About New Version of RSDS System

Zbigniew Suraj and Piotr Grochowalski

Institute of Computer Science
University of Rzeszów, Rzeszów, Poland
{zbigniew.suraj,piotrg}@ur.edu.pl

Abstract. The aim of this paper is to present a new version of a bibliographic database system - Rough Set Database System (RSDS). The RSDS system, among others, includes bibliographic descriptions of publications on rough set theory and its applications. This system is also an experimental environment for research related to the processing of bibliographic data using the domain knowledge and the related information retrieval.

Keywords: rough sets, data mining, knowledge discovery, pattern recognition, database systems.

1 Introduction

The presented RSDS system is a bibliographic system that includes bibliographical references aimed at disseminating information on publications on rough set theory and its applications [4–8]. The system is available for free at http://rsds. univ.rzeszow.pl. Currently about 4000 bibliographical descriptions of publications are collected in its database.

The RSDS system is also an experimental environment for carrying out research related to the broadly defined information processing based on the methods and techniques in the field of ontology and rough sets. In addition, it enables one to analyze the data contained in the database by using the advanced statistical and graphical methods and techniques.

Apart from the bibliographical descriptions, it contains additionally:

– information on software related to the theory and applications of rough sets,
– bibliographies of people who render outstanding services to the development of rough set theory and its applications,
– personal details about the authors of publications whose descriptions are included in the database of this system.

The system was developed in the client-server technology, i.e., the data for the system and mechanisms for handling such data are running on the server and the user with the help of the web browser is able to access the resources.

The work on the system began in 2002 and it is continued. During the works there were developed two versions of the system that have been made available for users [9]-[11],[14]-[15],[19]. In March 2013, the third version of the system
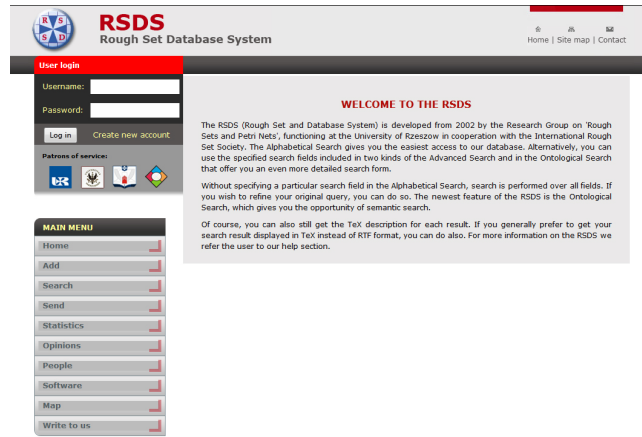
Fig. 1. The main window of the system.

was released. It has been completely rebuilt, allowing the introduction of larger number of facilities for users of the system with the use of modern technology.

In the current version the following changes was introduced:

- exchange and reorganization of the engine of the system (CMS - Drupal),
- rebuilding the website design,
- rebuilding and upgrading the functionalities of the system,
- increasing the role of the administrator(s) of the system - admin panel,
- larger number of facilities for registered users - user panel,
- introduction of the status of data and the possibility of its modification,
- system-to-user communication via e-mail (on important issues).

The rest of the paper is organized as follows. Section 2 describes the logical structure of the RSDS system. Functionalities of the system are presented in Section 3. Section 4 provides description of data for the system. Section 5 gives the system requirements. The future plans for the RSDS system are discussed in section 6.

2   The Logical Structure of the System

The RSDS system structure can be divided into four functional layers. Each of these layers includes modules and with the help of these modules the layers meet specified tasks (see Figure 2):

