method

2.1.2 Reduction of attributes

Reduction of attributes is available using:
- dynamic method – dynamic reducts i.e. reducts that remain to be such for many subtables of the original decision table
- non-dynamic such as exhaustive or genetic algorithm with the possibility of setting additional parameters, i.e. Full or Object discernibility, in case of the genetic algorithm: high/normal/low speed.

In addition to assigning reducts from the attributes in the data set, the core calculation was also implemented.

2.1.3 Generation of rules

Generation of rules (which later can be used for the classification problem):

- exhaustive algorithm
- genetic algorithm
- covering algorithm
- LEM2 algorithm.

Both for the method of covering and LEM2 a dedicated coefficient determining the degree of coverage can be set (coverage parameter). Algorithms based on the genetic and the exhaustive approach have the same coefficients for settings, as listed earlier.

2.1.4 Linear combinations

Regarding linear combinations, is the following are available:
- generation of linear combinations
- adding linear combinations as new attributes (one of the methods that create new attributes, alongside well-known ones: adding manually an attribute and its value based on expert knowledge in RSES).

2.1.5 Other methods

There are also other methods being the part of the data analysis process, such as:
- Local Transfer Function Classifier (LTF-C) based on radial neural network architecture (RBF)
- Missing Template Decomposition Classifier (MTD-C)
- Decomposition Tree.

Most of these methods can be implemented in classification problems: a model that has been previously prepared from these existing rules may be used in training.

RSES enables users to take advantage of practical possibilities of changing an existing data set in terms of attribute reduction, exchange of the attribute's significance, i.e. change of conditional attribute for a decision attribute (and vice versa), division of the decision table into smaller units. Later, such a division of the decision table is involved in process of training, testing the regularity and usability of the model in the classification problem, i.e. decision tree, neural network or others. Developers of the systems have implemented numerous options improving the quality of data analysis – statistics of useful parameters based on the rough set theory such as the calculation of a positive area or the comparison of either single attribute values from the table, or between many attributes. In this case, numerical values are obtained, defining the range of values, standard deviation, average value, type and status of the particular attribute (CONDITION/DECISION).

2.2 Rough Set Data Explorer (ROSE2)

ROSE2 [23, 24, 25] has been created at the Laboratory of Intelligent Decision Support Systems of the Institute of Computing Science in Poznań, Poland. It provides both basic and advanced data analysis methods based on the classical rough set theory and the variable precision rough set theory. It provides its user with a GUI window without command line, although less intuitive in use than RSES, nevertheless ROSE2 has more methods implemented than the RSES system.

2.2.1 Pre-processing stage

Pre-processing stage in ROSE2 consists in completing missing objects by using the most frequent value for the attribute or removing such objects.

2.2.2 Discretization

The following discretization methods are available:
- local method (entropy-based method)
- global method
- derived from norms (provided by a domain expert through a previously specified file)

2.2.3 Reduction of attributes

There are several methods of reduction of attributes available:
- calculation of reducts and of core (one of many possible options)
- based on lattice term (Lattice Search)
- based on discernibility matrix
- heuristic search
- manual search

2.2.4 Rule Induction Methods

There are several rule induction methods available:
- LEM2 algorithm (Basic minimal covering)
- ModLEM algorithm (Extended minimal covering) with possible evaluation measures: Laplace and Entropy-based
- algorithm satisfying the demands of the user (Satisfactory Description) with possible properties of rule to set as maximal length and minimal strength of generated rule or minimum discrimination level.

ROSE extends the methods of the classic rough set theory by a similarity relation. It is used instead of the indiscernibility relation and provides sufficient help when dealing with continuous values of examined data.

2.2.5 Options available for this extension

ROSE2 provides also some options that are available for extension:
- Similarity learning
- Similarity-based approximation
- Similarity-based minimal covering algorithm for rule induction

2.3 Rough Set Toolkit for Analysis of Data (ROSETTA)

ROSETTA [26, 27, 28, 29] is being developed by researchers of the University environment in Uppsala. Apart from the basic functionality as the import and export of data, it supports also the ODBC interface for extracting data from databases. The tool provides a GUI as well as a build-in command line.

