

# Analysis of the Web Sites Use as Learning Resources

Miloslava Kašparová, Martin Horčíčka

Institute of System Engineering and Informatics  
Faculty of Economics and Administration  
University of Pardubice  
Studentská 84, 532 10 Pardubice, Czech Republic

Jiří Křupka

Institute of System Engineering and Informatics  
Faculty of Economics and Administration, University of  
Pardubice  
Studentská 84, 532 10 Pardubice, Czech Republic

**Abstract**—This article deals with an analysis of the web sites use by the students for gaining the electronic study sources and discovering of associations in the data by selected algorithm. Association rules associate a particular conclusion with a set of conditions. The Generalized Rule Induction was applied. Models were designed in SPSS Clementine.

**Keywords**—data analysis; web sites; learning sources; association rule; generalized rule induction

## I. INTRODUCTION

Thanks to the Internet, it is possible nowadays to find a lot of information and data, which are suitable for studying as well as for leisure time. If we aim at the Universities and Colleges, students can attend the lectures and seminars. They also can borrow the specialized books and scripts from the library or to buy them in the specialized bookshops. This form can be then used for study in the printed form or in electronic form. Another form of gaining the information and data is the Internet. It consists of a huge amount of information connected to the specific fields or subjects, which are taught at the Universities or Colleges. However, there could be a problem in their order or lucidity or even in their veracity.

### A. Information Sources

Main and inevitable information sources in the process of individual gaining of the information and specialized knowledge are registered information. The name document is used for naming this registered information and forms. Documents consist of information of different characteristics and orientation. The unit that is formed is called natural fond of documents. This represents very complex, dynamic system, in which it is necessary to order the documents according to the different criteria for better orientation. Documents are according to their main communication function divided into: scientific, educational, popular science, political and administrative documents, which are described according to [1].

*Scientific documents* arise from entering the information, which are gained by the scientific activities of people. Their main function is to present the results of the scientific research, theoretical one or factual characteristics. In the scientific field every new data or pieces of knowledge are added to the previous one or it can deepen it, modify it or replace it.

Information from this field can be gained from two groups of scientific documents. First group can be formed by the former scientific documents. It is e.g. research reports, scientific articles and others. Second group can be formed by the synthesizing scientific documents. Those are qualified to the connection of former information with the new one, e.g. it could be scientific encyclopedia.

Scripts, students' books, collection of tasks or lectures differ not only from its content, but also from the way of presentation. Some of them are by its content closer to the scientific documents of the synthesizing characteristic and others impinge by its content to the popular science field.

*Popular science documents* consist of information from the all spheres of human activities, but modified to the popular form. This modification is done so that the information and pieces of knowledge are understood by the readers, who do not have any knowledge of this field. They are aimed at the satisfaction of the amateur cognitional or practical needs. Into this field it is possible to put e.g. popular introductions of the scientific fields, popular science encyclopedia or films.

*Political documents* represent very different and non-limited group. There belong official political documents (e.g. acts, materials from the meetings of the government bodies, documents from the parliament hearings, regulations of the public administration central authority or international agreements).

*Administrative documents* consist of results of the administration practice. They satisfy general and professional needs, which are connected to the administrative and economic relationships between the companies, companies and individuals and between individuals. Characteristic feature of these materials is their high level of communications data formalization. Typical example is official correspondence, requests, orders, confirmations and others.

### B. Study Sources

*Study source* is able to be understood as a document where we can gain the information for studies. Generally it can be divided into printed study sources and electronic sources.

A book is a typical example of *printed study source*. It fills the function of universal communication media. Nowadays, its influence is not as important as it used to be. Books are more

---

This work was supported by the project No. SGFES02 of the Ministry of Education, Youth and Sports of the CR with title Research and Development Activities in the area of System Engineering and Informatics at the Faculty of Economics and Administration, University of Pardubice in 2013.

and more substituted by the quicker media, i.e. the Internet. On the other hand, the book is suitable for its oneness in recording the wider pieces of information. It provides the systematic explanation and summarization of reached results. The examples are scripts, research papers, thesis and others.