- The presentation layer with the graphical interface module.
- The application layer with the modules of login, add/edit, search, graph and statistical, download, auxiliary (biographies of people, software, maps).
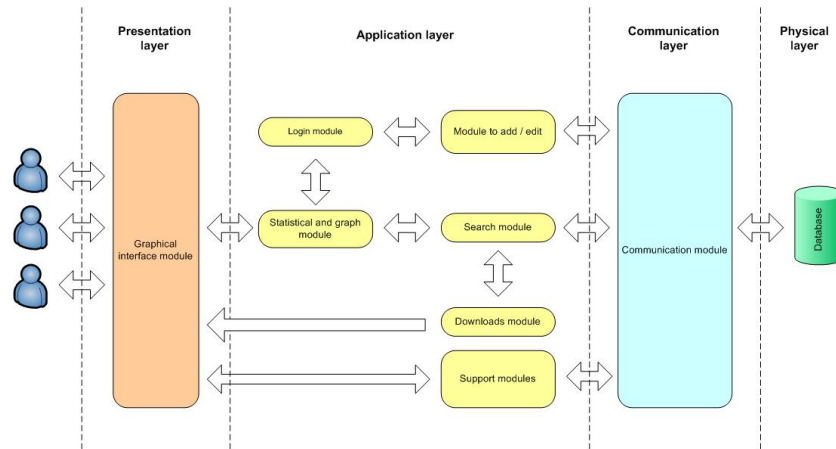- The communication layer with the module of communication with the database.

Fig. 2. The logical structure of system.

– The physical layer containing the database.

The purpose of the modules from the presentation layer is to communicate with the user with the usage of created interface.

In the application layer there are the modules implementing the main functionalities of the system. The login module is responsible for the correct handling of the login/logoff process of the users and storing information about the users logged in the system. The add and edit module supports the process of implementing the new data into the system or editing already existing ones. It will ensure the correctness of the input data and its correct assignment to the user-owner. The search module performs the search process of the publications descriptions meeting the criteria set by the user. The operation of this module has been improved by implementing the results of the research connected with the information retrieval in it. The graphic and statistical module is designed not only for the analysis of data in the system but also for the analysis of the activity of the users. The analysis is carried out in various aspects, such as publications, the authors of the publication and the relationships between them. The purpose of this module is the presentation of the results of these analyzes. The download module provides the users with different options of retrieving data from the system in the form of prepared bibliography. Auxiliary modules extend the basic functionality of the system to the biographies of people who render outstanding services to the development of rough set theory and its applications, and descriptions of available software related to the theory of rough sets and also a world map showing where the given problem is growing solved.

The communication layer possesses a module that is responsible for the proper communication with the database, which stores the data for the system.

The physical layer includes a relational database in which the data are stored and presented in the system.

3   System Functionalities

The basic functionalities of the RSDS system include:

 – Adding new data.
 – Editing existing data.
 – Data search.
 – Registration of users in the system.
 – Saving data to a file.
 – Sending data files to the administrator.
 – Service of user comments.
 – Statistics, analysis of statistical graphs, determining the Pawlak number of
   the first and second kind, classifier of publications.
 – Help.

Capabilities and the content of the RSDS system are constantly extended.

In order to store the held information in the simplest form and to exclude redundancy (redundancy of data), the data for the RSDS system are stored in a relational database. The database structure is based on the BibTeX format [22]. Well defined and uniform structures of the description decided about its choice. Additionally, the possibility of getting the bibliographical descriptions in the BibTeX format included in the system, allowing one to automatically generate bibliographies and attach them to the prepared publication, has been added to the system.

To share data, the system should be first equipped with it. Data entry and other operations allowing a modification of data, require user authentication by logging on to the system. New users, in order to get access to the full functionality of the system, need to register.

Data entry can be performed by two independent pathways:

 – by predefined forms;
 – with the usage of software able to read the files in the BibTeX format and
   storing information in the system in the appropriate way.

The usage of predefined forms allows registered users to introduce new data into the system individually. If one does not want to do this action individually or intends to enter a large number of new data, he or she can send the data to the system administrator and then administrator with the use of appropriate software will enter the received data into the system. The advantage of the individual data input by users is that they are assigned to them. In such case the users are authorized to edit the data in the future. This possibility is available only for registered users, in order to avoid entering incorrect information into the system. System with published descriptions (data) provides various options of searching.

Searching for information on the RSDS system was implemented so far in two main ways:

– Alphabetical search by certain keywords, such as titles of publications, their authors, editors, conference names, magazines, or year;
– Advanced search based on specified criteria, which sought description of publication has to fulfill.