ROSETTA uses the RSES library for elementary computation and adds its own implementations of well-known algorithms based on the classical rough sets theory and its extensions, namely the variable-precision rough sets approximation and based on tolerance relations. It supports both unsupervised and supervised learning methods, as well as user-defined notions of discernibility. To facilitate statistical analysis, the tool can generate confusion matrices, ROC curves and calibration curves.

2.3.1 Pre-processing stage

Pre-processing stage in ROSETTA consists in:
- completing missing objects using the mean value
- completing missing objects using the conditioned mean value
- completing missing objects using the combinatorial completion
- completing missing objects using the conditioned combinatorial completion
- removing missing objects

2.3.2 Discretization

The following discretization methods are available:
- naive algorithm
- semi-naive algorithm
- Equal frequency binning based method
- Entropy-based method (MDL algorithm)
- Boolean reasoning algorithm (OrthogonalScaler)
- derived from file with cuts (Orthogonal File Scaler)
- derived from template (Template Scaler)
- manual method

2.3.3 Reduction of attributes

ROSETTA supports several available methods of reduction of attributes:
- genetic algorithm
- Johnson algorithm
- Holte 1R
- manual method

2.3.4 Other methods

The considered tool also provides numerous methods of shortening, filtering the result to obtain even better quality, or testing the relationship between results, such as:
- Basic Filtering
- Cost Filtering
- Performance filtering
- Basic Shortening
- Rule Basic Filtering
- Quality Rule Filtering
- Quality Rule Loop Filtering

2.4 Waikato Environment for Knowledge Analysis (WEKA)

WEKA [30, 31] is a complex tool for data analysis and predictive modelling. The supported methods are implemented in the form of specially prepared filters to perform experiments. WEKA is based not only on the rough set theory, but also on other mathematical formalisms (in the form of implemented packages) or their extensions, such as fuzzy-rough extension. Thanks to a GUI consisting of Experimenter and Explorer windows, WEKA helps a user through a series of stages of data mining: pre-processing, feature selection, instance selection and classification giving many practical functionalities outside the classical rough sets theory.

2.4.1 Pre-processing stage

Pre-processing stage in WEKA consists in:

- removing objects with missing values of attributes using Weka's existing instance filters (RemoveMissing filter)
- replacing missing objects with the mean value (RemoveMissing filter)
- fuzzy-rough method: interval-valued approach named IVFRFS

Adding AddConditionalNoise/AddClassNoise filters specially prepared for this purpose and handling noise with available fuzzy-rough approaches called the Vaguely Quantified Rough Set (VQRS) or Ordered Weighted Average (OWA) is based on the fuzzy-rough set. In contrast to the previously used tools, the deliberate addition of noise in the data allows checking the defect of the tested algorithm in different environments. The robustness of the algorithm is one of the most important criteria along which we can decide whether to use it on a larger scale and in what conditions.

2.4.2 Reduction of attributes

In the case of reduction of attributes, the available methods relating to the rough set are as follows:

- based on Ant Search
- based on Genetic Search
- Johnson's algorithm
- PSO Search (Particle Swarm Optimization algorithm)
- SAT Search
- QuickReduct
- exhaustive algorithm

It is worth noting that there are special WEKA tool measures to be set for each method referring to RST, which should be set before using the given algorithm. Some of them are given below:

- boundary region
- fuzzy discernibility matrix
- VQRS/WeakVQRS (vaguely-quantified approach to FRST)
- The fuzzy-rough dependency measure (Gamma/WeakGamma measure)
- discernibility function (DiscernibilityF)
- fuzzy Entropy
- fuzzy gain ratio

Apart from many modules for varied data analysis, *i.e.* filters, shortening ratio, *etc*, WEKA contains as well numerous classifiers.