Among the *electronic study sources* belong books, scripts, lectures and others in electronic form. They are accessible via the Internet or other media for distribution of digital data (e.g. DVD, USB Flash disc). They also could be magazines, thesis of students, contributions from the proceeding papers and others, which students can gain e.g. from the university library, electronic encyclopedia [2] and dictionaries [1].

### C. Web Sites

New information on the Internet is appearing all the time, is changing and is being deleted. The common user gets the information usually via web sites.

Web site is hypertext document, which is displayed thanks to www browser after displaying the files saved at the web server. For displaying the hypertext document in the Internet environment serves the service WWW (World Wide Web). Thanks to its merit, the Internet was popularized and it was opened to all categories of users [3].

During the 1990s, the WWW emerged as the most popular medium of Internet. Today, the Web – just like the Internet – is used not only for consuming information passively, but for creating and publishing new content, distinguishing it from traditional media. This was reinforced by popular websites that simplify user contributions: Wikipedia, YouTube, and numerous social networking sites are typical examples. Increasing amounts of knowledge are created by individuals, leading to a phenomenon that has described as read-write culture [4].

The goals of this article are:

- To analyse the web sites use by the students for gaining the electronic study sources
- To search mutual associations in the gained data set by selected algorithm

## II. PROBLEM FORMULATION

The article deals with the analysis of the web sites use by the students during the gaining the study electronic sources including the search of associations in the data on the basis of selected algorithm. Analysis will be carried out on the data gained from the questionnaire, which was asked in January, February and March 2013. The questionnaire was placed on the social network Facebook into groups, which associate the students of the University Pardubice. It was also published within the Learning Management System (LMS) Moodle at the Institute of System Engineering and Informatics.

### A. Data Collection

Questionnaire consisted of twenty questions. Questions were mainly closed [5], i.e. alternative (offer the choice of two answers) and selective (offer more than two alternatives of

answers). Other types were closed multi-valued questions (two overall), where it is possible to choose from more answers. There were also placed two open questions, where the respondents formulated the answer independently. At the end of the questionnaire were four demographical questions. Examples of the questions are: Q2: How often do you use the Internet for independent gaining study sources (it is not related to the links and files (e.g. specialized articles), which were provided by the lecturers for self-study)? Q3: Which study source do you find the best for studying? Q7: Do you think there are enough authentic web sites related to your studies, which you can gain the study materials? Q8: What is an approximate ratio of web sites visits related to your studies to the visit of web sites for leisure time? Q10: Do you gain the study sources from the foreign web sites? Q12: What are your most used web sites for searching for study sources? Q13: Are gained study materials from the mentioned web sites an adequate tool for successful course completion? Q14: Does it happen to you that the found study sources on the Internet are misleading? Q15: Do you verify the content of the study materials gained independently on the Internet in the specialized literature (related to study materials which were not provided by the lecturer)? Q16: Do you think that searching for study materials on the Internet leads the students to the greater independency? Q17: What is your sex? Q18: How old are you?

### B. Data Analysis

Within the questionnaire there answered 400 respondents, from which 203 were men and 197 women. Age groups including their representation are shown in the Table 1.

TABLE I. AGE GROUPS OF THE RESPONDENTS

Age Group	Frequency	Frequency in [%]
19-20	32	8.00
21-22	142	35.50
23-24	174	43.50
25-26	47	11.75
27-28	1	0.25
Older than 28	4	1.00

Students between 23 and 24 years (43.50%) represented the most frequent age group and then students from the age of 21 to 22 years (35.50%). One student from the age of 27 to 28 (0.25%) represented the least frequent category. Preference of the used printed and electronic sources according to the age groups is shown in the Table 2. From the data analysis results that 56.34% respondents from the age of 21 or 22 and 59.2% respondents from the age of 23 or 24 uses rather printed study sources. From the gender point of view it was found out, that women prefer printed study sources (in total 72.1% from the overall number of asked women) and men electronic (in total 55.67% from the overall number of the men).

The Table 3 shows the frequency of answers to the questions Q2: How often do you use the Internet for independent gaining study sources? (in the Table 3 Use of the Internet) and Q3: Which study source do you find the best for studying? (in the Table 3 Study source). We can see that respondents, who sometimes use the Internet for the studies

prefer the printed study sources. However respondents, who use the Internet for studies very often, partly predominates the preference to the electronic study sources.

