American Journal of Intelligent Systems. 2011; 1(1): 37-42
DOI: 10. 5923/j.ajis.20110101.06

Knowledge Discovery from Repository of Web Information

Kristina Machova, Dominika Fodorová

Department of Cybernetics and Artificial Intelligence, Technical University of Košice, Košice, 04200, Slovakia

Abstract Paper focuses on the knowledge discovery from repository of web information and subsequent knowledge relationship discovery within the integrated data. The information repository model is described in it. The contribution introduces various approaches to knowledge relations discovery like the model creation, the exact comparison and the dynamic comparison. The implementation of the introduced approaches – WKID system – is also described. The paper contains also a comparison to similar approaches. The results of experiments with the system are described and discussed.

Keywords Semantic Web, Integration of Information, Knowledge Discovery, Relations Discovery, WKID System

1. Introduction

The Internet is a real phenomenon of these days. In spite of massive usage of the web, it suffers from many problems. The main problems of web usage according to reference[1] are the following: high recall connected with low precision, low or even zero recall, search results being web pages and not searched information, and search results depended on a used vocabulary. Solutions of these problems need intelligent approaches to information processing in semantically enriched web[6].The most popular browsers try to solve the problem with low precision of search results using page ranking and subsequent ordering the retrieved pages according to this rank. This solution supposes that web user would read only some of the first found pages and so it is important to locate the more precise web pages on the top of the resulting page list.

The second problem with low or even zero recall requires refining of web searching and understanding of user needs. A solution to this problem can be represented by semantic search according[2]. The semantic search can be an appropriate tool for making search results independent from used vocabulary. Mainly such semantic technologies as metadata (XML, RDF documents) and ontology (OWL documents) can be used for these purposes[9]. Ontology can represent the vocabulary of all words used in tags of RDF documents. Understanding of user needs in the frame of the semantic web requires also information integration according[3]. Information integration is an "ingredient" of semantic search, which would be able to recognize, that two pieces of information presented in various forms have the samemeaning. This paper focuses on information integration from RDF web documents.

The last problem is related with the fact that web search results are provided only in the form of web pages. People usually start searching web because of their information needs. They search for exact information instead of only web pages the information is buried somewhere within them. One attempt to cope with this problem is represented by the Wolfram Alpha system in reference[10], which formulated "answers" on web user demands in the form of selected information extracted from web pages. It has some drawbacks. For example, it is strong only in some domains (e.g. mathematics, geography and economics) and it needs the creation of large databases before. This creation means rather time consuming preparation activity.

2. Information Repository Model

The repository model is a system, which enables data storage and subsequent access to these data using queries. These queries can be based on specialization (which object corresponds to conditions from query) as well as on generalization (which facts are valid for all objects corresponding to queries). This system uses a binary matrix for representation of attributes and their values.

Some data are served from a source z from Z. Let us assume the scheme of each source is $S_z = (A_z, F_z)$, which covers at least a list of attributes A_z and functional relations

F_z (A_z x A_z) between attributes. Let these data are represented in the form of attribute – value pairs from a range of values A_z x D_z, where D_z represents all values (e from E) covered by the source z.

The data of this source can be represented as an instance of functional relations f from F_z using implications $e_i \rightarrow e_j$ be-

* Corresponding author:
kristina.machova@tuke.sk（Kristina Machova）
Published online at http://journal.sapub.org/ajis

tween elements. These implications for each source *l* can be expressed using a binary matrix of the repository

$\Phi_l = [\Phi^l_{ij}]$ defined in equation (1):

$$\Phi_{ij}^{1} = 1\ if\ z_i\ \mathrm{cov}ers\ e_i \rightarrow e_i\left(0\ else\right) \qquad (1)$$

Analogically, a matrix Δ_l = [δ^l_{ij}] of active domains of the attributes of the source z_l can be expressed as in equation (2):

