International Journal of Advanced Statistics and IT&C for Economics and Life Sciences

Volume 6
Number 2
December 2016

ISSN 2067 – 354X

Volume 6 **Number 2 2016** **ISSN 2067 – 354X**

International Journal of Advanced Statistics and IT&C for Economics and Life Sciences

Editor

Daniel Volovici

Lucian Blaga University of Sibiu, Romania

Associate Editor

R. Crețulescu, Lucian Blaga Univesity of Sibiu, Romania

International Journal of Advanced Statistics and IT&C for Economics and Life Sciences

ISSN 2067 – 354X

Frequency: Published 2 issues a year

ELEMENTARY SCHOOL PUPILS: FROM PASSERS-BY IN LIBRARIES AND OTHER CULTURAL HERITAGE INSTITUTIONS TO THEIR END-USERS

Denis Vincek [1]

[1] *"Osnovna škola Ante Kovačića, Vladimira Nazora 1, HR-49250 Zlatar, Croatia*

Abstract

In terms of class organization, an important factor for creating the conditions for the teaching process is the venue at which the teaching takes place. School field trips represent a special type of teaching, which involves visits to museums, libraries and other cultural heritage institutions. The aim of this research, conducted in a Croatian elementary school, is to determine how well the pupils remember their visits to these institutions and how motivated they are for such trips. Taking into consideration that the research involved 41 participants who visited a total of 31 museums, libraries and other institutions during the course of 8 years, the total number of possible answers in the survey was 1271 and the participants only provided 239 answers, which means that they remembered only 18.8 percent of institutions visited. The survey also probed the students to see how they prepared for these visits during their school excursions and field trips. Out of 41 participants, 16 reported that their homeroom or class teachers had acquainted them with the institution they would visit; only 4 students obtained the information on their own from online or some alternative sources; and 21 students, more than a half, did not try to obtain any information about the institution they would visit. A way to change this lies in appropriate intellectual and emotional motivation of students, and one of the possible concrete measures is the project "A Backpack Full of Culture", conducted by the Ministry of Culture of the Republic of Croatia.

Keywords: Libraries, Cultural Heritage Institutions, Motivation, Backpack Full of Culture.

1. Introduction

The aim of this research, conducted in an elementary school in the Krapinsko-zagorska county in Croatia, was to determine how well the elementary school pupils remember their visits to museums and other cultural heritage institutions and to what extent they are motivated for such visits. The intention was to test the following hypotheses: 1. The eighth grade pupils remember less than one quarter of libraries, museums and other cultural heritage institutions they had visited during their eight years of education; 2. The students were not prepared in schools for their trips to cultural heritage institutions; 3. After their visits to libraries, museums and other cultural heritage institutions, the students did not spend any more time on this topic.

2. Why visits to libraries, museums and other cultural heritage institutions

In terms of class organization, an important factor for creating the conditions for the teaching process is the venue at which the teaching takes place. The modern notion of student-oriented education thus allows the classes to be held not just within the school, which implies both the school premises and the facilities outside it, but also conducting practical education in different institutions and companies. School excursions and field trips represent a special type of classes. The difference between field trips and excursions is that trips are used as a form of psychological and physical recreation of students, while excursions are used to study certain parts of the curriculum in their essential form [1], which is something that cannot be accomplished in the framework of traditional classroom setting.

In accordance with the Primary and Secondary School Education Act and the School Statute, the School Board passes the School Curriculum based on the proposal of the Board of Teachers and a positive review from the Board of Parents. This document defines the syllabi for the elective subjects, extracurricular activities and other educational activities, programs and projects, according to the guidelines of the Croatian National Education Standard. Accordingly, all student excursions and field classes are listed in the School Curriculum of every school.

The Primary and Secondary School Education Act states that the educational activity in the school is based on the autonomy in planning and organization, and freedom of pedagogic and didactic work. This means that the plans for excursions and field classes differ from school to school, but have to be done in accordance with the national curriculum, the national pedagogical standards and the syllabi. The Primary and Secondary School Education Act stipulates that for every excursion and field trip the following aspects need to be listed and described in detail in the School Curriculum: aims, purpose, holders, means of realization, time schedule, detailed list of expenses and means of evaluation.

Within the scope of excursions and field classes, the elementary school students from the first to the eighth grade visit various museums, libraries and other cultural heritage institutions and their goal, that is, the didactic importance of these types of classes, is not just to learn about the cultural heritage at the place of its preservation, in order to enhance the intellectual and emotional experience of the students, but to motivate students for (subsequent) visits to museums, libraries and other cultural heritage institutions.

3. Research methodology

The survey questionnaire contained 14 questions, 13 of which were closed-type questions and only one of which was an open-type question in which the students had to list museums and other cultural heritage institutions that they had visited during the previous eight years of their education. Closed-type questions with multiple answers and answers for level of intensity were used. For the purpose of this research, intentional, convenience sample was used [2], which means that its representativeness and sufficiency should be taken somewhat loosely. The survey questionnaire was

filled out (during their homeroom classes) by 23 students of class 8a and 18 students of class 8b. Since no statistically significant difference in distribution of answers was noted between the two classes, all questionnaires (N = 41) were processed together.

4. Results and discussion

The first question required the students to list the museums, libraries and other cultural heritage institutions that they had visited during their school excursions and field classes during their eight-year education. As a help to remind them what this refers to, the cultural heritage institutions covered by this survey included [3], apart from museums, galleries, libraries, sacral objects, old towns, castles, ethno villages, ethnographic collections, national parks etc., which were also listed in the question for the students.

Figure 1. Number of participants that remembered a visit to a particular museum or other cultural heritage institution.

Figure 1 contains the names of all cultural heritage institutions with their original names in Croatian. Their English equivalents are provided here, in order in which they are listed in Figure 1: Church of st. Philip and Jacob, Vukovar; Ethno village Skradinski buk; **Sučić** Family Ethnological **Collection;** City Museum, Sisak; Croatian

War Museum, Karlovac; City Museum, Varaždin; Franciscan monaster, Vukovar; Ethno village, Rastoke; Peasants' Revolt Museum; Memorial Home, Vukovar; Homeland War Memorial; Old Town, Sisak; Old Town, Čakovec; National Park Brijuni; National Park Krka; Zagreb Cathedral; Vukovar Hospital Memorial; Vukovar Hospital – Place of Remembrance; The **Church** Of The **Mother** Of God Of Gorje, Lobor; Krapina Neanderthal Museum; Zagreb City Museum; National Park Plitvička jezera; Oršić Castle; Old Town, Varaždin; Ovčara; Croatian National History Museum; St. Jacob's Cathedral, Šibenik; "Staro selo" Museum, Kumrovec; Stork Village, Čigoč; Trakošćan Castle; Technical Museum, Zagreb; Archaeological Museum, Zagreb

As was mentioned, all museums, libraries and other cultural heritage institutions that the students had visited as a part of their excursions and field classes are listed in the School Curriculum. For the purpose of processing the answers provided for this question, the researcher conducted the interview with the homeroom teachers of both classes who taught students from the fifth to the eighth grade, their teachers from the first to the fourth grade and with two eighth-graders to generate a control list which contained all the museums and other cultural heritage institutions that the participants visited during their education, before administering the questionnaire. The list included 31 museums and other cultural heritage institutions.

The chart in Fig. 1 shows the number of participants that remembered and wrote the correct name of each cultural heritage institutions. It is obvious that out of 41 participants, the most (34) remembered the Archaeological museum in Zagreb. Out of 31 cultural heritage institutions visited, five institutions were not remembered by a single participant. These are the church of St. Philip and Jacob in Vukovar, Ethno village Skradinski buk, the Sučić family Ethnological collection, the Sisak Town Museum and the Croatian War Museum in Karlovac. Taking into consideration that the research involved 41 participants who visited a total of 31 museums and other cultural heritage institutions during the course of 8 years, the total number of possible answers in the survey was 1271 and the participants only provided 239 answers, that is, only 18.8 percent. One participant listed 13 institutions, which was the highest figure, while one participant listed only one institution (the Archaeological museum in Zagreb). No participants left this question unanswered. Taking into consideration the number of participants and given answers, the average number of museums and other institutions that the participants were able to name is 5.829 (out of 31), i.e. 18.8 percent of all museums and other cultural heritage institutions that they visited during their elementary school education. When asked how long on average their visits to particular institutions were, 80 percent of participants stated that the visit lasted more than 45 minutes, 33 participants (80 %) thought that this was enough time to see the exhibits, while 8 participants did not share this sentiment. As much as 95 percent of participants said that they always, or in most cases, had a guide during such visits. The questions about the duration of visit and the professional guides for cultural heritage institutions were asked to gain further information about the quality of organization of the visit itself. These were actually control questions as the answers provided were not in correlation with the hypothesis, but they still indirectly indicate that a visit to a cultural heritage institution was organized with the intention of users learning as much as they could during their visit, that is, with the intention of achieving the learning outcomes as defined by Bloom and other taxonomies for individual learning domains. As much as 83 participants reported that they heard the