TABLE II. FREQUENCIES OF USAGE THE STUDY SOURCES ACCORDING TO THE AGE GROUPS

Q18: Age Group	Q3: Study Source		Total
	Electronic	Printed	
19 -20	10	22	32
21 - 22	62	80	142
23 - 24	71	103	174
25 - 26	25	22	47
27 - 28	0	1	1
Older than 28	0	4	4
Total	168	232	400

TABLE III. FREQUENCIES OF USAGE THE STUDY SOURCES AND THE INTERNET FOR INDEPENDENT GAINING THE STUDY SOURCES

Q2: Use of the Internet	Q3: Study Source		Total
	Electronic	Printed	
Very often	111	91	202
Often	56	89	145
Sometimes	1	47	48
I don't use it	0	5	5
Total	168	232	400

From the use of the Internet point of view for independent gaining the study sources (Q2) and use of the foreign web sites (Q10) was found out that 47% students who use the Internet very often also use the foreign web sites (Table 4). This information is also valid for 52.4% respondents, who use the Internet often for independent gaining the study sources.

TABLE IV. FREQUENCIES OF THE INTERNET USE AND USE OF FOREIGN WEB SITES

Q2: Use of the Internet	Q10: Use of Foreign Web Sites		Total
	Yes	No	
Very often	95	107	202
Often	76	69	145
Sometimes	18	30	48
I don't use it	0	5	5
Total	189	211	400

From the data analysis results that in total 80% respondents, who use the Internet very often for their independent gaining the study sources, consider the web sites to be adequate tool for the course completion of the selected subject. This is also valid for 77.2% respondents, who often use the Internet. In total 54.2% asked students, who sometimes use the Internet, do not consider web sites to be adequate tool for the course completion. Then it was found out, that 71.8% respondents verify the veracity of the content of the study sources gained independently on the Internet in the specialized literature. 40.75% students sometimes found misleading information on the Internet from the point of view to gain the study source; 55% found always the relevant sources for their studies. Nearly 60% students think that searching for study sources on the

Internet leads to the greater independence of the students. The Fig. 1 shows the frequency of answers to the question Q8: What is an approximate ratio of web sites visits related to your studies to the visits of web sites for leisure time?

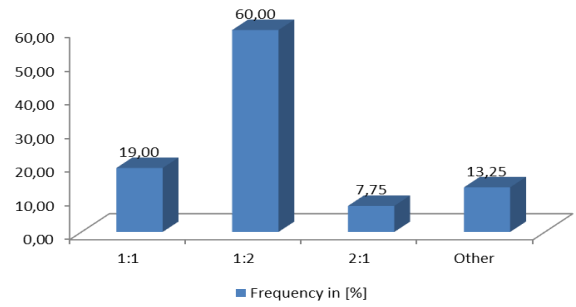


Fig. 1. Frequencies of answers to the question Q8

### C. Data Pre-processing

Data gained from the questionnaire was evaluated and then processed to the required data matrix. At the basis of initial 20 questions and possible answers was processed the data matrix of the range (400×54). It contained 54 nominal variables (attributes) with values 0 and 1. The list of variables is in the Table 5 and 6.

TABLE V. LIST OF ATTRIBUTES (PART I)

ID attribute	Description of an Attribute	Values
$a_1$	Q2: Use of the Internet: Very often	{0, 1}
$a_2$	Q2: Use of the Internet: Often	{0, 1}
$a_3$	Q2: Use of the Internet: Sometimes	{0, 1}
$a_4$	Q2: Use of the Internet: I don't use it	{0, 1}
$a_5$	Q4: Use of electronic services of the university library: Searching of books, textbooks, scientific articles and other documents by criteria	{0, 1}
$a_6$	Q4: Use of electronic services of the university library: Reservation of the study literature	{0, 1}
$a_7$	Q4: Use of electronic services of the university library: Download of an electronic literature	{0, 1}
$a_8$	Q4: Use of electronic services of the university library: Possibility to extend a reservation of a borrowed study literature	{0, 1}
$a_9$	Q4: Use of electronic services of the university library: I don't use them	{0, 1}
$a_{10}$	Q5: Other services of the university library: To inform students by SMS relative to date of reclaim the book	{0, 1}
$a_{11}$	Q5: Other services of the university library: browsing of on-line books	{0, 1}
$a_{12}$	Q5: Other services of the university library: I don't request any new service	{0, 1}
$a_{13}$	Q5: Other services of the university library: I don't know	{0, 1}
$a_{14}$	Q6: Use of the reader's account of the university library	{0, 1}
$a_{15}$	Q7: Enough authentic web sites related to your studies	{0, 1}
$a_{16}$	Q10: Use of study sources from the foreign web sites	{0, 1}