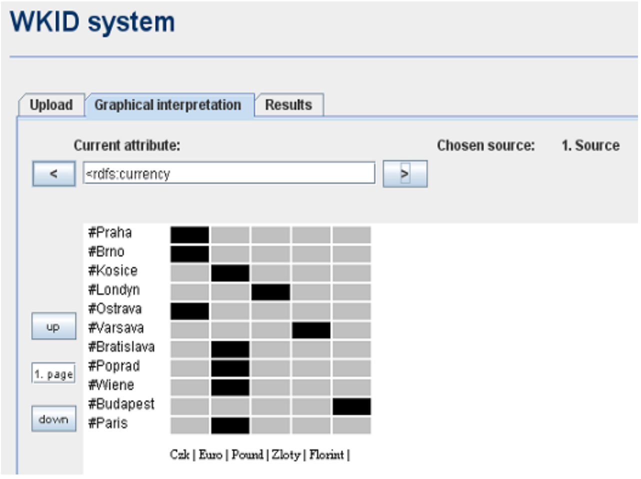
$$\Phi_{ij}^{l} = 1\ if\ e_i = A_{ij}\left(0\ else\right) \qquad (2)$$

Each non zero cell of the matrix $\Phi^l_{ij} > 0$ is an instance of some functional relation f = ($A_i \rightarrow A_j$) Є F_l.

The information repository model (illustrated in Figure 1) is a model of obtained data and information representation. These data are distilled from RDF and XML sources. Data from each source can be represented by a binary repository matrix. This repository matrix contains all attributes' values from the actual source. The attribute – value pairs can be read from this matrix. For example, the attribute ”elements” has values Prague, Brno, Košice, Bratislava, Poprad, Vienna, Budapest and Paris. The attribute “capital city” has values: yes and no. The attribute “currency” has values: CZK, Euro and Forint and so on. Some sell has the value „1“in some sub matrix if this town is capital city, or its currency is Euro and so on. If not, this cell has value “0”.

Sub-matrixes located on the main diagonal are generated by the WKID System. They represent some meta information about attributes and their values. All sub- matrixes located outside main diagonal represent deduced information after processing.

When the process of data importing is finished, some space of the repository matrix remains empty. This empty space is used for information obtained during the processing and so it is changed in an active part of the matrix. These active (white) sub-matrixes are used for searching if–then rules using method on the first level of processing. The active part of the repository matrix enables effective subsequent processing and creation of graphical information representation. The graphical representation depends on chosen attribute. It is implemented in the WKID system and will be described in next chapters. Figure 2 illustrates graphical interpretation of data.