2.4.3 Classification algorithms

The following classification algorithms are available:

- FuzzyRoughNN
- FuzzyNN
- DiscernibilityClassifier
- QuickRules (hybrid feature selection/rule induction algorithm)
- VQRules
- VQNN

2.5 Rough Set Data Analysis Framework (jMAF)

JMAF [32, 33] is a tool suitable for analysis of data gathered in the decision table with predefined profiles that are based on background knowledge about ordinal evaluations of objects from a universe, and about monotonic relationships between these evaluations. The tool draws from jRS library, which is the computational core for all methods in jMAF. The tool supports the methods that are the extensions of the classical rough sets theory, such as the methods based on DRSA (Dominance-based Rough Set Approach) and VC-DRSA (Variable Consistency-based Rough Set Approach) approaches. Compared to other systems, it implements fewer algorithms. However, as a small and easy-to-use tool, jMAF is dedicated for non-professionals.

2.5.1 Calculation of Dominance Cones (P-Dominating Sets and P-Dominated Sets)

The calculation of Dominance Cones is performed from following statements:

- $D_P^+(x) = \{y \in U : yD_P\,x\}$,
  a set of objects dominating x [32]
- $D_P^-(x) = \{y \in U : xD_P\,y\}$,
  a set of objects dominated by x [32]

2.5.2 Calculation of approximation

Calculation of approximation of upward and downward unions of decision classes is performed by granules (dominance cones) of knowledge generated by attribute criteria [32].

2.5.3 Induction algorithm available

The available induction algorithm is VC-DOMLEM algorithm.

2.5.4 Classification methods

There are two available types of classification methods available in jMAF [33], based on DRSA and based on VC-DRSA.

2.6 Comparison of selected aspects of the tools

Table 1 gives a brief comparison of the selected aspects of the presented tools, such as the programming language in which the tool was implemented, the supported operating systems, availability of the open-source license, installation requirement, as well as the possible extensions of the rough sets theory implemented in the considered tools.

Table 2, in turn, provides a list of various input and output formats available in each of the presented tools. This can be helpful in selecting the proper tool, depending on the available input data formats or required output formats.

3 Conclusion

The paper presents selected features of the existing ready-made systems that facilitate conducting in-depth data analysis in the form of decision tables using methods based on their complexity on the rough sets approach.

Table 1. Comparison of the presented tools.

Tool	Language	OS			License	Cross-validation	Installation	Approach				
		Windows	Linux	Mac	Open Source			CRSA	VP-CRSA	DRSA	VC-DRSA	FRSA
RSES	C++/Java	✓	✓	✗	✓	✓	✓	✓	✗	✗	✗	✗
ROSE2	C++	✓	✗	✗	✓	✓	✓	✓	✓	✗	✗	✗
ROSETTA	C++	✓	✓	✓	(✓)	✓	✓	✓	✓	✗	✗	✗
WEKA	Java	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✓
jMAF	Java	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✗

✓ - *Yes*, (✓) - *with some restrictions or limitation*, ✗- *No*
CRSA – *classical rough sets approach*, VP-CRSA – *variable precision CRSA*, DRSA – *dominance-based rough sets approach*,
VC-DRSA – *variable-consistency DRSA*, FRSA – *fuzzy-rough sets approach;*
Columns: OS – *operating system*, License – *open source or with limitation*, Cross-validation – *method of validation*, Installation - *required/not required,* Approach – *the possible extensions of rough sets theory implemented in appropriate tools*

Table 2. Input/output formats available in the presented tools.

Tool	INPUT FORMAT							OUTPUT FORMAT									
	ARFF	ISF	ROS	TAB	XLS	CSV	M	CPP	PL	XML	ISF	ROS	IG	DF	CM	RLS	CSV
RSES	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
ROSE2	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗
ROSETTA	✗	✗	✓	✗	✓	✓	✗	✓	✓	✓	✗	✓	✓	✓	✓	✗	✗
WEKA	✓	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓
jMAF	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✓	✓	✓

ARFF – *WEKA data format (.arff)*, ROS – *Rosetta data plain format (.ros)*, TAB – *RSES (RSES 1.0) data format (.tab)*,
ISF - *Information System File (.isf)*, XLS – *Excel format (.xls)*, CSV - *comma-separated values (.csv)*, M – *Matlab ASCII format (.m)*,
CPP – *C++ format file (.cpp)*, PL – *Prolog format (.pl)*, XML - *Extensible Markup Language (.xml)*, RLS – *Rule File Format;*
other elements of output: IG – Indiscernibility graph, DF – Discernibility function, CM – confusion matrix

The presented tools are easy to use and allow at the same time executing non-trivial experiments. In particular, a friendly GUI affects the RSES utility (each tool has them) allowing fast transition between windows. Using WEKA requires moving between the Explorer and Experimenter windows.