TABLE VI. LIST OF ATTRIBUTES (PART II)

ID attribute	Description of an Attribute	Values
$a_{17}$	Q11: Use of the translator: <a href="http://translate.google.com">http://translate.google.com</a>	{0, 1}
$a_{18}$	Q11: Use of the translator: <a href="http://online-slovník.cz/prekladac">http://online-slovník.cz/prekladac</a>	{0, 1}
$a_{19}$	Q11: Use of the translator: <a href="http://onlineprekladac.com">http://onlineprekladac.com</a>	{0, 1}
$a_{20}$	Q11: Use of the translator: I don't use it	{0, 1}
$a_{21}$	Q11: Use of the translator: Another	{0, 1}
$a_{22}$	Q12: Most used web sites for searching for study sources: <a href="http://www.google.cz">http://www.google.cz</a>	{0, 1}
$a_{23}$	Q12: Most used web sites for searching for study sources: <a href="http://www.primat.cz">http://www.primat.cz</a>	{0, 1}
$a_{24}$	Q12: Most used web sites for searching for study sources: <a href="http://www.wikipedia.cz">http://www.wikipedia.cz</a>	{0, 1}
$a_{25}$	Q12: Most used web sites for searching for study sources: <a href="http://www.finance.cz">http://www.finance.cz</a>	{0, 1}
$a_{26}$	Q12: Most used web sites for searching for study sources: <a href="http://www.vutbr.cz">http://www.vutbr.cz</a>	{0, 1}
$a_{27}$	Q12: Most used web sites for searching for study sources: Another	{0, 1}
$a_{28}$	Q13: Study materials from the mentioned web sites an adequate tool for successful course completion	{0, 1}
$a_{29}$	Q15: Verification the content of the study materials gained on the Internet in the specialized literature	{0, 1}
$a_{30}$	Q16: Searching for study materials on the Internet leads the students to the greater independency	{0, 1}
$a_{31}$	Q17: Sex	{0, 1}
$a_{32}$	Q8: An approximate ratio of web sites visits related to studies to the visit of web sites for leisure time: 1:1	{0, 1}
$a_{33}$	Q8: An approximate ratio of web sites visits related to studies to the visit of web sites for leisure time: 1:2	{0, 1}
$a_{34}$	Q8: An approximate ratio of web sites visits related to studies to the visit of web sites for leisure time: 2:1	{0, 1}
$a_{35}$	Q8: An approximate ratio of web sites visits related to studies to the visit of web sites for leisure time: Another	{0, 1}
$a_{36}$	Q14: An experience with misleading study sources on the Internet: Often	{0, 1}
$a_{37}$	Q14: An experience with misleading study sources on the Internet: Sometimes	{0, 1}
$a_{38}$	Q14: An experience with misleading study sources on the Internet: they were not misleading.	{0, 1}
$a_{39}$	Q14: An experience with misleading study sources on the Internet: Study sources from Internet I don't use	{0, 1}
$a_{40}$	Q3: Study source the best for studying: electronic	{0, 1}
$a_{41}$	Q3: Study source the best for studying: printed	{0, 1}
$a_{42}$	Q4: Use of electronic services of the university library: I use them	{0, 1}
$a_{43}$	Q4: Use of electronic services of the university library: I don't use them	{0, 1}
$a_{44}$	Q4: Use of electronic services of the university library: I don't know them	{0, 1}
$a_{45}$	Q18: Age group: 19 -20 years	{0, 1}
$a_{46}$	Q18: Age group: 21 -22 years	{0, 1}
$a_{47}$	Q18: Age group: 23 -24 years	{0, 1}
$a_{48}$	Q18: Age group: 25 -26 years	{0, 1}
$a_{49}$	Q18: Age group: 27 -28 years	{0, 1}