WKID system

Current attribute: <rdfs:currency

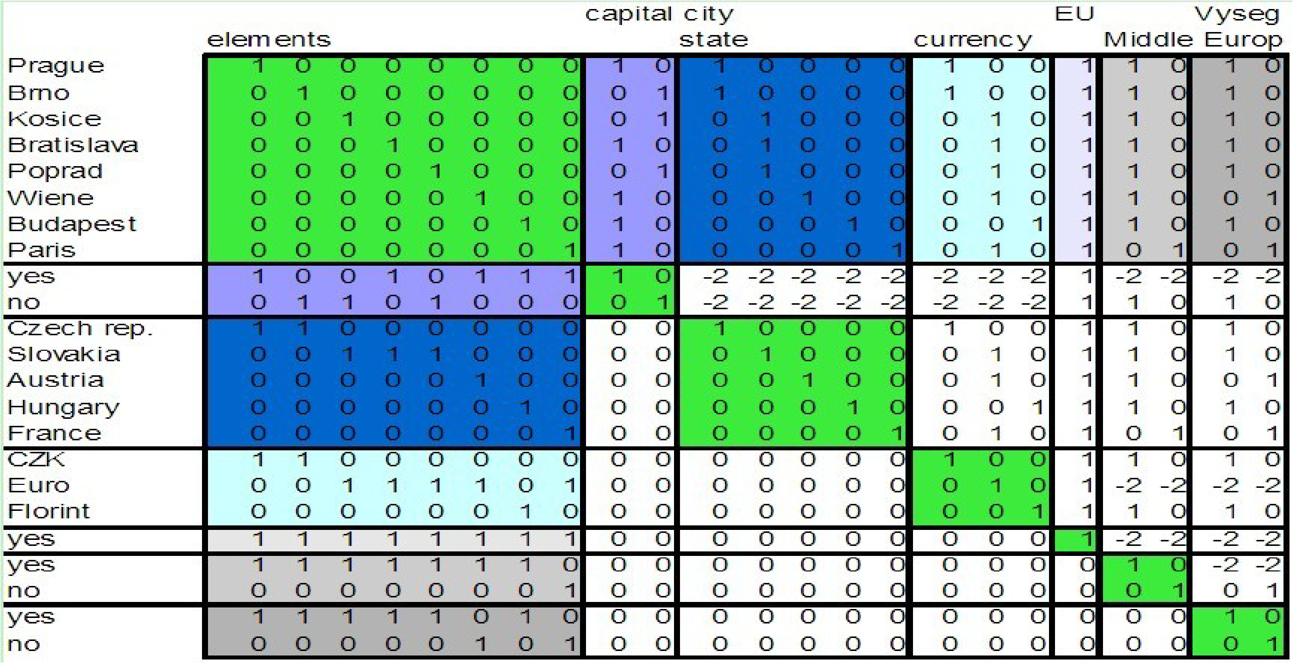
Chosen source: 1. Source

	Czk	Euro	Pound	Zloty	Florint
#Praha	■				
#Brno	■				
#Kosice		■			
#Londyn			■		
#Ostrava	■				
#Varsava				■	
#Bratislava		■			
#Poprad		■			
#Wiene		■			
#Budapest					■
#Paris		■			

Figure 2. The graphical interpretation of data in WKID system.

3. Comparison to Similar Approaches

Similar approach was used in[4], where a transposed matrix was also generated for querying the represented data. Our approach is different. No transposed matrix is used and the binary repository matrix is modified. In our approach all information got from an RDF source is downloaded in the first thick row of sub- matrixes. This representation is used because of the possibility to create effective query algorithms performing over this structure.



									capital city										EU			Vyseg	
	elements										state					currency				Middle Europ			
Prague	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	1	0	1	0
Brno	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	1	0	1	0
Kosice	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	1	0
Bratislava	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	1	0	1	0
Poprad	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	1	0
Wiene	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	1	1	0	0	1
Budapest	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	1	1	0	1	0
Paris	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	1	0	1	0	1
yes	1	0	0	1	0	1	1	1	1	0	-2	-2	-2	-2	-2	-2	-2	-2	1	-2	-2	-2	-2
no	0	1	1	0	1	0	0	0	0	1	-2	-2	-2	-2	-2	-2	-2	-2	1	1	0	1	0
Czech rep.	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	0	1	0
Slovakia	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0
Austria	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	1	1	0	0	1
Hungary	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	1	1	0	1	0
France	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1
CZK	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0
Euro	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	1	-2	-2	-2	-2
Florint	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0
yes	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	-2	-2	-2	-2
yes	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	-2	-2
no	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
yes	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
no	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 1. A sample of binary repository matrix as a data model.

Reference[8] introduces a new approach for data mining by using a computer representation form – binary 1 and 0 digits. One of the differences between our approach and their approach is the view on matrix representation. In reference[8] a binary matrix model is presented, however, rows represent entities and columns represent all possible attribute values of entities. It shows relation between object and context. We used binary matrix model, where both rows and columns represent attribute-value. In this case we try not to find context pattern of the entity, but data pattern of attribute in relation to all other attributes in a source.