ROSE2 and ROSETTA can be perceived as the least intuitive and hence unfriendly, although the latter is characterised by the largest number of available options, as well as the number of methods implemented in WEKA (fuzzy-rough approach). JMAF, as the only tool, implements the classic RST expansion approach, called DRSA and Variable-Precision DRSA, which treats the indiscernibility relation in a specific way. WEKA offers, besides the rough approach, also the fuzzy-rough one. Certain methods implemented in one tool can be implemented in another, i.e. feature selection methods (determination of reducts) RSES and ROSETTA, with ROSETTA offering many more (both ROSETTA and RSES use the same library to perform calculations). Dominance-based RST in jMAF is very limited in the sense of functionality (computation, i.e. P-Dominance and P-Dominated sets, approximations, etc, induction algorithm VC-DOMLEM).

Exchange of data between tools for in-depth analysis (the use of algorithms not implemented in the current tool but others) is burdened with the drawback of the various file formats established by the developers. This

is because there is no way to use an .arff file containing a well-known data set in the required version of ROSE2. As part of a significant range, ROSETTA is a collection of data from, inter alia, Oracle, SQL Server, XLS, etc. Within the mentioned range, the ROSETTA tool is the most adapted, which extends the import of files with datasets originating, among others, from recognised databases.

Our goal of providing a comparative overview of the rough set toolkit systems was accomplished by highlighting selected system features, i.e. available algorithms, license or basic requirements. However, this does not exhaust the rich subject matter and requires further study.

References

1. Z. Pawlak. Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing, Dordrecht, 1991

2. W. Dauben, Joseph. (2005), Georg Cantor, paper on the Foundations of A General Set Theory (1883), 600-612

3. Ziarko W. (2005) Probabilistic Rough Sets. In: Ślęzak D., Wang G., Szczuka M., Düntsch I., Yao Y. (eds) Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. RSFDGrC 2005. Lecture Notes in Computer Science, vol 3641. Springer, Berlin, Heidelberg

4. D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–209

5. D. Dubois, H. Prade, Putting rough sets and fuzzy sets together, in: R. Slowinski (Ed.), Intelligent Decision Support: Handbook ofApplications and Advances of the Sets Theory, Kluwer, Dordrecht, 1992, pp. 203–232

6. Greco, S., Matarazzo, B., Słowiński, R.: Rough sets theory for multi-criteria decision analysis. European Journal of Operational Research, 129, 1 (2001) 1–47

7. Greco, S., Matarazzo, B., Słowiński, R.: Multicriteria classification by dominance-based rough set approach. In: W.Kloesgen and J.Zytkow (eds.), Handbook of Data Mining and Knowledge Discovery, Oxford University Press, New York, 2002

8. Słowiński, R., Greco, S., Matarazzo, B.: Rough set based decision support. Chapter 16 [in]: E.K. Burke and G. Kendall (eds.), Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques, Springer-Verlag , New York (2005) 475–527

9. Greco, S., B. Matarazzo, R. Slowinski and J. Stefanowski: Variable consistency model of dominance-based rough set approach. In W.Ziarko, Y.Yao (eds.): Rough Sets and Current Trends in Computing. Lecture Notes in Artificial Intelligence 2005 (2001) 170–181. Springer-Verlag

10. J.G. Bazan, M. Szczuka, "The Rough Set Exploration System", Transactions on Rough Sets III, ser. LNCS, vol. 3400, Springer, Heidelberg, 2005, pp. 37-56.