$a_{50}$	Q18: Age group: Older than 28 years	{0, 1}
$a_{51}$	Q20: Accommodation: hall	{0, 1}
$a_{52}$	Q20: Accommodation: lodgings	{0, 1}
$a_{53}$	Q20: Accommodation: I commute	{0, 1}
$a_{54}$	Q20: Accommodation: I study in town where I have my place of residence	{0, 1}

### III. PROBLEM SOLUTION

The article deals with a search for associations in the data by selected algorithm [6 - 10]. There were specified three goals (concerning searching of associations in the data in regard to the following questions):

- Q2: How often do you use the Internet for independent gaining study sources (it is not related to the links and files (e.g. specialized articles), which were provided by the lecturers for self-study)? (1<sup>st</sup> goal)
- Q10: Do you gain the study sources from the foreign web sites? (2<sup>nd</sup> goal)
- Q7: Do you think there are enough authentic web sites related to your studies, which you can gain the study materials? (3<sup>rd</sup> goal)

It was used association rules on the basis of above mentioned goals. They associate a particular conclusion with a set of conditions. Association rules are statements in this form (1) [11]:

$$\text{If } antecedent(s) \text{ Then } consequent(s). \quad (1)$$

It was applied the Generalized Rule Induction (GRI). We can see an application of GRI for example in [12 - 15]. The GRI extracts a set of rules from the data. Each rule in the final rule set has associated instances, support and confidence. *Instances* is calculated as the number of records for which the antecedent is true [16]. Rule support  $m_{RS}$  and *confidence*  $m_C$  are two measures of rule interestingness. They respectively reflect the usefulness and certainty of discovered rules [17].

*Confidence*  $m_C$  [16, 17] displays the ratio of rule support to antecedent support. This indicates the proportion of IDs (Rule ID displays the rule ID assigned during model building) with the specified antecedent(s) for which the consequent(s) is/are also true [11]. Equation (2) of *confidence*  $m_C$  is the following:

$$m_C = m_{RS} / m_S, \quad (2)$$

where *support*  $m_S$  displays antecedent support; *rule support*  $m_{RS}$  displays the proportion of IDs for which the entire rule, antecedents, and consequent(s), are true [11].

Software SPSS Clementine 10.1 was used. Achieved results of models were analysed. The design of model creation is in the Fig. 2.

#### A. Modelling and Evaluation

At the start of modeling nine variables were excluded on the basis of too many records falling into the same category relative to the total number of records (maximum percentage of

records in a single category was 95%; it means this percentage was greater for variables  $a_4, a_{11}, a_{19}, a_{21}, a_{26}, a_{36}, a_{39}, a_{49}, a_{50}$ . The initial set of variables included 46 variables. The set of attributes described in the Table 5 was the following (3):

$$A = \{a_1, a_2, a_3, \dots, a_{54}\}. \quad (3)$$

To create GRI association rules we needed some input attributes and one output attribute. Default setting of the GRI was the following: maximum antecedent support (%) is 0%, minimum rule confidence is 50%, maximum number of antecedent is 3 and maximum number of rules is 100.