4. Knowledge Relations Discovery

Reference[7] introduces approaches to knowledge mining from text documents, which can be from web area. Our approach differs from it. We transform RDF documents from web into binary repository matrix and consequently discover knowledge from this matrix. The information repository model affords opportunity to process data and to discover information on different levels. At first, two basic principles have to be distinguished: local and global point of view.

On the local level, various methods for knowledge extraction and discovery from one or more sources are applied.

On the global level, some conclusions from local level are interpreted using several techniques responsive to a user request. On the global level, higher level methods will be used for more sources processing.

The local methods of the knowledge relationship discovery on various levels of processing are implemented in the WKID system. There is if-then rules generation on the first level of processing. It is implemented within "Model creation". On the second level of processing, method "Exact comparison" is implemented and method called "Dynamic comparison" represents processing on the third level. The menu of these methods in the WKID system can be seen in Figure 3.

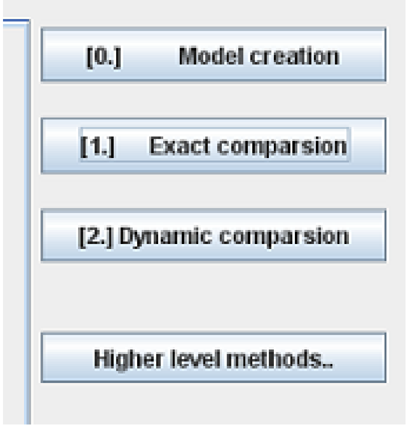


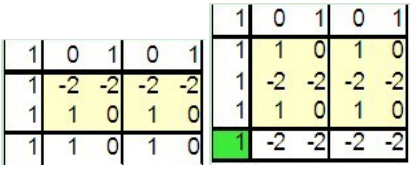
Figure 3. Menu of methods of the WKID system.

4.1. First Level Method

The most elementary method is the first level method for if-then rules generation, which is used during ***"Model creation"***. This method is able to process only one source. It is used to fill empty sub-matrixes after importing data from an actual source. These sub-matrixes (illustrated in Figure 1) represent relations between attributes related to the position. The most important part of the algorithm of the first level method is connected to if-then command. It depends on the position in matrix which is supposed to be filled. This position (matrix cell) contains information about two attributes currently being compared. If all elements (towns), which have the same value of the first attribute, are in the same group, which represents values of the second attribute, then value "1" is imported to the actual position (cell) in a selected empty sub-matrix. The value "1" is written particularly into the row with actual group position and column represents value of the first attribute. The value "0" is written to all other cells in the same row in the sub matrix. The value "-2" is written to the cell, if there is no relation. This value represents the fact that the algorithm is not able to find any relation in the process of comparing of attributes. The result of this method is generation of several if-then rules, which primary can be used like inputs for more sophisticated methods.

4.2. Second Level Method

The second level method, called ***"Exact comparison",*** is working still only with one information source. The inputs of this method are results from the first level processing, so another source uploading is not needed and a lot of time is saved. This method investigates dependence of two attributes by monitoring their behaviour on different starting conditions. This method is looking for the same behaviour represented by similar sub matrix created within the first level processing. Figure 4 illustrates two examples – sub-matrix selected from data model in repository matrix illustrated in Figure 1. The founded couple with the same behaviour detects relation, for example relation between Middle Europe and Vinegar deducted from sub-matrixes in Figure 4.



1	0	1	0	1
1	-2	-2	-2	-2
1	1	0	1	0
1	1	0	1	0

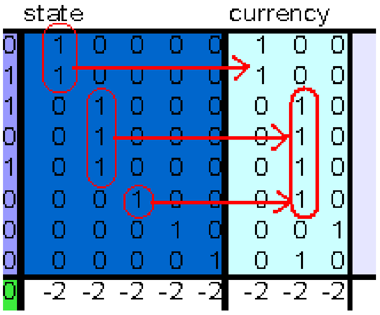
1	0	1	0	1
1	1	0	1	0
1	-2	-2	-2	-2
1	1	0	1	0
1	-2	-2	-2	-2

Figure 4. Similar sub-matrixes, which were generated from data model.