Each of the currently available options of finding information in the RSDS system has both positive and negative aspects. Alphabetical search works when you know for example: the author of sought publication, the name of the journal in which the publication was published, who published a publication, or when one knows the year of the publication. The weaker part of this search method, however, is that in the absence of precise information about sought publication the system provides the large number of publication descriptions meeting the search criteria, which often have to be further analyzed by a painstaking selection process. However, during advanced search, the user defines the criteria which have to be fulfilled for the sought publication and, depending on the accuracy of the selection of these criteria, he or she obtains more or less adequate results. The problem of the further selection of obtained results still exists in many cases. This process involves directing a user's query to the database system to find the matching (appropriate) data (publication) for your search pattern. This matching is based on the finding of exact pattern in the data in the database. If the matching to the pattern data is found, it will be annexed to the result. This is repeated for all data located in the database. Then the edited result set (publication) is sent to the user (see Figure 3).
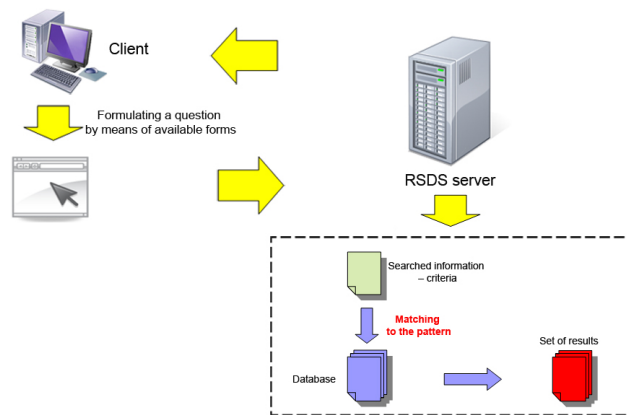

Fig. 3. The current course of the process of searching for the information in the RSDS system.

Finding information in the system process supplemented with the additional knowledge (using ontology and methods of rough set theory) is presented in the general form shown in Figure 4. This process in comparison with the mapping

specified search pattern is different in this that the system after receiving the user's query accede to its confrontation with the information included in the system [1–3, 17, 21]. This is done as follows: from the resources of the system are retrieved the specific ontologies for publication and they are verified for their belonging to the inner circle of the sought information in the domain of general ontology (taking into account the relationships between different types of concept). In the case of positive verification the approximations are appointed for a specific publication. They are then aggregated into a single value representing the total adjustment to the published information. When appointed value is greater than the threshold value, the given publication will be included in the result set. This process is repeated for all the data (publications). Next, after the result set processing (division of publication into groups) it is sent to the user. The developed methodology has been implemented in the system and made available to the users in the ontological search section. In order to streamline the process of building queries, for the system has been prepared the editor to assist the creation of query-by-user. The functionalities of this editor are: auto completion of entered concepts, defined relational operators such as AND, OR, NOT, and taking into account the priority of the process by using parentheses.
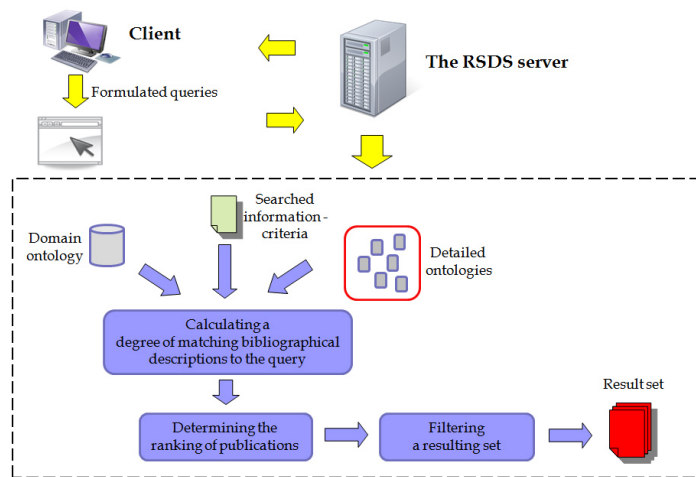

Fig. 4. The process of searching information in the RSDS system based on additional knowledge.