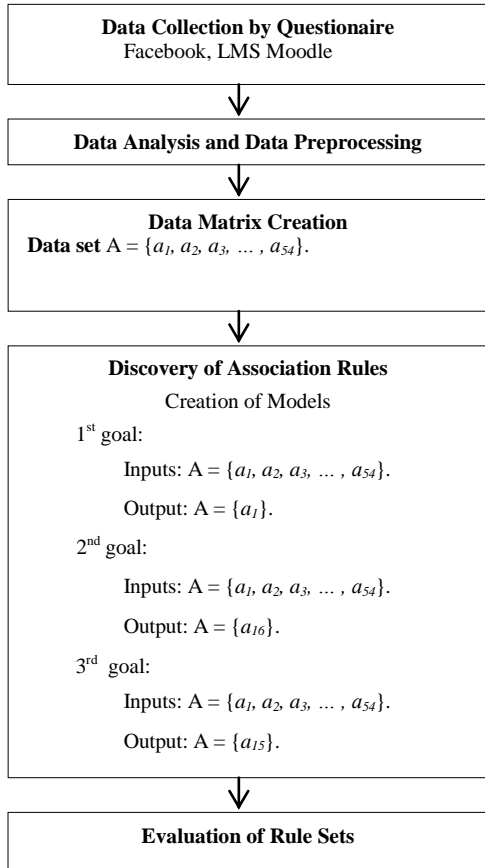


Fig. 2. Design of models creation

**1<sup>st</sup> goal:** For realization of the 1<sup>st</sup> goal the  $a_1$  attribute was the output variable. The remainder attributes were inputs. It was achieved 100 rules. On account of a decreasing of rules number some unimportant variables were excluded; it related to variables  $a_5, a_6, a_7, \dots, a_{15}$ , that informed about services of the university library. But it was generalized 100 rules again. Examples of 5 rules sorted by rule support  $m_{SR}$  are in the Table 7.

In this table we can see antecedents, consequent, instances and rules support  $m_{RS}$ . By application of the GRI it was discovered, that from point of view of very often use the Internet for independent gaining study sources:

TABLE VII. FIVE RULES BY GRI (1<sup>ST</sup> GOAL)

Consequent	Antecedent	Instances	Rule Support in [%]
$a_1 = 1$	$a_{15} = 1$ and $a_{28} = 1$	192	27.75
$a_1 = 1$	$a_{40} = 1$	168	29.5
$a_1 = 1$	$a_{28} = 1$ and $a_{40} = 1$	147	24.25
$a_1 = 1$	$a_{15} = 1$ and $a_{28} = 1$ and $a_{29} = 1$	149	23.5
$a_1 = 1$	$a_{29} = 1$ and $a_{40} = 1$	136	22.75

**Rule 1:** 111 students think, that it is enough authentic web sites related to their studies, which you can gain the study materials and find mentioned web sites an adequate tool for successful course completion ( $m_C = 61.46\%$ ,  $m_S = 48\%$ ).

**Rule 2:** 118 students find electronic study sources the best for studying ( $m_C = 66.07\%$ ,  $m_S = 42\%$ ).

**Rule 3:** 97 students find mentioned web sites an adequate tool for successful course completion and electronic study sources as the best for studying. ( $m_C = 65.99\%$ ,  $m_S = 36.75\%$ ).

**Rule 4:** 94 students think, that it is enough authentic web sites related to their studies, which you can gain the study materials, find mentioned web sites an adequate tool for successful course completion and verify the content of the study materials gained independently on the Internet in the specialized literature ( $m_C = 63.09\%$ ,  $m_S = 37.25\%$ ).

**Rule 5:** 91 students verify the content of the study materials gained independently on the Internet in the specialized literature and find electronic study sources the best for studying ( $m_C = 66.91\%$ ,  $m_S = 34\%$ ).

**2<sup>nd</sup> goal:** For realization of this 2<sup>nd</sup> goal the  $a_{16}$  attribute was the output variable. The remainder attributes were inputs. The number of rules was 100 again. After the elimination of attributes informed about services of the university library 11 rules were discovered. Examples of 5 rules sorted by rule support  $m_{RS}$  are in the Table 8.

TABLE VIII. FIVE RULES BY GRI (2<sup>ND</sup> GOAL)

Consequent	Antecedent	Instances	Rule Support in [%]
$a_{16} = 1$	$a_{17} = 1$	259	44.25
$a_{16} = 1$	$a_{17} = 1$ and $a_{29} = 1$	181	34.00
$a_{16} = 1$	$a_{17} = 1$ and $a_{28} = 1$	184	33.75
$a_{16} = 1$	$a_{17} = 1$ and $a_{28} = 1$ and $a_{29} = 1$	134	27.25
$a_{16} = 1$	$a_{17} = 1$ and $a_{15} = 1$	143	25.75

It follows from the results of given method that students using the foreign web sites to study:

**Rule 1:** use the translator on <http://translate.google.com> (177 students,  $m_C = 68.34\%$ ,  $m_S = 64.75\%$ ).

**Rule 2:** use the translator on <http://translate.google.com> and verify the content of the study materials gained independently on the Internet in the specialized literature (136 students,  $m_C = 75.14\%$ ,  $m_S = 45.25\%$ ).