It holds that the last two columns are the same in both sub-matrixes in Figure 4. These last two columns represent attributes: Middle Europe and Vinegar. This relation is well known for humans, but it was not explicitly represented in the used source. The disadvantage of this method is a need for the same number of attributes' values in compared sub-matrixes. Another method from higher level called "Dynamic comparison" has not this disadvantage.

4.3. Third Level Method

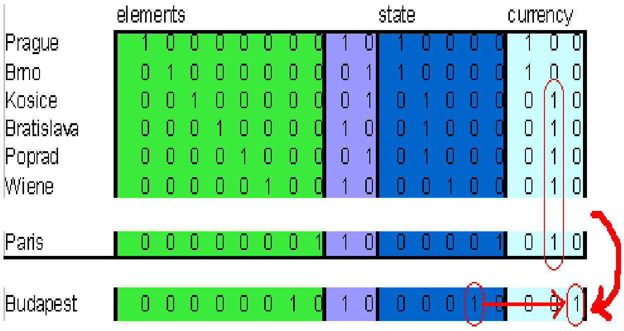
The third level method - ***"Dynamic comparison"*** uses specific groups of data from repository matrix. It distinguishes two categories: individuals and groups. The groups are created in columns from adjacent values "1" in a sub-matrix. In the process of comparison two cases can occur. At first, there is no difference in structure (values) of compared attributes and so there exist a relation between them. At second, there are some differences in the structure of compared attributes. In this case, the algorithm searches for individual value "1", which can be connected to bigger group, as it is illustrated in Figure 5.



	state					currency		
0	1	0	0	0	0	1	0	0
1	1	0	0	0	0	1	0	0
1	0	1	0	0	0	0	1	0
0	0	1	0	0	0	0	1	0
1	0	1	0	0	0	0	1	0
0	0	0	1	0	0	0	1	0
0	0	0	0	1	0	0	0	1
0	0	0	0	0	1	0	1	0
0	-2	-2	-2	-2	-2	-2	-2	-2

Figure 5. Bigger group creation from given individuals in a sub matrix.

The algorithm also tries to change rows ordering in this case. For example, let us consider two attributes state and currency, represented in the third and fourth sub-matrixes in the first thick row in the repository matrix. In the currency sub-matrix there is individual value "1", which can be connected to bigger group after ordering change between seventh and eight rows (see Figure 6). This change can be made because of the occurrence of individual values "1" in these rows in the both sub-matrixes and so this change does not cause disorder in the actual groups of values "1".



	elements										state					currency		
Prague	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0
Brno	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0
Kosice	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0
Bratislava	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0
Poprad	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0
Wiene	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0
Paris	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0
Budapest	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1

Figure 6. Ordering change between seventh and eighth rows in a sub matrix.

This method is able also to extract information about dominance of one attribute in the case when bigger group is connected with smaller one or with few individuals. It unites values of the weaker attribute. The mentioned example illustrates search for relation between state and currency. This relation is following. The groups of cities in the same country as well as in some different countries have the same currency. For example, some countries of the Europe have the same currency (Euro). But one state can has only one currency.

One fact, which should not be ignored in the process of dynamic comparison, is the number of allowed changes. A limit adequate to the dimension of used source has to be defined. Too many changes can damage attribute patterns and can lead to misinformation.