guides well, 12 percent did not hear the guides as they were too far from them and 5 percent said they did not hear the guide well as they were talking too quietly. When asked about their opinion on tour-guided visits to institutions, 71 percent of the participants said they were a good thing as they allowed them to learn more, 24 percent thought they were not a good thing as the guides gave too many information in too little time, while the words that the guides used were too complex and incomprehensible for 5 percent of the participants. The students were also asked how they prepared for these visits to cultural heritage institutions during their school excursions and field trips. Out of 41 participants, 16 reported that their homeroom or class teachers had acquainted them with the institution they would visit; only 4 students obtained the information on their own from online or some alternative sources; and 21 students, more than a half, did not try to obtain any information about the museum or cultural heritage institution they would visit. It is interesting to note that as much as 21 (out of 41) participants did not know that the institutions had guest books in which they could leave their opinions on the visit. The following range of questions wanted to determine whether the students stopped thinking about what they saw in the cultural heritage institutions after leaving their premises. The answers have shown that, just as was the case with the previous question, more than half of students (24 of them, to be precise) stop reading about the cultural heritage institution after their visit. Only four students read about them on websites or through other sources – and these are the same students that look for information before the visit – while 13 students reported that they talked about what they saw in their respective classes. The next question the students were supposed to answer was whether they had to write a school report on what they saw during their visits. Only one student said "yes", 8 students said "sometimes", another 8 students said "rarely", while 24 students, that is more than a half, said they never had to write a report. The one student that does write the reports is the student that covers the visits for the school website and the school newspaper. The websites of many school post photo-galleries from excursions and field classes, which use photos from visits to museums and other cultural heritage institutions. The Internet is obviously the medium that the students visit frequently – and this applies to the school website as well – 14 students said they browse the photos on the school website after a visit to the cultural heritage institution, 13 students reported they browse them occasionally, 8 students rarely browse them, and 6 students never browse them. Another piece of data indicates that the students want to have a memento of the visit. As much as 83 percent of students said they buy souvenirs from the cultural heritage institutions in order to have something to remember their visit by, the remaining students buy them as presents, while only one student covered by this study never buys souvenirs. The last question wanted to check the attitude that the students have towards visiting libraries, museums and other cultural heritage institutions. The results are as follows: for 6 students, these are the most boring parts of visits; 11 students reported that they are not interested for such visits, but they are nonetheless part of their field classes; while 12 students do not regard them as very important. Only 12 students, i.e. 29 percent of participants, states they are very interested in visiting cultural heritage institutions.

5. Conclusion

This study, albeit conducted on a small sample, wanted to investigate whether the students are at all motivated for visiting cultural heritage institutions as a part of their school excursions and field classes during elementary school education and whether the students are the real end-users of these cultural heritage institutions or whether they are mere passers-by in museums, libraries, galleries, churches, nature parks, ethological villages, national parks etc. simply because someone else decided they should visit them. After all, the author of this paper has on more than one occasion noticed while correcting the student reports for school newspaper and website that the visits to cultural heritage institutions are given the same amount of words in a text as a description of a restaurant where they had lunch that day and/or a McDonald's restaurant, which is an inevitable part of every such field class. The fact is that these outings have to be in line with the school syllabi as their purpose is not, as was already mentioned, primarily recreational. The research has confirmed the first hypothesis. Not only do the eighth-graders remember less than a quarter of the cultural heritage institutions they had visited during their eight years of education, the results are even more disastrous – they were able to name only 18.8 percent of visited museums and other cultural heritage institutions. The remaining two hypotheses have also been confirmed: more than a half (51 percent) of participants did not seek any information about the museum or the cultural heritage institution before the visit; more than a half (59 percent) of participants did not spend any time after the visit to find any information about the institution they had visited recently. As much as 59 percent of students stated that they did not have to write a school report about their visit.

How to change this? Can this really be changed if only 29 percent of students claim that the visits to museums, churches, galleries, ethno villages, ethnographic collections, national parks are very interesting to them?

That things are not necessarily so bleak is reflected in the fact that 83 percent of students buy a souvenir during such visits as they want to have a memory of the cultural heritage institution they had visited, and 85 percent of students always, sometimes or rarely browse the photos from these visits on the school website. The answer lies in the motivation as the school excursions and field classes that involve visits to museums and other cultural heritage institutions need to contain an appropriate motivating factor as it is also necessary that motivation is the introductory part of every class. Here, we should differentiate between intellectual motivation which implies introduction into what is going to be seen in the library, museum, or other cultural heritage institution at the cognitive level – e.g. by asking questions to which the students will find answers during the visit – and emotional motivation, which implies "creating the emotional environment in the class, as well as positive surroundings and incentive for learning" [4]. However, this is not something that can be achieved overnight. An excellent project that could help motivate students to visit museums, libraries and other cultural institutions is “A Backpack Full of Culture” [5]. “A Backpack Full of Culture” is a program that allows the children and youth from places with limited availability art and culture programs to get acquainted with them better. This is a joint program conducted by the Ministry of Culture and the Ministry of Science, Education and Sports as a supplementary program for kindergarten, elementary and high school curricula. The program's activities are: theatre, film,

music, dance, visual arts, literature, cultural heritage, and programs from students of the art academies. The implementing actors of the program are professional artists and students of the art academies with their professors and (most commonly) librarians acting as mentors. One of the four expected outcomes of this project is to make pupils more aware of arts and culture; in other words, to stop them from being mere passers-by in museums, libraries and cultural heritage institutions and to help them become the real end-users of these institutions.

6. References

[1] V. Poljak, Didaktika, Školska knjiga, Zagreb, 1985.
[2] A. Tkalac Verčić; D. Sinčić Čorić; N. Pološki Vokić. Priručnik za metodologiju istraživačkog rada u društvenim istraživanjima: kako osmisliti, provesti i opisati znanstveno i stručno istraživanje, M. E. P., Zagreb, 2011.
[3] I. Maroević, Uvod u muzeologiju, Zavod za informacijske studije Odsjeka za informacijske znanosti, Filozofski fakultet Sveučilišta, Zagreb, 1993.
[4] M. Matijevič, D. Radovanović, Nastava usmjerena na učenika : prinosi razvoju metodika nastavnih predmeta u srednjim školama, Školske novine, Zagreb, 2011.
[5] Ruksak pun kulture. URL: http://www.min-kulture.hr/default.aspx?id=8787. (20.03.2016.).

PERCEPTION OF LIBRARY 2.0: THEORETICAL CONCEPT OR PRACTICAL MODEL FOR INNOVATIVE AND SERENDIPITOUS DISCOVERY SERVICES? DRIVERS AND IMPACTS IN THE ROMANIAN LIBRARIES

Voichița Dragomir [1]
[1] *"Carol I" Central University Library - Bucharest, Romania*

Abstract

Library 2.0 is a theoretical concept launched by Casey in 2005, which has been transformed into a practical model applied worldwide by the libraries ready to move toward major change and innovation, willing to redefine and revitalize their services in accordance with users' needs and expectations, accepting users' participation as communication partners and knowledge contributors. Library 2.0 represents a virtual set of library user-centered services, built on solid principles like radical trust, collective intelligence, creative thinking, collaboration, openness, community knowledge management, content free use and reuse, social networking, people interactivity and feedback encouragement. It is supported by the Web 2.0 technologies, configured as a public sphere, providing an innovative platform for knowledge dissemination for library users and nonusers with the aim of bringing them into the library virtual or physical environment. Library 2.0 uses both the Web 2.0 tools and specialized instruments like the Next Generation Catalogue or the open access information repository which offer unexpected, serendipitous ways of information discovery.
The Library 2.0 challenge has been taken up by the Romanian libraries. The paper presents a brief overview of the Romanian libraries response to this challenge, investigating the status of Library 2.0 tools acceptance and implementation, pointing out the Library 2.0 drivers and impacts in the Romanian information and documentary structures.

Keywords: Library 2.0, Romanian libraries, Next Generation Catalogue, OPAC 2.0, Discovery Layer, Technology usage.

1. Introduction

Since 2005, when Michael Casey[2] came up with the "Library 2.0" term, the library professionals within the library environment worldwide have been very active in debating the role of Library 2.0 model, its functional characteristics and social features, its benefits and impacts on the library services and users, its relation to other traditional or modern library models. The libraries have adopted and implemented in different ways the Library

2.0 model, willing to overcome the threat coming from the Internet search engines, applications, and services, being ready to accept fundamental changes and innovation, to improve the library services and the library users relationship, to transform the library into a more flexible, visible, open, communicative, socialized and intelligent organization, to demonstrate that the Library 2.0 approach can really "make the library human, ubiquitous, and user-centered".[5]

The Library 2.0 model is supported by the Web technologies, following the same Web 2.0 underlying principles, it opens up unforeseen possibilities of sharing ideas, and knowledge, of discovering and accessing information which can be freely used and reused. It is a community-building environment which enables, beyond asynchronous and synchronous interaction, a real bi-directional communication between the library and the patrons. Both the users and the nonusers are invited to participate and contribute, they are getting a face and an identity emerged from the compact mass of the people, becoming equally consumers and producers of information, the so-called prosumers [12] Toffler was talking about.

2. Library 2.0 – theoretical concept. Literature review

Charles Cutter wrote in 1876, in his "Rules for a Dictionary Catalogue", about the principle of the "convenience of the public" [3]. More than one hundred years later, the Functional Requirements for Bibliographic Records conceptual model was set to better respond to the users' needs in their attempt to find, identify, select and obtain information, while IFLA takes up, in 2009, in the "Statement of International Cataloguing Principles" document, the Cutter's concern regarding the convenience of user in the bibliographic catalog construction principles.

Following this principle, the Library 2.0 model is above all, a user-centered platform, focused on its customers' expectations. Casey and Savastinuk state that "the heart of Library 2.0 is the user-centered change" [2], considering this model as a virtual and physical service for the next generation libraries. They define Library 2.0 by three elements: constant change, users' empowerment through participation and cumulative customer-driven services attempting to reach the potential users and better serving the current ones.

According to Maness, Library 2.0 can be defined as "the application of interactive, collaborative, and multimedia web-based technologies to web-based library services and collections." [8] Another point of view is provided by Brevick [1] who considers that Library 2.0 represents a natural evolution of the library services at a level in which the user has the control over the access to information and library services in terms of time, and method.