In order to minimize the time complexity of the process, the determining of the detailed ontology is carried out once. Determination of the detailed ontology is a part of the process relying on the preparation for the implementation of the system of bibliographic information for retrieval based on domain knowledge. This process can be carried out in several steps, which also has been shown (see Figure 5). The first step is to develop the general ontology by the domain

expert. Then, on the basis of the system resources, supported by the domain knowledge (general ontology) the detailed ontologies are determined. In the process of the detailed ontology generation new concepts or relationships between concepts can occur. In that case the system in cooperation with the domain expert is able to include such concepts in the general ontology, thus creating extended domain ontology. The final step in this process is to determine the degrees of bibliographical descriptions match to the elementary concepts from the general ontology. This is done because of the minimization of time required for determining the response to a query from user.
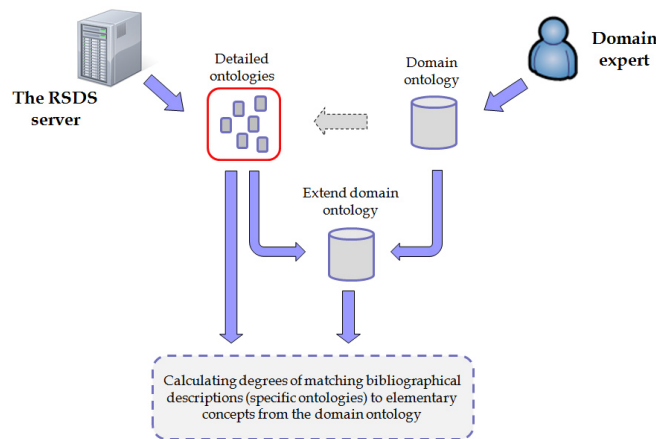

Fig. 5. The process of preparing the bibliographic system for searching information based on additional knowledge.

The system, in addition to bibliographical data (represented in the form of descriptive or BibTeX format) provides a fairly wide range of different statistics and the results of the analyzes of the data [12, 13, 16, 18].

In the carried out research the data collected from the year 1981 to the present are taken into account. They are analyzed in two ways: statistical and graphic. In relation to statistic data are processed in different compartments of time:

1. till the defined periods, five-years in the incremental relation, i.e., 1981-1985, 1981-1990, 1981-1995, etc.
2. in certain five-years, i.e., 1981-1985, 1986-1990, 1991-1995, etc.

In terms of graphic analysis it is made on the basis of the defined cooperation CG graph. The graph vertices are the authors of the publications included in the bibliographic system. Two vertices are connected by the edge when two authors have written at least one common publication.

In statistical analysis were determined the following values characterizing the examined data set, i.e., the number of authors in respect to the number of

written publication, various kinds of means, standard deviations associated with a particular medium, the number of works in respect to the number of authors creating them.

Graph analysis of the defined CG graph lies in its overall analysis that is the appointment of the average degree of vertex, of isolated vertices, etc. CG graph after the rejection of isolated vertices has been further analyzed. It was based on the determination of the components of the graph and the analysis of the largest component. Determination of components allowed on determining the groups of authors writing joint publications - the cooperating authors. These groups are reflected in reality. The analyzed parameters of the largest component allows for its accurate interpretation.

Additionally, the system allows users to read the Pawlak number of the first and second kind, which values indicate the strength of the proximity of publishing author's work with prof. Z. Pawlak, i.e., the less value of the Pawlak number represents the stronger relation between the author of published work and Professor Pawlak [18, 20].

All analyzes are performed dynamically, i.e., the calculation of parameters is taking into account any change in the data collected in system.

4   Input-output Data

Bibliographical descriptions are described in the system according to specifications of BibTeX [22]. This means that the description of each publication is divided into elements defined by BibTeX, such as title, publisher, year of publication, the keywords, abstract, etc. The prepared descriptions are placed in a relational database. Each component is stored in the database structure defined as a string, the importance of which, unfortunately, neither database nor database languages can understand. An example of a bibliographic description located in the system is presented in Table 1. Descriptions of publications are formulated in English.