5. Experiments

The designed methods were implemented within the WKID system. This system was subsequently tested. We wanted to prove in the ***first experiment***, that our information repository model supports quick and effective information retrieval. Information retrieval was provided by the matrix and vector multiplication. The matrix played the role of an information repository and the vector represented the query used for searching relations of a particular element (attribute) to others elements in the source. The experiments with implementation were provided on a binary repository matrix, which is presented in Figure 7. This matrix represents such elements respectively attributes as: Sport, Players number, Playground, Kind, Playing area, Playing style and Tool. These attributes have the following values:

Sport = {snowboarding, skiing, ice hockey, volleyball, football, basketball}

Players number = {less than 4, more than 4}

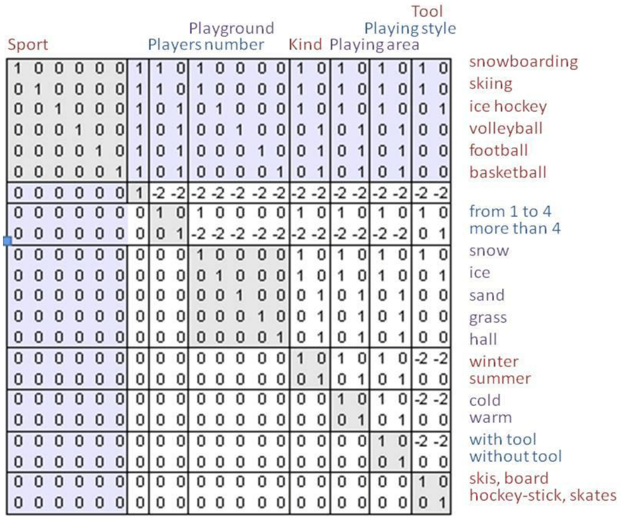
Playground = {snow, ice, sand, grass, hall}

Kind = {winter, summer}

Playing area = {cold, warm}

Playing style = {with tool, without tool}

Tool = {skis, board, hockey-stick, skates}.



Sport							Players number		Playground					Kind		Playing area		Playing style		Tool		
1	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	1	0	1	0	1	0	snowboarding
0	1	0	0	0	0	1	1	0	1	0	0	0	0	1	0	1	0	1	0	1	0	skiing
0	0	1	0	0	0	1	0	1	0	1	0	0	0	1	0	1	0	1	0	0	1	ice hockey
0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	1	0	1	0	1	0	0	volleyball
0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0	football
0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	1	0	1	0	1	0	0	basketball
0	0	0	0	0	0	1	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	
0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0	1	0	1	0	from 1 to 4
0	0	0	0	0	0	0	0	1	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	0	1	more than 4
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	1	0	snow
0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	ice
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	sand
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	grass
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	0	hall
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	-2	-2	winter
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	summer
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	-2	-2	cold
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	warm
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	-2	-2	with tool
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	without tool
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	skis, board
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	hockey-stick, skates

Figure 7. Binary repository matrix from the sports domain.

The matrix in Figure 7 was obtained by the application of the first level method of the WKID system. The seventh row and seventh column represent [Type = element].

The first query in the first experiment was: What information about element "ice hockey" can be obtained from the repository matrix? This query can be represented as the following vector:

q = [0010000000000000000000].

It is the vector of element "ice hockey" activation, which is in the third row of the matrix. This vector was used in the transposed form in this experiment (as a column vector). Within similar approach introduced in[4], there was provided the process of searching by repeated multiplication of a matrix and a vector. In our approach, only one multiplication of the binary matrix and the query vector is sufficient for finding all information, which can be obtained from the source. The result of this multiplication is the vector:

o = [0010001010100010101001].

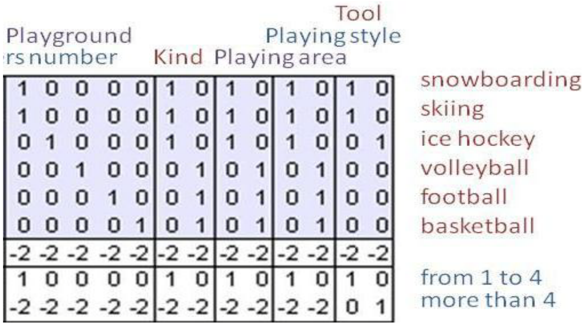
This vector represents all active domains: [Sport = ice hockey], [Type = element], [Players number = more than 4], [Playground = ice], [Kind = winter], [Playing area = cold], [Playing style = with tool] and [Tool = hockey-stick & skates]. Some of these domains [Players number = more than 4] and [Kind = winter] represent relation between winter sports and sports with more then 4 players. This result does not mean that another alternative (for example figure skating as a stand up winter sport) must be refused.