There are some critical voices which argue that Library 2.0 is a slogan without substance [4] and represents for the librarians only a game with the Web 2.0 technologies [4]. Nesta and Mi [9] are not enthusiastic about the Library 2.0 benefits, pointing out that libraries have always been focused on the users' needs by opening their doors to longer

hours and offering open stacks, computers labs, online access catalogues, group study rooms, instruction sessions, reference services by e-mail, forums and discussion lists, library tours, brochures. Even the Ranganathan's laws are a proof of the reader-oriented library approach. More than that, Gorman and Crawford [10] added other five laws ("Libraries serve humanity", "Respect all forms by which knowledge is communicated", "Use technology intelligently to enhance service", "Protect free access to knowledge", "Honor the past and create the future"), Noruzi proposed a new interpretation of the laws in the web environment and Simpson [11] suggested an updated form determined by the media richness ("Media is for use", "Every patron his information"). An OCLC report suggests that the 5 laws have to be reinterpreted „to reflect the today's library resources and services, as well as the behaviors that people demonstrate when engaging with them".[13]

The Library 2.0 model [7] is, according to Lankes, a participatory library, enabling an interactive communication between users and library that become, alternatively, transmitter and receptor of a message. The Academic Library 2.0 Concept Model v2 proposed by Michael Habbib [6] refers to the Library 2.0 as a collaborative space where the boundary between physical and virtual has vanished. This approach emphasizes the fact that Library 2.0 does not replace the library; it is a part of it, a subset of the library services, which supplements and improves them. The Farkas [5] model underlines the essence of Library 2.0 which consists in focusing on patrons' needs, not on the Library 2.0 tools.

3. Library 2.0 – practical model

Resuming the definitions and the theoretical models shaped by different authors, the Library 2.0 model can be considered a Habermasian public sphere transposed into the virtual environment of knowledge. In this context, the most important elements of the Library 2.0 model could be: users' empowerment through participation in the creation of content, a continuously growing collection of full-text electronic resources accessible online, a more interactive communication between users and library, change and innovation, next-generation library tools.

In the real life, the Library 2.0 practical model is sustained by the following solid pillars:

- Web 2.0 technologies, as part of everyday life online;
- Extensive Websites for information dissemination, online access to local&worldwide resources and services, communication with the library users and nonusers, specialized assistance through e-mail and chat, digital marketing of library services, events and exhibitions promotion, international visibility;
- Discovery tools for innovative and serendipitous resource dicovery services:
    - Discovery Interfaces, so called Next Generation Catalogues or OPAC 2.0 – tools which operate at a similar level of sophistication and atractivity as Google, Amazon or other popular Web sites, providing an intelligent and efficient platform for resource discovery;

    - Web-scale discovery services or Index - based discovery services, like Summon, Ebsco Discovery Systems, WorldCat Discovery Service and Primo Central which provide a unique point of entry to all library resources through a single consolidated index including bibliographic descriptions of the collections items, locally managed by the library integrated system, electronic bibliographic and full-text resources harvested from the subscribed databases and from the library's digital repositories;
- Federative search tools, like Metalib and Millenium Access Plus, for information searching and retrieval across multiple, ***heterogeneous*** and distributed library databases;
- Digital platforms made up of digital libraries and institutional repositories, including and mixing digital objects in various formats, which are described using technical, preservation and administrative metadata, representing digitized items selected from the library's collections and born-digital resources supplied by the librarians, researchers, professors and students.

Being worldwide accepted and adopted by academic and public libraries, the model demonstrates its viability in the real life, providing enhanced library services, changing attitudes, giving voice to the library users and putting in place new information discovery tools customized for the users' needs and their search behavior.

4. Drivers and impacts in the Romanian libraries

The Romanian libraries are trying to keep pace with the rapidly changing of technological environment and to develop the library services according to the new trends on the international level.

In this respect, the Romanian public libraries have highly adopted Web 2.0 tools. Among the 42 public libraries which have been analyzed (National Library and Metropolitan Library are included), there are only 2 libraries with no web 2.0 technology, 40 libraries are using Facebook (in 2015, 20 public libraries implemented Facebook), there are 14 blogs, 9 YouTube, 5 Twitter, 4 Flikr, 4 IM, 3 RSS and 1 installation of Picasa, Slideshare, Scribd, G+. In total, the public libraries are using 13 Web 2.0 technologies. There are 17 libraries interested in implementing only one Web 2.0 tool, 13 libraries have adopted 2 Web 2.0 technologies, two public libraries offer 6 Web 2.0 tools and one library is using 7 Web 2.0 tools. The graphic indicates that the interest of the Romanian public libraries in setting up a participatory institution is growing from one year to another.

Library integrated systems are implemented in 31 public libraries (73%), 21 of these choosing to work with Next-Generation Catalogues: TinRead is used in 17 libraries, eBibliophil is implemented in 6 libraries and Qulto in 2 libraries. There are also 3 Liberty installations, 2 Aleph library systems, 2 TinLib systems and 2 installations of Qulto - the older version.

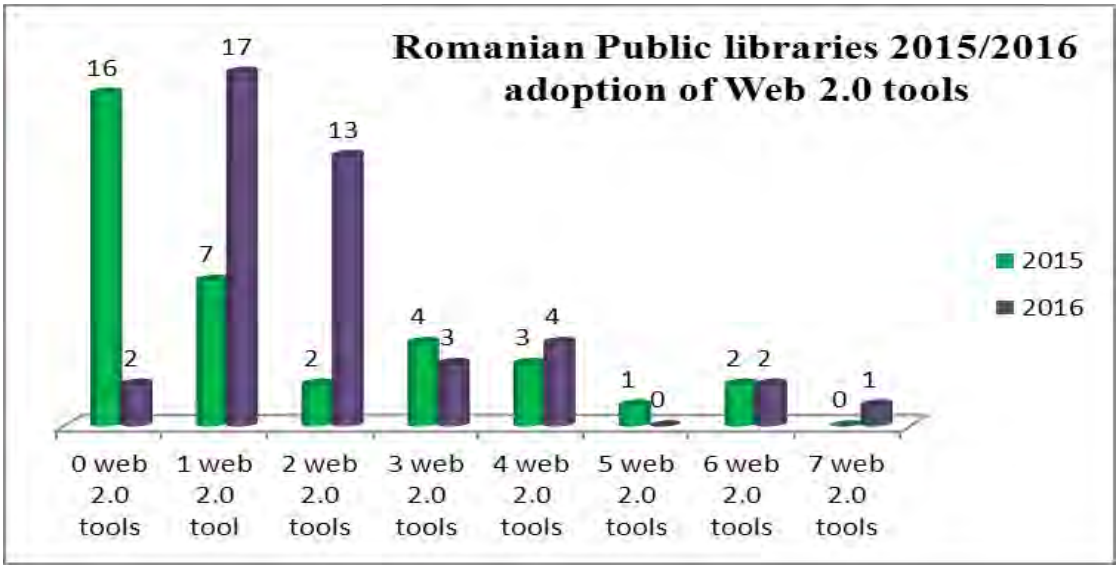

Figure 1 Adoption of Web 2.0 tools by the Romanian Public Libraries – 2015/2016

In contrast to public libraries, the rate of adoption of the Library 2.0 model by the Romanian academic libraries is quite low. The assessment of 58 university libraries, including the four central university libraries in Bucharest, Iasi, Cluj and Timisoara reveals that 36 libraries (62%) have no Web 2.0 tools, 19 libraries have 2 Web 2.0 tools and only 3 libraries are interested in sharing information through 2 web 2.0 channels. The range of Web 2.0 technologies chosen by the academic libraries is narrow, only 4 web 2.0 tools being used in this environment.

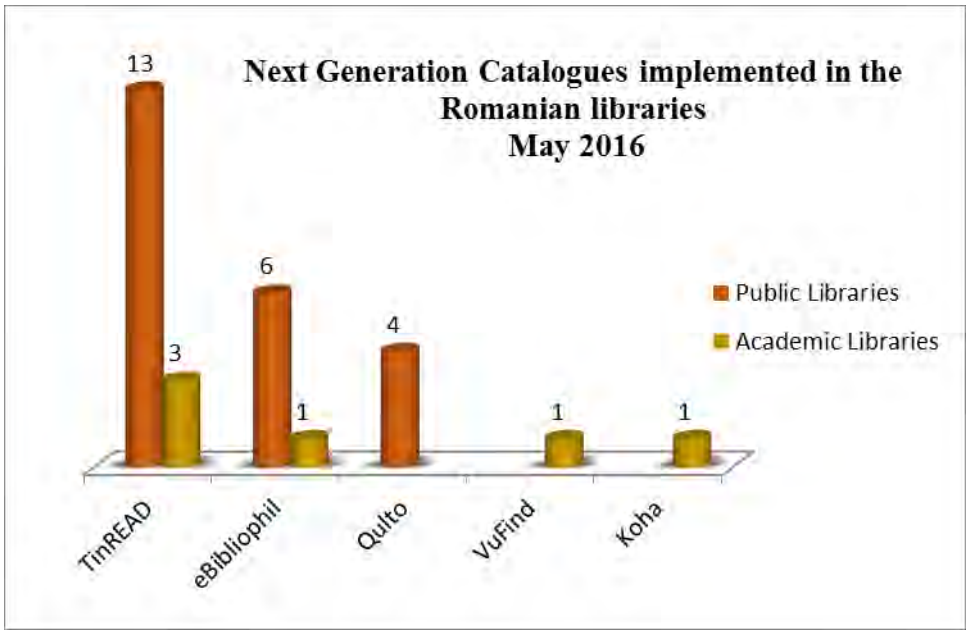

Figure 2 Next Generation Catalogues in the Romanian Libraries

Regarding the implementation of discovery layers systems, there are four Next Generation Catalogues functioning in 6 libraries: TinRead is used in 3 libraries,

eBibliophil is installed in 1 library, Koha was recently implemented in 1 library and VuFind is used in 1 library. The other academic libraries are using automated systems like Liberty (19 installations), Aleph (8 installations), TinLib (3 installations), VubisSmart (1 installation), Alice (1 installation) and Alephino (1 installation).