**Rule 3:** use the translator on <http://translate.google.com> and students find mentioned web sites an adequate tool for successful course completion (135 students,  $m_C = 73.37\%$ ,  $m_S = 46\%$ ).

**Rule 4:** use the translator on <http://translate.google.com> and find mentioned web sites an adequate tool for successful course completion and verify the content of the study materials gained independently on the Internet in the specialized literature (109 students,  $m_C = 81.34\%$ ,  $m_S = 33.50\%$ ).

**Rule 5:** use the translator on <http://translate.google.com> and think, that it is enough authentic web sites related to their studies, which you can gain the study materials (103 students,  $m_C = 72.03\%$ ,  $m_S = 35.75\%$ ).

**3<sup>rd</sup> goal:** The attribute  $a_{15}$  was the output for realization of 3<sup>rd</sup> goal. Under this question (Q7: Do you think there are enough authentic web sites related to your studies, which you can gain the study materials?) the associations in the data were searched. Total 100 rules were in the rule set (also after the elimination of above mentioned variables). Examples of 5 rules sorted by rule support  $m_{SR}$  are in the Table 9.

TABLE IX. FIVE RULES BY GRI (3<sup>RD</sup> GOAL)

Consequent	Antecedent	Instances	Rule Support in [%]
$a_{15} = 1$	$a_{28} = 1$ and $a_{29} = 1$	222	37.25
$a_{15} = 1$	$a_{28} = 1$ and $a_{33} = 1$	181	31
$a_{15} = 1$	$a_{28} = 1$ and $a_1 = 1$	161	29.5
$a_{15} = 1$	$a_{28} = 1$ and $a_{30} = 1$	173	29.5
$a_{15} = 1$	$a_{40} = 1$	168	29

In the Table 9 we can see if the consequent is the attribute  $a_{15}$  then:

**Rule 1:** 149 students think that gained study materials from the mentioned web sites are an adequate tool for successful course completion and verify the content of the study materials gained independently on the Internet in the specialized literature ( $m_C = 67.12\%$ ,  $m_S = 55.5\%$ ).

**Rule 2:** 124 students think that gained study materials from the mentioned web sites are an adequate tool for successful course completion and an approximate ratio of web sites visits related to their studies to the visits of web sites for leisure time is 1:2 ( $m_C = 68.51\%$ ,  $m_S = 45.25\%$ ).

**Rule 3:** 118 students think that gained study materials from the mentioned web sites are an adequate tool for successful course completion and use the Internet for independent gaining study sources very often ( $m_C = 73.29\%$ ,  $m_S = 40.25\%$ ).

**Rule 4:** 118 students think that gained study materials from the mentioned web sites are an adequate tool for successful course completion and think that searching for study materials on the Internet leads the students to the greater independency ( $m_C = 68.21\%$ ,  $m_S = 43.25\%$ ).

**Rule 5:** 116 students find an electronic study sources the best for studying ( $m_C = 69.05\%$ ,  $m_S = 42\%$ ).

## IV. CONCLUSIONS

Goals of this article were to analyze the web sites use by the students for gaining the electronic study sources and to search mutual associations in the gained data set by selected algorithm.

On the basis of questionnaire among mainly students of University Pardubice data matrix of the range (400×54) was created. The answers of students were analysed. Selected results of analyse we can see in the Table 1 (Age groups of respondents), in the Table 2 (Frequencies of usage the study sources according to the age groups), in the Table 3 (Frequencies of usage the study sources and the internet for independent gaining the study sources) and in the Table 4 (Frequencies of the internet use and use of foreign web sites).

Under the search for association in the data GRI was used. There were three partial goals specified that focused on associations in the data in regard to the following questions Q2, Q7 and Q10. There were discovered several rule sets. They contented big number of rules (100) except rule set that was generated after an elimination of attributes  $a_5, a_6, a_7, \dots, a_{15}$  informed about services of the university library. Each rule in the final rule set was described by support  $m_S$ , confidence  $m_C$ , rule support  $m_{RS}$  and by number of instances. Examples of five rules were described with respect to specific goal.