The ***second experiment*** was focused on the designed methods' functionality and their behaviour. The functionality was tested on the same source as the first experiment. This source was originally represented in the RDF form and it was consequently transformed into the binary matrix (see Figure 7) using the *first level method*. The attributes of this source were set in the way, which ensure dimensionally compatible sub matrixes for the *second level - exact comparison method* testing. Also one additional parameter was set. It was a limit for the *third level - dynamic comparison method* testing to prevent samples destruction. The attributes: [Kind = winter or summer] and [Playing area = cold or warm] were introduced for testing ability of redundancy finding. These two attributes have different names but the same property. There were two main problems concerning data in the binary matrix:

1) mistaken data (some facts are added incorrectly),
2) missing data (nonexistent or undetected – they exist but we do not know about them).

The redundancy can be classified to the first kind of problems, when the same attribute has more different names. We tried to test both kinds of problems and to detect their impact on our system functioning.

The *exact comparison method* processed data from binary matrix. Its output consisted of sub matrixes with derivative information. This method iterated through the source systematically and denoted all existing relations between elements in this source, what is illustrated in Figure 8.



Playground rs number					Kind		Playing area		Playing style		Tool		
1	0	0	0	0	1	0	1	0	1	0	1	0	snowboarding
1	0	0	0	0	1	0	1	0	1	0	1	0	skiing
0	1	0	0	0	1	0	1	0	1	0	0	1	ice hockey
0	0	1	0	0	0	1	0	1	0	1	0	0	volleyball
0	0	0	1	0	0	1	0	1	0	1	0	0	football
0	0	0	0	1	0	1	0	1	0	1	0	0	basketball
-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	
1	0	0	0	0	1	0	1	0	1	0	1	0	from 1 to 4
-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	0	1	more than 4

Figure 8. Data processing by the exact comparison method.

One advantage of this method is that it could find each hidden relation in the examined binary matrix. On the other hand, some relations between elements were unknown. But it does not need to be a problem, because patterns of behaviour could contain also some uncertain regions.

The value "1" in the sub matrices represented if-then rule. For example:

If sport has [Tool = skis or board]
then [Players number = from 1 to4]
for the given source.

The output of the *dynamic comparison method* was the list of attributes couples, which there were relations between them. In the experiments with this method, the output was generated in the following form:

Found: <rdfs: Players_number vs. <rdfs: Kind
Found: <rdfs: Players_number vs. <rdfs: Playground
Found: <rdfs: Players_number vs. <rdfs: Playing_style
Found: <rdfs: Players_number vs. <rdfs: Tool ...

These findings can be interpreted as relations of the attribute "Players number" to the attributes "Kind ", "Playground ", "Playing style "and "Tool". Some more relations were found within this experiment, as a relation between "Kind " and "Playground ","Kind " and "Playing style" and also a relation between "Playground " and "Playing style". These relations were found in the source repeatedly.

The second experiment also showed redundancy between two attributes: [Kind = winter or summer] and [Playing area = cold or warm]. This experiment indicated one drawback of the given source, which was not complete. It resulted in the fact that all sports with tool are only winter sports. This mistake was imported into the source purposely to investigate the ability of our methods to eliminate the impact. Consequently, one mistaken relation was found: [Playing style = with tool] is equal to [Kind = winter]. Dealing with missing data is a general problem in the field of knowledge discovery. It showed an importance of correctly and completely created source and a need to develop higher level methods.