In the academic environment, there are two Web-scale Discovery services which are used within the ANELIS PLUS consortium, providing users with a one-stop shop searching over the virtual collection of the library's accessible and subscribed resources. EDS (Ebsco Discovery System) and Summon are the Web-scale discovery services available for the affiliated libraries, offering to their users access to the subscribed scientific databases based on IP address within the library premises and mobile access through a personal account and password.

Federated search services are provided at national level through Rolinest and Biblio.ro. The Rolinest portal is built on Metalib product, whereas Biblio.ro is sustained by TinRead, both platforms allowing the users to submit a single query which performs a search in multiple distributed and heterogeneous databases, displaying real-time, aggregated results through a unique interface.

Another important pillar of the Library 2.0 practical model consists of digital platforms built for storing and managing digital objects and the associated metadata, representing the cultural heritage hosted by the library and/or the intellectual output of both library and host university. In Romania, there are 8 institutional repositories registered in ROAR (Registry of Open Access Repositories) based on Dspace open software: APAS/SNSPA, ARTHRA/Dunarea de Jos University - Galati, IRCULB / Central University Library "Carol I" Bucharest, ASsee Online Series – SNSPA, ICESBA / Fundatía Romania de maine and NOS / World Economy Institute.

There are also several digital libraries developed by different libraries and cultural institutions. Among them, National Digital Library developed by the National Library, DacoRomanica set up by the Metropolitan Library, Restitutio set-up by the Central University Library "Carol I", e-Patrimoniu developed by CIMEC- the national aggregator for Europeana. At the moment, the Romanian contribution to Europeana is very low - 172.186 digital objects [14], in January 2016, representing 0.4% of total minimum contribution set in the Commission Recommendation. The most significant contribution comes from the Central University Library "Lucian Blaga". The Culturalia project, initiated by CIMEC, has as main objective for 2020 to open up a significant critical mass of over one million digital objects supplied by the Romanian libraries and cultural institutions.

5. Conclusion

The Library 2.0 challenge has been taken up by the Romanian libraries in different manners. The public libraries are more flexible and active, much more open to innovation and change in opposition to academic libraries which are still conservative,

know-it-all and rigid institutions, still struggling to adopt the 2.0 technologies, with some exceptions.

Recognizing that the main mission of the library can be better served by providing access to information and facilitating creation of knowledge through participation, communication and content sharing, the Romanian libraries are engaged in setting up the Library 2.0 model as a user-centered, network-based platform delivering rich library content and library enhanced services.

6. References

[1] BREVIK, Thomas. Library 2.0 = MyLibrary, 2006, Available at: https://lib1point5.wordpress.com/2006/04/12/library-20-mylibrary/

[2] CASEY, Michael. Working towards a definition of Library 2. In: LibraryCrunch, October 21, 20015. Available at: http://www.librarycrunch.com/2005/10/working_towards_a_definition_o.html

[3] CUTTER, Charles. Rules for a Dictionary Catalogue. Available at: https://archive.org/details/rulesforadictio06cuttgoog, p. 6

[4] CRAWFORD, Walter. Library 2.0 and 'Library 2.0". In: Library 2.0: A Cites and Insights Reader, Vol. 6, No. 2: Midwinter 2006

[5] FARKAS, Meredith. Web/Library 2.0 backslash. In: Information Wants to be Free, 1.12.2005. Available at: http://meredith.wolfwater.com/wordpress/2005/12/01/weblibrary-20-backlash/

[6] HABIB, Michael. Conceptual model for academic library 2.0. In: LIS, 2006. Available at: http://www.mchabib.com/2006/06/09/conceptual-model-for-academic-library-20/

[7] LANKES, David ; Silverstein, Joanne ; Nicholson, Scott Participatory networks: the library as conversation. In: Information Technologa and Libraries, december 2007, p. 17-33. Available at: http://ejournals.bc.edu/ojs/index.php/ital/article/view/3267/2880

[8] MANESS, Jack. Library 2.0 Theory: Web 2.0 and Its Implications for Libraries. Available at: http://www.webology.org/2006/v3n2/a25.html

[9] NESTA, Frederick ; Mi, Jia. Library 2.0 or Library III: returning to leadership. In: Library Management, Vol. 32 Iss ½, 2011. p. 85 – 97. Available at: http://dx.doi.org/10.1108/01435121111102601

[10] SEN, B. K. Ranhanathan's five laws. In: Annals of Library and Information Studies, Vol. 55, June 2008, p. 87-90. Available at: http://nopr.niscair.res.in/bitstream/123456789/1769/1/ALIS%2055%282%29%2087-90.pdf

[11] SIMPSON, Carol. Five laws. In: Library Media Connection, April/May 2008. Available at: http://www.carolsimpson.com/5laws.pdf

[12] TOFFLER, Alvin.Al treilea val. Bucuresti: Antet, [s.a.], 320 p.

[13] Reordering Ranganathan: Shifting User Behaviors, Shifting Priorities. OCLC, 2014, Available at: http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-reordering-ranganathan-2014.pdf

[14] http://ec.europa.eu/information_society/newsroom/image/document/2016-6/contribution_to_europeana_13733.pdf

THE REFERENCE INSTRUMENTS IN SCIENTIFIC PAPERS

Teodora Sabău [1]
[1] *"Central University Library, "Lucian Blaga" University of Sibiu, Romania*

Abstract
The article mainly refers at the reference instruments as bibliographic elements which are being used in the elaboration of the scientific papers. The aim of the study is to establish the parts and the corresponding format for a good bibliographic use of the documentary sources used in the scientific activity. Are highlighted the differences: citation /bibliographic description for reference, reference list / bibliography and the role of the ISO 690 as a flexible standard which does not impose a fixed style but basic recommendations in the bibliographic structure of used sources, resulting in a variety of international editorial styles. The major conclusion of the study is that there is the possibility of adapting the international standard requirements to a personal style of using the reference instruments.

Keywords: Reference instruments, Citation, Bibliographic description, Reference list, Bibliography, ISO 690, Editorial styles.

1. Argument

The scientific papers need references, which are being used in some specific ways that bibliographically represent the consulted documentary sources in study elaboration.

These ways can be considered instruments and need to be known enough for a good use:

- The citation and its methods
- The reference list
- The bibliography
- The standard of ISO 690
- The editorial styles or the recommendations for authors

2. The reference: definition, features, types and role

2.1 Definition:

The reference represents a set of bibliographic information about the source of a quotation, concept or of an idea mentioned in the paper or the source of a simple documentation for the study.

2.2 Features:

- It covers all types of documents in any medium;
- It consists of the bibliographic description of documents in accordance with the international standard of ISO 690.[1]

2.3 Types:

The types of reference depending on placing it:

- in text (between brackets);
- in footnotes or endnotes of a chapter;
- in the end of whole text: reference list or bibliography.

2.4 Role:

- to investigate the information flows;
- to recognise the documentation, placing the own work in the context of research.[2]

3. The reference instruments

3.1 The citation: definition, elements, methods and role

3.1.1. Definition:

The citation is a short reference inserted in text according to citation methods.

3.1.2. Elements:

The elements of citation are:

- idea or paraphrase - to introduce the idea / ideas in your own words;
- a direct quote: short (used with quotation marks) or long (in a new paragraph).

3.1.3. Methods:

a. the first element and date: author - date (Harvard), including author - title or author - page (for human sciences);
b. the numeric citation or sequential numeric system (Vancouver), with numbers in square/round brackets, or superscript;
c. the footnotes / endnotes citation.[1]

3.1.4. Role:

- to emphasize the author work
- to respect the copyright
- to highlight the scientific productivity of authors expressed by the Hirsch index or H-index [1]

3.2 The reference list and the bibliography: similarity and difference

A. The reference list:

- All cited sources with an extended bibliographic description and in relation with citation methods:
    - in alphabetical order of the first element: the surname of the author/ first word of the title;
    - in numerical order of citation in text (for numerical method).

B. **The bibliography**:

- List of consulted sources in the study elaboration or having relation with the research subject, not necessarily cited

Similarity:

- Both have the same bibliographic format.

Difference:

- The reference list consists only of cited sources.

3.3 ISO 690 Standard – the bibliographic regulatory instrument

Is addressed to the authors and editors and gives general rules in bibliographic description for the reference in all documents (book, magazine, chapter, article, electronic documents, web page etc.) about content, structure and form given by:

- identification elements of a document (author, title, publication data);
- order of rendering the bibliographic elements;
- specific/conventional punctuation.

Provides the citation methods of references, but it **does not impose a determined / exact style** in citation systems or in bibliographic format, only a few conditions about the consistency of style, sufficiency and accuracy of information. [1]

A limited bibliographic description for the most usual type of document (book) can be rendered by the following scheme:

Author Surname, First name. ***The title of document: Information about the title.*** **Place of publication: Publisher, Year of publication.**

4. The editorial styles: citation systems with illustration

In the current editorial practice there are a variety of formats and citation styles.

Also in the citation styles is being used one or more methods or systems citation.