Table 1. The example of bibliographic description in the BibTeX format.

```
@INPROCEEDINGS{,
author     = {Hu, Xiaohua Tony and Cercone, Nick},
title      = {Mining knowledge rules from databases: A rough set approach},
booktitle  = {Proceedings of the 12th International Conference on Data Engineering},
conference = {International Conference on Data Engineering (CDE), New Orleans,
             USA},
pages      = {96-105},
publisher  = {IEEE Computer Society Press},
address    = {Los Alamitos, CA, USA},
month      = {February},
year       = 1996,
isbn       = {0-8186-7240-4},
abstract   = {In this paper, the principle and experimental results of an attribute-
             oriented rough set approach for knowledge discovery in databases are de-
             scribed. ... },
keywords   = {knowledge mining and discovery},
}
```

5   System Requirements

The RSDS system can be run on any computer that is connected to the Internet. The computer must have an operating system equipped with web browser. The presented above requirements must be met, as the RSDS system is an online system and requires a permanent connection to the Internet. In addition, a web browser, in which the system will be running, must support JavaScript scripting language, CSS style sheets and cookies. The system has been tested with the following browsers: Internet Explorer 9, Mozilla Firefox 17, and Chrome 23.

6   Plans for the Future

The directions of further research and work related to the system will be:

- development of a method for the formal verification of the correctness of the defined relations in the general ontology,
- increase of the degree of efficiency of information retrieval,
- automation of the process of processing of the owned information,
- attempt to improve the quality of semantic analysis,
- development of new functionalities of the system increasing its features such as: automatic discovery of scientific user profile, finding new data from Internet resources, extending the analysis of owned data, etc.

References

1. Grochowalski P. and Pancerz K., The outline of an ontology for the rough set theory and its applications. In Czaja L., Penczek W., Salwicki A., Schlingloff H., Skowron A., Suraj Z., Lindemann G., Burkhard H.-D. (Eds.), *Proceedings of the Workshop on Concurrency, Specification and Programming, CS&P2008*, Gross-Vater See, Berlin, 29 September - 1 October, 2008, vol. 1-3, pp. 192–204. Informatik-Berichte, 2008.
2. Grochowalski P. and Pancerz K., The outline of an ontology for the rough set theory and its applications. *Fundamenta Informaticae*, 93(1-3):143–154, 2009.
3. Pancerz K. and Grochowalski P., Matching Ontological Subgraphs to Concepts: a Preliminary Rough Set Approach. In *Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA'2010)*, Cairo, Egypt, November 29 - December 1, 2010, pp. 1394–1399, IEEE Xplore, 2010.
4. Pawlak Z., Rough Sets. *International Journal of Computer and Information Sciences*, 11:341–356, 1982.
5. Pawlak Z., Grzymała-Busse J.W., Słowiński R., and Ziarko W., Rough Sets. *Communications of the ACM*, 38(11):88–95, November 1995.
6. Pawlak Z. and Skowron A., Rough Sets and Boolean Reasoning. *Information Sciences*, 177(1):41–73, 2007.

7. Polkowski L.T., *Rough Sets. Mathematical Foundations.* Advances in Soft Computing. Physica-Verlag, Heidelberg, 2002.

8. Skowron A. and Pal S.K. (Eds.), Special Volume: Rough Sets, Pattern Recognition and Data Mining, vol. 24(6), *Pattern Recognition Letters.* North Holland, 2003.

9. Suraj Z. and Grochowalski P., The Rough Sets Database System: An Overview. In Komorowski J., Grzymała-Busse J.W., Tsumoto S., Słowiński R. (Eds.), *Proceedings of the 4th International Conference on Rough Sets and Current Trends in Computing, RSCTC 2004*, Uppsala, Sweden, June 2004, vol. 3066, *Lecture Notes in Artificial Intelligent*, pp. 841–849, Springer-Verlag, 2004.

10. Suraj Z. and Grochowalski P., The Rough Set Database System: An Overview. Transactions on Rough Sets III, *Lecture Notes of Computer Sciences*, vol. 3400, Springer-Verlag, Berlin, pp. 190–201, 2005.