11. M. Sudha, A. Kumaravel, Performance Comparison based on Attribute Selection Tools for Data Mining, Indian Journal of Science and Technology, Vol 7, November 2014, pp. 61–65

12. M. Mohamad, A. Selamat (2018), An Analysis of Rough Set-Based Application Tools in the Decision-Making Process, in: Saeed F., Gazem N., Patnaik S., Saed Balaid A., Mohammed F. (eds) Recent Trends in Information and Communication Technology. IRICT 2017. Lecture Notes on Data Engineering and Communications Technologies, vol 5. Springer, Cham

13. C.R. Kavitha, T. Mahalekshmi, Performance Comparison based on Attribute Selection Techniques of WEKA and ROSE Tools, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, October 2016

14. A. Janusz, M. Szczuka, S. Stawicki, D. Ślęzak, Rough Set Tools for Practical Data Exploration, Proc: Rough Sets and Knowledge Technology: 10th International Conference, RSKT 2015, Held as Part of the International Joint Conference on Rough Sets, IJCRS 2015, Tianjin, China, November 20-23, 2015, Proceedings, pp.77-86

15. L.S. Riza, A. Janusz., C. Bergmeir, C. Cornelis, F. Herrera, D. Ślezak, J. M. Benítez (2014), Implementing algorithms of rough set theory and fuzzy rough set theory in the R package "RoughSets", Inf. Sci., 287, pp. 68-89

16. Zain Abbas, Aqil Burney, A Survey of Software Packages Used for Rough Set Analysis, Journal of Computer and Communications, 2016, vol. 4, pp. 10-18

17. Alekhya Cherukri, Madhuri Doguparthi, Comprehensive Analysis of various rough set tools for data mining, International Journal of Advances in Electronics and Computer Science, Vol. 4, April 2017

18. Pięta P., Analysis of data mining algorthms based on Rough Sets theory, Master Thesis (Supervisor : Prof. Tomasz Szmuc), AGH University of Science and Technology, Cracow, Poland, 2018

19. J. Komorowski, Z. Pawlak., L. Polkowski, A. Skowron, Rough Sets: A tutorial, 1997

20. RSES_doc_eng.pdf, User Guide's

21. J. G. Bazan, M. Szczuka, J. Wróblewski (2002) A New Version of Rough Set Exploration System, In: Alpigini J.J., Peters J.F., Skowron A., Zhong N. (eds) Rough Sets and Current Trends in Computing. RSCTC 2002. Lecture Notes in Computer Science, vol 2475. Springer

22. http://logic.mimuw.edu.pl/~rses/

23. http://idss.cs.put.poznan.pl/site/rose.html

24. B. Predki, R. Slowinski, J. Stefanowski, R.

Susmaga, S. Wilk: ROSE - Software Implementation of the Rough Set Theory. In: L.Polkowski, A.Skowron, eds. Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, vol. 1424. Springer-Verlag, Berlin (1998), 605-608

25. B. Predki, S. Wilk: Rough Set Based Data Exploration Using ROSE System. In: Z.W.Ras, A.Skowron, eds. Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence, vol. 1609. Springer-Verlag, Berlin (1999), 172-180

26. A. Øhrn, J. Komorowski (1997), ROSETTA: A Rough Set Toolkit for Analysis of Data, Proc. Third International Joint Conference on Information Sciences, Fifth International Workshop on Rough Sets and Soft Computing (RSSC'97), Durham, NC, USA, March 1-5, Vol. 3, pp. 403-407

27. A. Øhrn, J. Komorowski, A. Skowron, P. Synak, The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets: The ROSETTA System, Rough Sets in Knowledge Discovery 1: Methodology and Applications, L. Polkowski and A. Skowron (eds.), Studies in Fuzziness and Soft Computing, Vol. 18, Chapter 19, pp. 376-399

28. A. Øhrn (2000), ROSETTA Technical Reference Manual, Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

29. http://bioinf.icm.uu.se/rosetta/downloads.php

30. R. Jensen, Tutorial on fuzzy-rough data mining using Weka, http://users.aber.ac.uk/rkj/Weka.pdf

31. http://users.aber.ac.uk/rkj/book/wekafull.jar

32. J. Błaszczyński, S. Greco, B. Matarazzo, R. Słowiński, M. Szeląg, jMAF - Dominance-Based Rough Set Data Analysis Framework, Rough Sets and Intelligent Systems - Professor Zdzisław Pawlak in Memoriam, Springer, 2012

33. http://www.cs.put.poznan.pl/jblaszczynski/Site/jRS.html