4.4 International editorial styles

We list some of the most popular citation styles belonging to some prestigious institutions, indicating the citation systems used [1]:

1. **ACS Style** (American Chemical Association): the largest scientific society in the world:

- All three citation methods: in the chemical, physical sciences;
- In magazines: sequential numeric system is preferred.

2. **AMA Style** (American Medical Association):

- Sequential numeric system: medical sciences.

3. **APA Style** (American Psychological Association):

- Author – date: psychology and social sciences (education, communication, politics, economics etc.).

4. **Chicago/Turabian Style**: Natural and Social Sciences:

- Author – date: natural sciences, physical and social sciences;
- Note - Bibliography (University of Chicago): human sciences (literature, history, art).

5. **CSE / CJE Style** (Council of Science Editors / Publishers Council in Biology):

- Author – date;
- Sequential numeric system.

6. **MLA Style** (Association of Modern Languages of America):

- Author - page: human sciences (literature, art, philosophy, religion).

7. **Vancouver Style** (The International Committee of Medical Journals Editors):

- Sequential numeric system: medical, biological and physical sciences.

Each style uses distinct bibliographic formats according to the specific of each type of document.

APA	**in text:** (Austin, 1998) **reference list:** Austin, J. H. (1998). *Zen and the brain: Toward an understanding of meditation and consciousness.* Cambridge, MA: MIT Press.
Chicago (Note)	**note:** 1. Tom Nairn, *Faces of Nationalism: Janus Revisited* (London: Verso, 1997), 17. **bibliography:** Nairn, Tom. *Faces of Nationalism: Janus Revisited*. London: Verso, 1997.
Chicago (Author-Date)	**in text:** (Nairn 1997) **reference list:** Nairn, Tom. 1997. *Faces of Nationalism: Janus Revisited*. London: Verso.
MLA	**parenthetical reference:** (Perle 183-185) **bibliography:** Perle, George. *Serial Composition and Atonality: An Introduction to the Music of Schoenberg, Berg, and Webern*. 6th ed. Berkeley: University of California Press, 1991. Print.

Fig. 1 Book: single author - Williams College Libraries. Books. 9 December 2011. Available: http://library.williams.edu/citing/formats/books.php, accessed March 15, 2016

APA	**in text:** (Henry, 1990) **reference list:** Henry, W. A., III (1990, April 9). Beyond the melting pot. *Time, 135*(4), 28-31.
Chicago (Note)	**note:** 3. Scott Spencer, "Childhood's End," *Harper's*, May 1979, 16. **bibliography:** Spencer, Scott. "Childhood's End." *Harper's*, May 1979, 16-19.
Chicago (Author-Date)	**in text:** (Birnbaum 1998) **reference list:** Birnbaum, Jeffrey H. 1998. "How to Succeed at Damage Control." *Fortune*, March 30, 173-176.
MLA	**parenthetical reference:** (Lerner 42) **bibliography:** Lerner, Barbara. "America's Schools: Still Failing After All These Years." *National Review* 15 Sep. 1997: 42+. Print. *Note: use the + sign if the pages are not consecutive; 42-44 would be the correct way if this article ran on consecutive pages; do not give the volume and issue number, even if they are available.*

Fig. 2 Magazine articles - Williams College Libraries. Books. 9 December 2011. Available: http://library.williams.edu/citing/formats/books.php, accessed March 15, 2016

APA	**in text:** (Centers for Disease Control, 2009) **reference list:** Centers for Disease Control and Prevention. (2009). 2009 H1N1 Flu ("Swine Flu") and You. Retrieved from http://www.cdc.gov/h1n1flu/qa.htm
Chicago (Note)	**note:** 1. Norman R. Yetman, "An Introduction to the WPA Slave Narratives," *Born in Slavery: Slave Narratives from the Federal Writers' Project, 1936-1938*, last modified March 23, 2001, http://memory.loc.gov/ammem/snhtml/snintro00.html. **bibliography:** Yetman, Norman R. "An Introduction to the WPA Slave Narratives." *Born in Slavery: Slave Narratives from the Federal Writers' Project, 1936-1938*. Last modified March 23, 2001. http://memory.loc.gov/ammem/snhtml/snintro00.html. Also include a publication date or date of revision or modification (see 14.8); if no such date can be determined, include an access date (see 14.7).
Chicago (Author-Date)	**in text:** (Yetman) **reference list:** Yetman, Norman R. 2001. "An Introduction to the WPA Slave Narratives." *Born in Slavery: Slave Narratives from the Federal Writers' Project, 1936-1938*. http://memory.loc.gov/ammem/snhtml/snhome.html.
MLA	**in text:** (Pilgrim) **bibliography:** Pilgrim, David. "The Brute Caricature." *Jim Crow Museum of Racist Memorabilia*. Ferris State University. Nov. 2000. Web. 20 May 2009.

Fig. 3 Web page - Williams College Libraries. Books. 9 December 2011. Available: http://library.williams.edu/citing/formats/books.php, accessed March 15, 2016

4.5 Local citation system and editorial style

Because of the flexible ISO 690, which does not impose a determined style but essential recommendations (of elements, order and punctuation), it is possible an individual/ized use of citation system rules and bibliographic format requirements by the different authors and local editors.

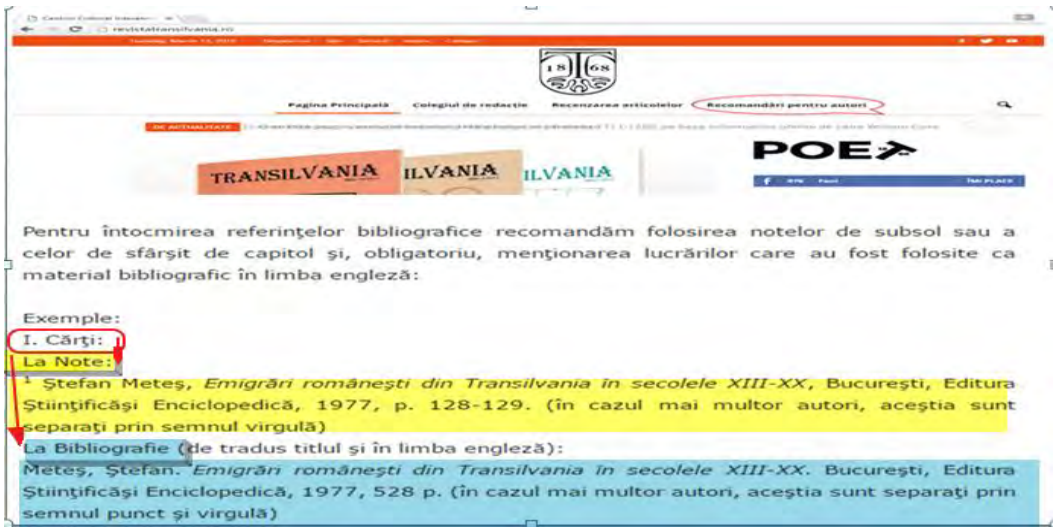

Fig. 4 Recommendations for authors from the Interethnic Cultural Center "Transilvania"

5. Use of Microsoft Office Word 2007 for generating the references

This informatics application allows the elaboration of references by using some steps:

- References:
  - Styles ▸ Insert Citation ▸ Add New Source ▸ Create Source
  - Bibliography ▸ Bilt-In: Bibliography /Works Cited ▸ Insert Biblography

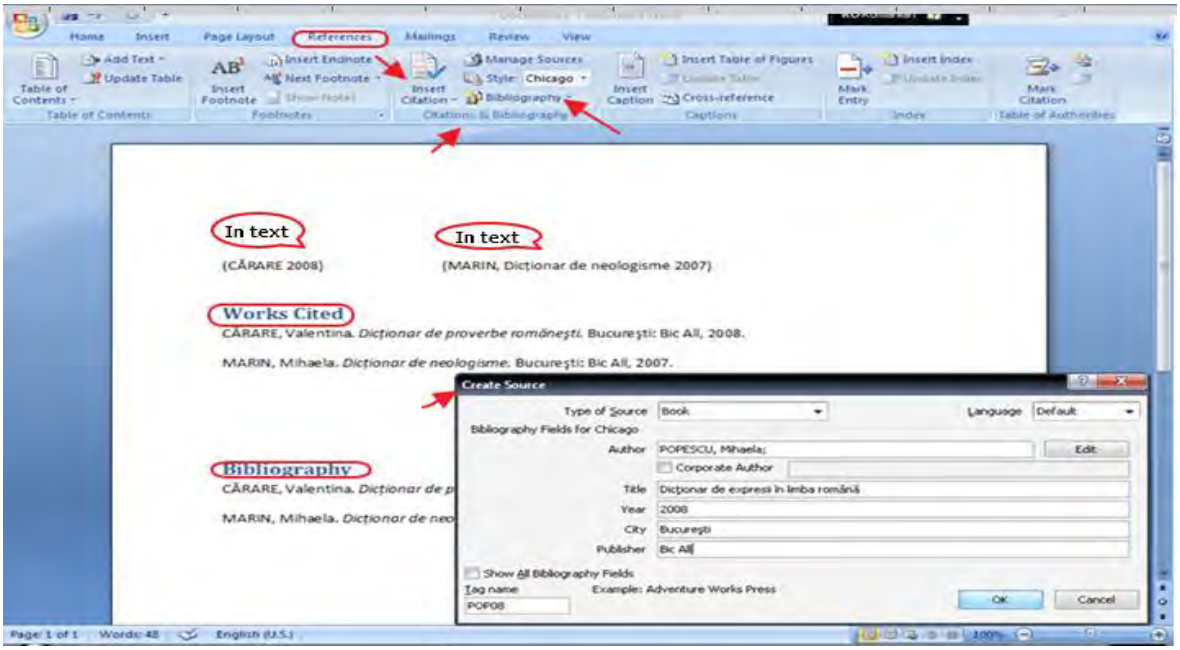

Fig. 5 Generating of references in Word.