On the basis of confidences  $m_C$  in the discovered rule sets that were created after above elimination, numbers of rules in intervals of confidence were the following:

- 1<sup>st</sup> goal: This rule set contented 14 rules with 100% confidence ( $m_C = 100\%$ ), 6 rules had confidence in range from 99.9% to 90%, 15 rules in range from 89.9% to 80%, 30 rules in range from 79.9% to 70%, 35 rules in range from 69.9% to 60%
- 2<sup>nd</sup> goal: This rule set contented 2 rules wit confidence in range from 99.9% to 90%, 4 rules in range from 89.9% to 80%, 4 rules in range from 79.9% to 70%, 1 rule in range from 69.9% to 60%
- 3<sup>rd</sup> goal: This rule set contented 27 rules with 100% confidence ( $m_C = 100\%$ ), 11 rules had confidence in range from 99.9% to 90%, 16 rules in range from 89.9% to 80%, 38 rules in range from 79.9% to 70%, 7 rules in range from 69.9% to 60%

We can state that association rules are usually easy to interpret. Rules in a rule set can overlap such that some records may trigger more than one rule. This allow the rule set to make rules more general than is possible with a decision trees [11].

In the future work we will focus on application of other methods for discovering of association rules and creation of classification models by selected algorithms of decision trees.

## REFERENCES

- [1] M. Nováková, Z. Jurčáková, E. Sakálová, M. Mathaeidesová. Využívání informačních zdrojov. Bratislava: Slovenská technická knižnica - Centrum VTI SR, 1995. (in Czech)

## Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)

- [2] Druhy dokumentů. Knihovna Akademie věd ČR [online]. 2013 [cit. 2013-03-15]. URL: <http://www.lib.cas.cz/asep/prozpracovatele/manual/popis-databaze/> (in Czech)
- [3] V. Sklenák et al. Data, informace, znalosti a Internet. Praha: C. H. Beck, 2001. (in Czech)
- [4] P. Hitzler, M. Krötzsch, S. Rudolph. Foundations of Semantic Web Technologies. Boca Raton: CRC Press, 2010.
- [5] H. Řezánková, Analýza dat z dotazníkových šetření. Příbram: Professional Publishing, 2007. (in Czech)
- [6] R. Nisbet, et al. Handbook of Statistical Analysis and Data Mining Applications. Academic Press, 2009
- [7] P. Guidici. Applied Data Mining: Statistical Methods for Business and Industry, West Sussex: Wiley, 2003.
- [8] O. Maimon, L. Rokach. The Data Mining and Knowledge Discovery Handbook, New York: Springer, 2005.
- [9] O. Maimon, L. Rokach. Decomposition methodology for knowledge discovery and data mining, World Scientific Publishing: London, 2005.
- [10] I.H. Witten, E. Frank. Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufman Publishers: San Francisco, 2005.
- [11] Clementine® Clementine 10.1 Desktop Node Reference, 2006.
- [12] A. Mirabadi, S. Sharifian. Application of Association Rules in Iranian Railways (RAI) Accident Data Analysis, Safety Science, Vol. 48, No. 10, 2010, pp. 1427-1435.
- [13] A. Haeri, R. Tavakkoli-Moghaddam. Developing a hybrid data mining approach based on multi-objective particle swarm optimization for solving a traveling salesman problem, Journal of Business Economics & Management, Vol. 13, No. 5, 2012, pp. 951-967.
- [14] E.M. Theodoraki, S. Katsaragakis, Ch. Koukouvinos, Ch. Parpoula. Innovative Data Mining Approaches for Outcome Prediction of Trauma Patients", Journal of Biomedical Science & Engineering, Vol. 3, No. 8, 2010, pp. 791-798.
- [15] P. Jirava, J. Mandys, J. Křupka, M. Kašparová. Modelování kvality života u osob se zdravotním postižením v kontextu regionálních strategických dokumentů, Scientific Papers of the University of Pardubice, Series D, Faculty of Economics and Administration, Vol. 17, No. 2, 2010, pp. 76-86. (in Czech)
- [16] Clementine® Clementine 10.1 Algorithms Guide, 2006.
- [17] J. Han, M. Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers. San Francisco, 2001.