6. Conclusions

The novelty of this paper is in using binary matrix representation in combination with the semantic web vision and in the transforming of information from RDF documents into binary repository model. The current work obtains implementation of local level methods unable knowledge relationship discovery. Particularly, model creation and if-then rules generation, exact comparison and dynamic comparison were implemented in the WKID system. These methods were designed to process one source and generate all possible knowledge pieces from it. Deducted knowledge seems to be adequate to the real word knowledge.

Our modifications of binary matrix used as a repository model, increased effectiveness of the process of relations mining from a source. Some more modifications can be implemented, for example using the binary matrix for the representation of the WKID system methods results. It can be the way how to represent meta-data from more resources. The system can be extended by an ability to discover the relations between frequently used words and so it can be used e.g. for newsgroup discussions analysis[5].

Our system enables creation of methods levels, which can be independent and they can have access to the results from various levels of sources processing (various methods). Our system also represents semantic approach to information integration because it enables synthesis and comparison of more sources from a semantic aspect.

In near future, our effort will concentrate on higher level methods development and implementation and on more RDF sources processing. Also more sources formats, obvious for classical web, can be used in the WKID system in the future to enable a fluent collaboration between semantic approach and actual web. We intend to increase applicability of our system for more expansive and more complicated web domain. Maybe an idea of patterns discovery can be implemented in the WKID system. Sub-matrixes of attributes can be modified in dependence on different attributes. The structure of sub-matrixes can be simplified and can offer more unknown knowledge.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic within the 1/0685/12 project ”Methods for analysis, extraction, modelling and prediction of collaborative information networks” (50%).

This work is also the result of the project implementation Development of the Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

REFERENCES

[1] G. Antoniu, F. van Harmelen, “A Semantic Web Primer”, Massachusetts Institute of Technology, USA, 2004, 238 pp., ISBN 0-262-01210-3

[2] D. Fensel, at al., “Enabling Semantic Web Services”, Springer-Verlang, Berlin, 2007, 188 ps., ISBN 3-540-34519-1

[3] T. Finin, at al., “Information Integration and the Semantic Web”, Workshop on Information Integration, 2006

[4] Z. Linková, M. Řimnáč, “Computerized Rules Design for Data Integration and Sémantic Web”, Proc. of the 7th annual international conference Znalosti 2008, Bratislava, February 13-15, 2008, Publisher: Slovak Technical University, Slovak Republic, 2008, 124-135, ISBN 978-80-227-2827-0

[5] M. Mach, G. Lukáč, “A dedicated information collection as an interface to newsgroup discussions”, IIS 2007 - 18th international confe-rence on Information and Intelligent Systems, September 12-14, 2007,Varazdin, Croatia, 163-169, ISBN 978-953-6071-30-2

[6] P. Návrat, M. Bieliková, D. Chudá, V. Rozinajová, “Intelligent Information Processing in Semantically Enriched Web”, Foundations of Intelligent Systems. Lecture notes in Computer Science, Springer Berlin Heidelberg, Vol. 5722, 2009, 331-340, ISSN 1867-8211

[7] J. Paralič, K. Furdík, G. Tutoky, P. Bednár, M. Sarnovský, P. Budka, F. Babič, “Knowledge mining from texts“, Equilibria, s.r.o., Košice, 2010, 80 ps, ISBN 978-80-89284-62-7

[8] I. Spiegler, R. Gelbard, “A Binary Model and Methodology to Represent Knowledge for Data Mining”, US Patent, No. 2002/0087567 A1. Granted April 27, 2004

[9] J. Vrana, M. Mach, “Key concepts extended by vector descriptions to interpret the meaning of ontologies” Acta Electrotechnica et Informatica, Vol. 11, no. 3, 2011, 57-63, ISSN 1335-8243

[10] S. Wolfram, “Computational Knowledge Engine – Wolfram Alpha”, http://www.wolframalpha.com/, Accessed: 9th November, 2010