6. Conclusions

1. It is necessary to distinguish between reference instruments in order to use them properly in scientific papers and in order to avoid the confusion caused by their similarity.

The definition of the related concepts establishes the differences, which individualize the terms and eliminate the confusions generated by their similarity, facilitating the appropriate use of these concepts.

2. Because of the flexibility of ISO 690 standard, which does not impose a determined style, we can talk of the following possibilities:

2.1. Adapting standardized requirements to the own system: citation – bibliographic format;

2.2. Taking one of the known editorial styles, depending of the specific research subject (example: for Natural and Social Sciences: author – date; for Human Sciences: note – bibliography);

2.3. Applying the recommendations for authors from a publisher (example: Recommendations for authors from the Interethnic Cultural Center "Transilvania“).

References

[1] Central Scientific Library (CSL) "Andrei Lupan“; Public Institution "Encyclopaedia of Moldova". *References: rules and citation styles: a practical guide*. Chisinau, 2010, available on the website: http://bsclupan.asm.md/img/userfiles/file/GHID_FINAL.pdf, accessed January 30, 2016

[2] Limerick Institute of Technology (LIT), Library and Information Resource Centre. *Write it right! A Guide to the Harvard ('Author- Date') referencing system*. Romanian translation coordinated by prof. Dr. Angela Repanovici, "Transilvania" University of Brasov, 2015 available: http://aspeckt.unitbv.ro/jspui/bitstream/123456789/1864/1/SCRIE%20CORECT-RIGHT% 20IT% 20RIGHT.pdf, accessed January 30, 2016

THE DIGITAL COLLECTIONS OF THE LBUS LIBRARY IN THE EUROPEANA CLOUD 2013-2016 PROJECT, AN OPENING TO SCIENTIFIC RESEARCH ACTIVITY. CASE STUDY: THE IMPACT ON CONSUMER OF INFORMATION

Rodica Volovici[1], Camelia Volosciuc[2], Elena Marginean[3], Ioan Visa[4]
[1234]"Central University Library, "Lucian Blaga" University of Sibiu, Romania

Abstract
This spring (2016) ends Europeana Cloud, a project of great importance for the development of the digital platform Europeana.eu (1). The LBUS Library is the only contributor from Romania, with three digital collections, which covered the topic of multicultural coexistence of the ethnics in Sibiu's historic past. The paper emphasizes our contribution, quantitative and especially qualitative through its cultural value and diversity of materials. The next step is to disseminate the most important aspects of the project, the structure, technical requirements, a description of the entire environment and the way of working in order to achieve objectives. The major orientation envisaged was supporting the scientific research, providing working- tools for user-consumer of information. As a case study, it was presented the questionnaire conducted in the LBUS on library users.

Keywords: Europeana Cloud, digital library, cloud platform, research tools.

1. About the project

Funded by the **European Union** and coordinated by the **Europeana Foundation**, the Europeana Cloud is the strategic project for the years 2013-2016. It aimed to create a new data infrastructure for cultural content in digital format, to be used by the Europeana and its partners. More, this shared data are opened both to the professional community of European institutions and to the users.

"*This infrastructure will exploit latest technological advances in the domain of cloud computing to provide new abilities for efficiently storing metadata and content, easily sharing cultural assets between institutions, improving abilities to access these assets and research them using innovative tools.*" [2]

The Digital Library **Europeana**, has continuously grown quantitatively and existing infrastructure has become insufficient and complicated, a conglomeration of various technologies and multiple standards with costs.

It needed a new infrastructure for maintaining and distributing the european cultural content.

In addition, they were required new tools to facilitate research and how to access this repository of culture, the primary consumers of digital content.

So, the project has the ambition to share data handling tools, to enrich, transform and preserve - which will allow some institutions that do not have these resources, to benefit from them and to enrich their data and services.

"***Europeana*** *has become a network, a community. This community shares its experience professional, scientific and technical, but also the motivation to get involved and to make widely accessible to european culture, thanks to digitization*".[2]

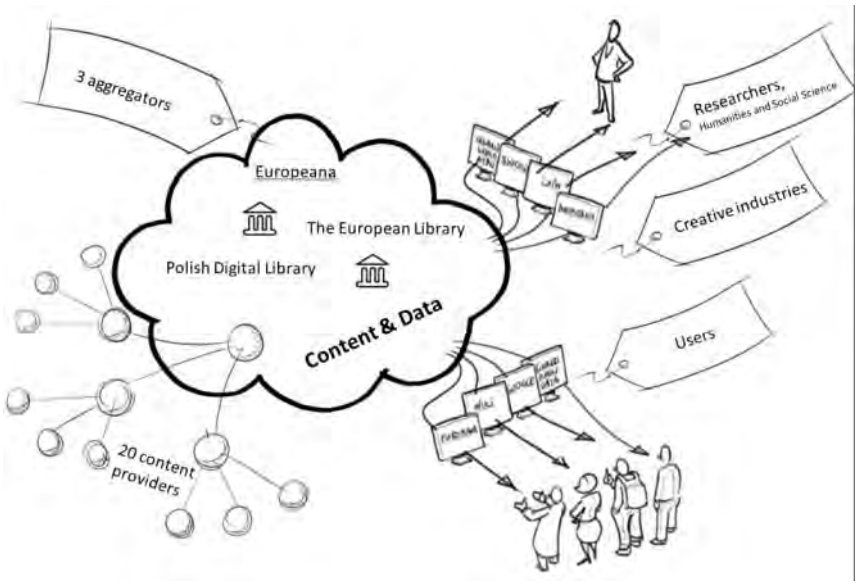

Fig.1 The Europeana Cloud environment

The project was focused precisely on these aspects:

- To follow the basic needs of data providers and researchers,
- To develop a new technical solution based on cloud technology,
- To develop useful research tools,
- Research economic and legal aspects related to efficiency and sustainability of the project.

Work packages in the project were aligned to these issues.

2. Project objectives:

- „contribution with new records, preparing and adding new data (2.4 million new metadata records and 5 million digital objects)
- creating a new workflow between content providers and aggregators for processing, storing data efficiently, plus provide access to digital cultural heritage
- orientation to researchers with new services and tools for access, work on and share the content stored in the Cloud
- allowing loading back into the cloud of new works from research and data enriched by scholars to be used by the institutions and researchers alike".[3]

The project was divided into seven main areas of interest, known as work packages:

- "WP1 - Assessing Researcher Needs in the Cloud and Ensuring Community Engagement
- WP2 - Developing the Infrastructure for Europeana Cloud
- WP3 - Exploiting Europeana Cloud with services and tools for researchers
- WP4 - Ingestion of Content and Metadata Development
- WP5 - Sustaining the Europeana Cloud: Legal, Strategic and Economic Issues
- WP6 - Dissemination and Networking
- WP7 - Project Management".[2]

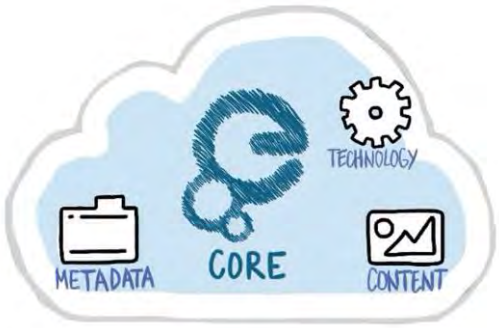

Fig.2. Europeana Cloud Architecture

3. LBUS Library participation in the project

Following the successful participation in the previous project "**Europeana Libraries 2011-2012**" (4) and joining the professional network "**The European Library**"(5), the Library of the LBUS was included as a partner - data provider.

The working team of specialized librarians, initially approached two digital collections of an overall theme, specific area of Sibiu, namely multicultural and multiethnic coexistence of residents, reflected in old books and periodicals.

The LBUS Digital Collections initially proposed for Europeana Cloud project, were:

1-***Brukenthal National Museum - calendars and journals***

(The oldest german calendars and journals)

2-***Sibiu - Historical and multicultural coexistence***

(The oldest romanian journals and books)

Later, were added two other sub-collections:

3-***Sibiu - Hermannstadt - Siebenbürgen, in chronicles and documents***

4-***Luceafărul - revistă literară*** (Revistă pentru literatură şi artă - tipărită in Sibiu)

12.	"Lucian Blaga" University of Sibiu (Romania)	**274** a1145 (187) a1146 (53) a1417 (34)	**274** 9200340 (187) 9200400 (53) 9200469 (34)		*Bust (1920): Mitropolitul Andrei Saguna (1808-1873), from the Biblioteca Facultatii de Teologie "Andrei Saguna" din Sibiu*

Fig.3. Initially proposed collection, caption from the Europeana Cloud text project (contain: name of contributor, collections number, items no, sample digital object)