11. Suraj Z. and Grochowalski P., Functional extension of the RSDS system. In Hirano S., Inuiguchi M., Miyamoto S., Nguyen H.S., Słowiński R., Greco S., Hata Y. (Eds.), *Proceedings of the 5th International Conference on Rough Sets and Current Trends in Computing, RSCTC 2006*, Kobe, Japan, November 2006, vol. 4259, *Lecture Notes in Artificial Intelligent*, pp. 786–795. Springer-Verlag, 2006.

12. Suraj Z. and Grochowalski P., Patterns of collaborations in rough set research. In Gomez V., Bello R., Falcon R. (Eds.), *Proceedings of the International Symposium on Fuzzy and Rough Sets, ISFUROS 2006*, Santa Clara, Cuba, December 5-8, 2006, pp. 1–7.

13. Suraj Z. and Grochowalski P., Patterns of Collaborations in Rough Set Research. In Bello R., Falcon R., Pedrycz W., Kacprzyk J. (Eds.), Granular Computing: at the Junction of Fuzzy Sets and Rough Sets, *Studies in Fuzziness and Soft Computing*, vol. 224, Springer-Verlag, 2008, pp. 79–92.

14. Suraj Z. and Grochowalski P., The Rough Set Database System. Transactions on Rough Sets VIII, *Lecture Notes of Computer Sciences*, vol. 3400, Springer-Verlag, pp. 307–331, 2008.

15. Suraj Z., Grochowalski P., Garwol K., and Pancerz K., Toward intelligent searching the rough set database system: an ontological approach. In Szczuka M., Czaja L. (Eds.), *Proceedings of the CS&P'2009 Workshop*, Kraków-Przegorzały, 28-30 September, 2009, vol. 1-2, pp. 574–582, Warsaw University, 2009.

16. Suraj Z. and Grochowalski P., Some Comparative Analyses of Data in the RSDS System. In Yu J., Greco S., Lingras P., Wang G., Skowron A. (Eds.), *Proceedings of the 5th International Conference on Rough Sets and Knowledge Technology, RSKT 2010*, Beijing, China, October 15-17, 2010, *Lecture Notes in Artificial Intelligence*, vol. 6401, Springer-Verlag, pp. 8–15, 2010.

17. Suraj Z. and Grochowalski P., Toward intelligent searching the Rough Set Database System (RSDS): an ontological approach. *Fundamenta Informaticae*, 101(1-2):115–123, 2010.

18. Suraj Z., Grochowalski P., and Lew Ł., Pawlak Collaboration Graph and Its Properties. *Proceedings of the 13th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC'2011)*, Moscow, Russia, June 27-30, 2011.

19. Suraj Z. and Grochowalski P., RoSetOn: The Open Project for Ontology of Rough Sets and Related Fields. In J.T. Yao et al. (Eds.), *Proceedings of the 6th International Conference on Rough Sets and Knowledge Technology, RSKT 2011*, Banff, Canada, October 9-12, 2011, *Lecture Notes in Computer Science*, vol. 6954, Springer-Verlag, pp. 414–419, 2011.

20. Suraj Z., Grochowalski P., and Lew Ł, Discovering Patterns of Collaboration in Rough Set Research: Statistical and Graph-Theoretical Approach. In J.T. Yao et al. (Eds.), *Proceedings of the 6th International Conference on Rough Sets and Knowledge Technology, RSKT 2011*, Banff, Canada, October 9-12, 2011, *Lecture Notes in Computer Science*, vol. 6954, Springer-Verlag, pp. 238–247, 2011.

21. Suraj Z., Grochowalski P, and Pancerz K., Knowledge Representation and Automated Methods of Searching for Information in Bibliographical Data Bases: A Rough Set Approach. In Skowron A., Suraj Z. (Eds.), Rough Sets and Intelligent Systems - Professor Zdzisław Pawlak In Memoriam, *Intelligent Systems Reference Library*, vol. 43, pp. 515–538, 2012.

22. BibTeX. Available: *http://www.bibtex.org/*