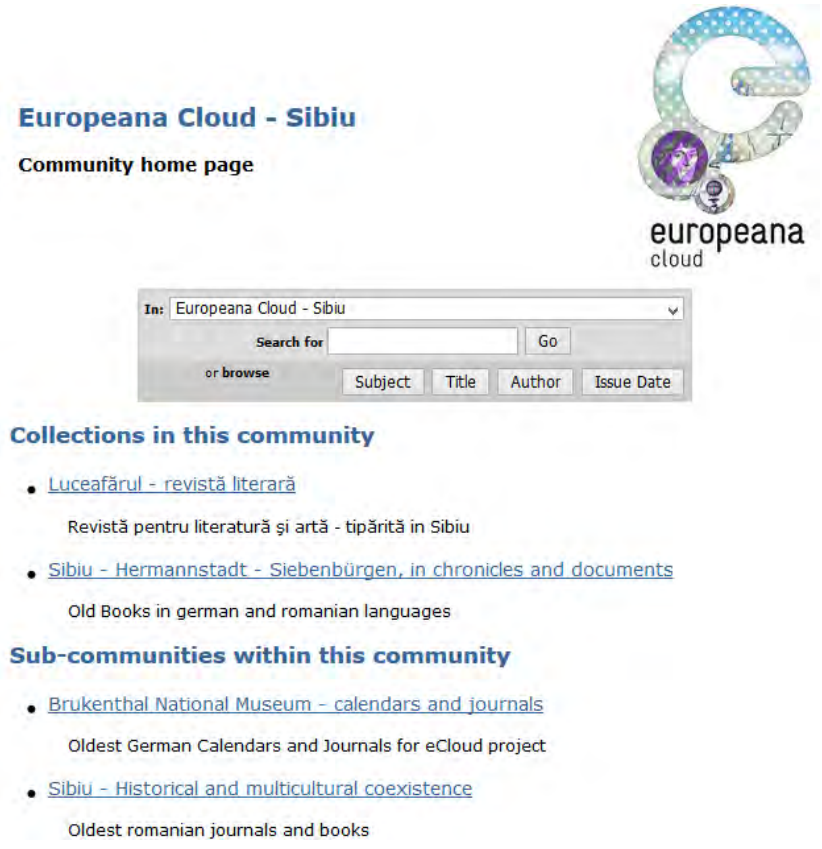

Fig. 4. LBUS Digital Library - collections in the local DSpace repository
http://digital-library.ulbsibiu.ro/dspace [6]

3.1 Digital collections, with titles, amount of pages scanned:

I. Collection 1: Brukenthal National Museum - calendars and journals (7)

- - Neuer Siebenbürgischer ProvinzialKalender, 1787 – 1871
- - Neu und alter Siebenbürgischer Provinzial Kalender, fur das Jahr Christi 1891
- - Der Siebenbürger Bote - (Nr.1, am 3, Januar 1841 - Nr. 97, am 2. December 1841)

Total = 4231 scanned pages

II. Collection 2: Sibiu - Historical and multicultural coexistence

Old books titles:

1. PUŞCARIU, Ilarion - Biserica Catedrală din Sibiiu. Sibiu: Tipografia Arhidiecezana, 1908
2. PUŞCARIU, Ilarion - Metropolia românilor ortodocşi din Ungaria şi Transilvania, Sibiiu 1900

3. XENOPOL, A.D. - Teoria lui Rosler, Iaşi 1884
4. LAURIANU, Treboniu - Istoria Romaniloru, Bucuresci 1862
5. Istoria biblica, pentru prunci, Sibiu 1858
6. BARITIU, Georgie - Istoria Transilvaniei, Sibiu, 1889
7. Mandinescu - Elemente de Istoria Universale
8. Zur Frage uber die Herkunft der Sachsen in Siebenburgen, 1856
9. Den Mitgliedern Schaessburg, Hermannstadt, 1867
10. unSächsisches Führertum, Hermannstadt, 1933
11. Episteln und Evangelia, Hermannstadt, 1824
12. Molitvelnik Sibiu, in Tipografia arhidiecezana, 1874
13. Puscariu, Ilarion, Dr. - Contributiuni istorice privitoare la trecutul Romanilor de pe pamantul craiesc, Sibiiu, Tiparul tipografiei arhidiecezane, 1913
14. Puscariu , Ilarion, Dr. - Documente pentru Limba si Istoria - Tom I , 1889
15. Puscariu , Ilarion, Dr. - Documente pentru Limba si Istoria - Tom II , 1897
16. Schriftliches Gebetbuch zum Gottesdienstichen Gebrauch der Evangelischen Gemeinden in Siebenburgen, Hermannstadt, Buchdruckerei v. Closius'sche Erbin, 1863

Total = 6309 scanned pages

III. Collection 3: Sibiu - Hermannstadt - Siebenbürgen, in chronicles and documents

Scanned pages:

- Old Book = 1043 pages
- Telegraful Român – years: 1868 – 1879, 1887, 1888 = 6450 pages
- Luceafarul 1908 + 1909 = 596 + 572 = 1168 pages
- Luceafarul 1853 - transliterat = 541 pages
- Vatra = 384 pages

Total = 9586 scanned pages

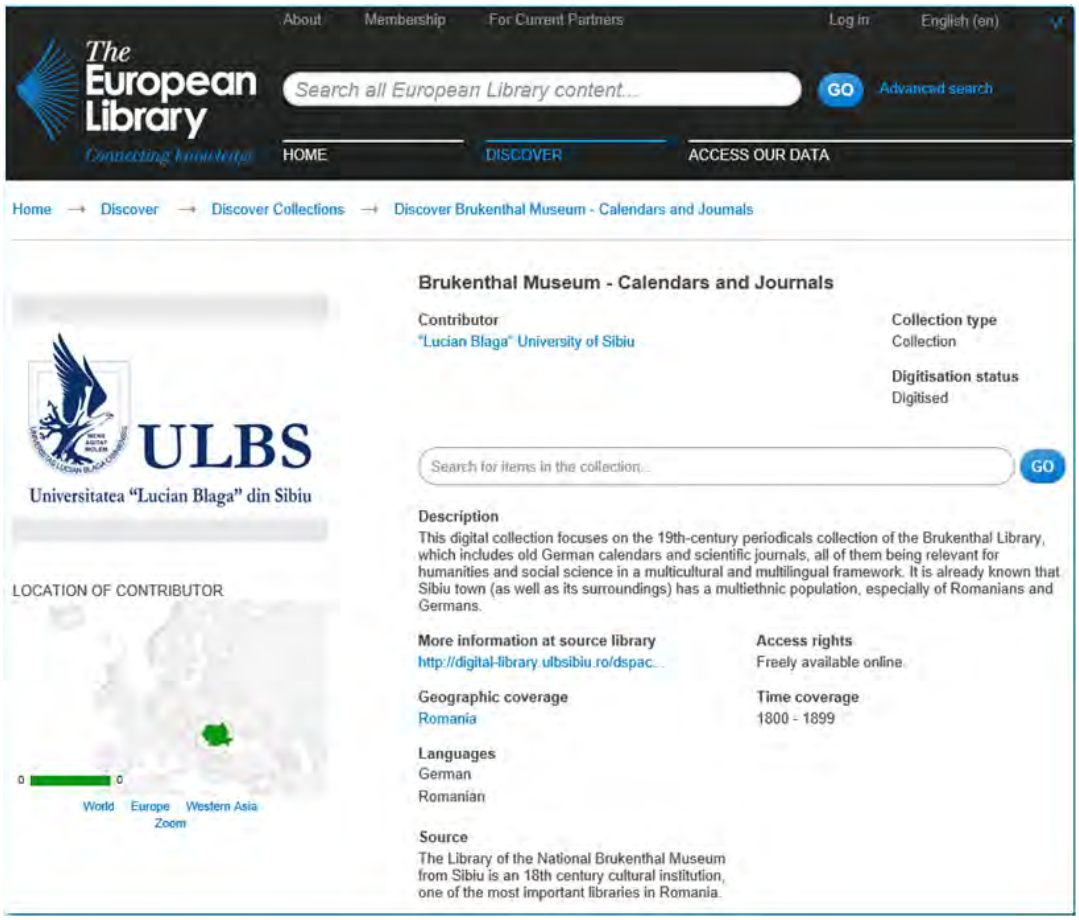

Fig. 5. ULBS page – the 1-st collection in the European Library – processing system [5]

Summary:

Collection1 = 4231 scanned pages

Collection2 = 6309 scanned pages

Collection3 = 9586

Total = *20.126 pages* / vs. Initial proposed = 10.000 pages.

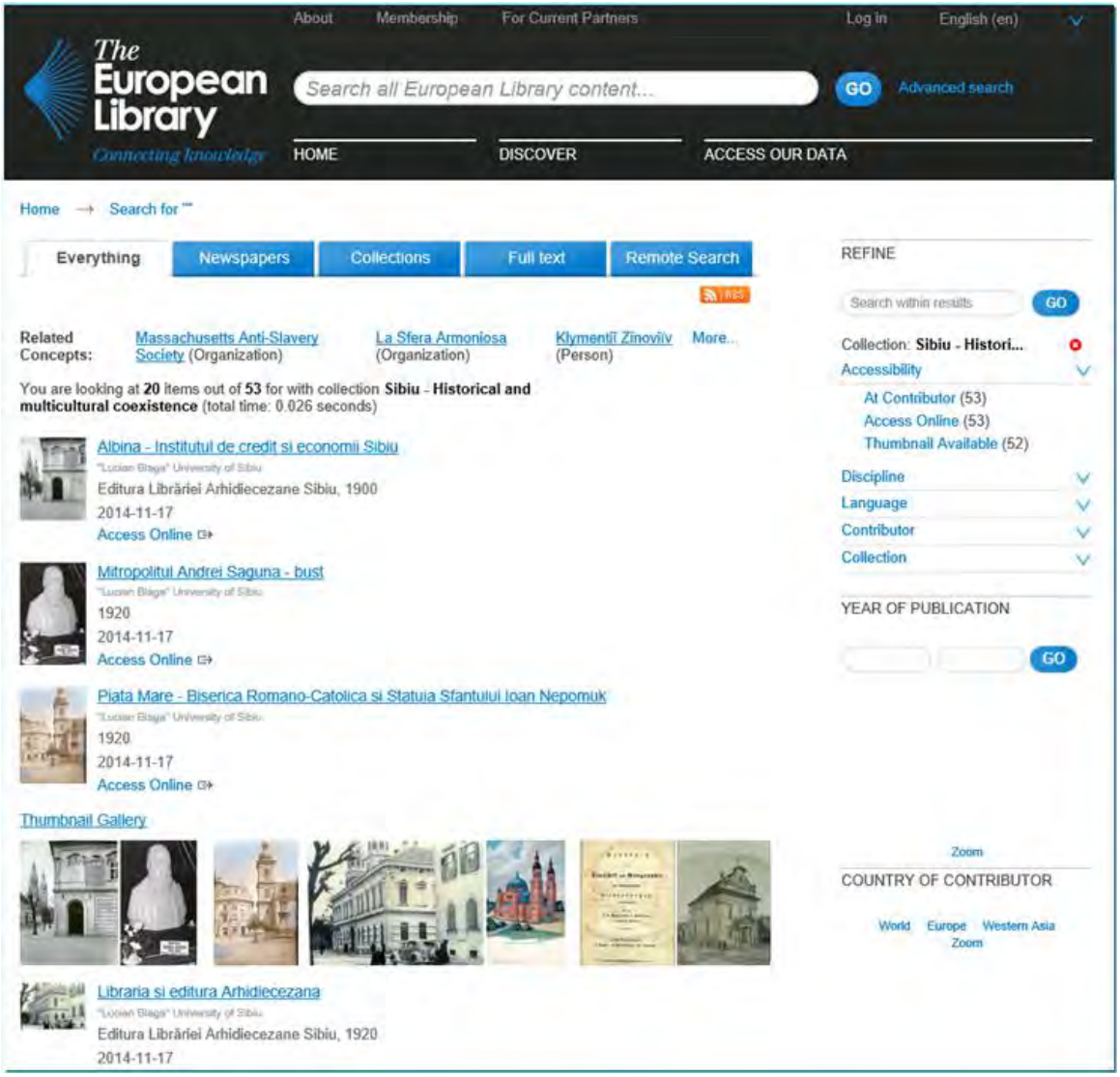

Fig. 6. LBUS – digital records in the European Library – processing system

3.2 Quiz about the User Profile, LBUS Library

The marketing Department from LBUS Library suggested a questionnaire for library users in order to know how they use the services, how easily find information on traditional or electronic support.

The subjects of this quiz are:

- accessibility to use library services
- communication traditional documents or of electronic documents
- other services (ex. Information retrieval services).

4. Conclusions

The great success of "Europeana" is not only managed to gather a mass of data, but because it gave a huge boost to the European cultural community and digitize heritage.

Countries and institutions that lacked a priority digitization, were held to obtain financing and to launch projects. Those who had already begun brought their collections, but also know-how and expertise. This effort was transversal, involving libraries, archives, museums, audiovisual and facilitated the emergence of concerns for interoperability collections.

"*A key strand of the Europeana Cloud project is the creation of services and tools targeted at researchers, collectively labelled Europeana Research, to allow scholarly use and re-use of the data held in our newly-built cloud.*"[3]

The LBUS Library is an active member of the European Library Network (TEL) and is the only romanian data provider in the Europeana Cloud project, with three digital collections in amount of *20.126 pages scanned* as digital objects.

Through this article we promote this european cultural treasure of Europeana digital library in the romanian academic area. Also, we invite users to search, process them critically and to use open data in their scientific papers.

5. Bibliography:

[1] Europeana portal: http://www.europeana.eu/portal/
[2] Europeana Cloud project: http://pro.europeana.eu/europeana-cloud
[3] http://pro.europeana.eu/get-involved/projects/project-list/europeana-cloud
[4] LBUS Library – Europeana Libraries project page - http://bcu.ulbsibiu.ro/europeana.html
[5] The European Library - Connecting knowledge, http://www.theeuropeanlibrary.org/tel4/
[6] LBUS Digital Library - the local DSpace repository, http://digital-library.ulbsibiu.ro/dspace
[7] http://digital-library.ulbsibiu.ro/dspace/handle/123456789/471

USING WEKA FRAMEWORK IN DOCUMENT CLASSIFICATION

Radu G. Creţulescu[1], Daniel I. Morariu[2], Macarie Breazu[3]

[123] *"Lucian Blaga" University of Sibiu, Engineering Faculty, Computer Science and Electrical and Electronics Engineering Department, Romania*

Abstract

Text document classification problem is a special case of a supervised data mining problem. In order to solve a text document classification problem some steps are required to fulfill. The common steps are: feature extraction, feature selection, classification, evaluation and visualization. The WEKA is a framework that helps us with all these steps. WEKA was initially developed as a library of java classes that help us to implement data mining applications. In the last years, in order to avoid java programming skills, the components from WEKA are also available into a visual form inside "WEKA Knowledge Flow Environment". We have studied and present in this paper some of the most important visual components that are available in the WEKA framework for the previously presented steps. These components are: "Arff Loader", "Attribute Selection", "Normalize", "Train Test Split Maker", a lot of classifier algorithms, "Performance Evaluator" and "Text Viewer". In order to prove the functionality of the visual framework in text document classification we have made and present some experiments. The most important advantage of the visual WEKA framework is the possibility to test different approaches without programming abilities..

Keywords: Document Classification, WEKA Framework, Visual Tools.

1. Introduction

Text document classification problem is a special case of a supervised data mining problem. In order to solve a text document classification problem some steps are required to fulfill. The common steps are: feature extraction, feature selection, classification, evaluation and visualization. The WEKA is a framework that helps us with all these steps. WEKA was initially developed as a library of java classes that help us to implement data mining applications. In the last years, in order to avoid java programming skills, the components from WEKA are also available into a visual form inside "WEKA Knowledge Flow Environment"[7].

2. WEKA framework

WEKA (Waikato Environment for Knowledge Analysis [7]) offers a framework that contains a collection of machine learning algorithms implemented for solving a lot of data mining problems. The algorithms are written in java and are available open source for integrate in your projects. But, for avoiding the programing part, WEKA offers also a framework that permits you to describe your data mining application as a flow of actions and to evaluated it, without the need to write code. If you want to use WEKA for analyze your specific data, you need to write some code in order to transform your dataset into a format that is accepted by WEKA.

WEKA offers four different options for realizing your data mining process. The WEKA ***Knowledge Explorer*** is an easy to use framework with a graphical user interface that offers all the facilities of WEKA package, grouped in some general steps as preprocessing, classification, clustering, attribute selection, etc. For an unexperienced user, that has only some knowledge in data mining, this framework helps you, because it forces you to consider all the steps that need to be used into a standard data mining process.

Another framework is Weka ***Experiment Environment*** that permits you to create, run and modify an experiment into a simple manner. The experiment can be described into a text file and tested into the WEKA framework. This is for a more experimented user.

WEKA ***KnowledgeFlow Environment*** permits you to describe your experiment as a flow of steps with some visual connections between them.

The last framework is the WEKA ***Workbench*** that contains a lot of state of the art data preprocessing and machine learning algorithms. In this framework the user can quickly try out existing machine learning methods on new datasets in a very flexible way.

In this paper, for feature selection, classification and evaluation steps, we have used the WEKA KnowledgeFlow Environment [1]. The project flowchart realized in WEKA is presented in Fig. 1.

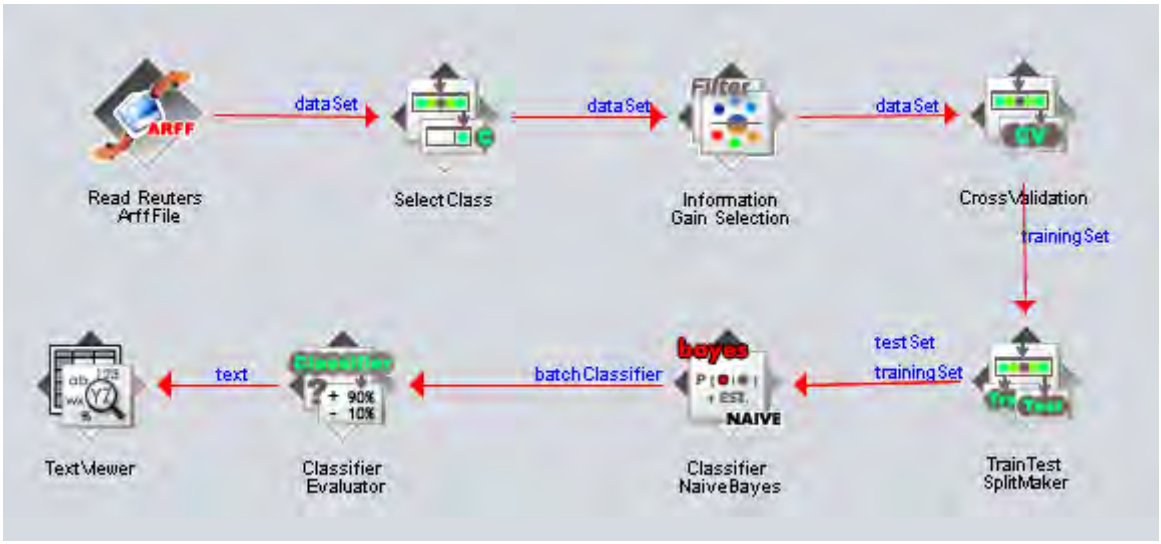

Figure 1 The WEKA project

In the following we present the components used in this project and a brief description of them as given by the WEKA framework help [8].

2.1 The ArffLoader Component

The first think that need to be done in each text mining project is to specify the dataset. For loading the dataset the WEKA framework has a lot of components as ArffLoader, DatabaseLoader, LibSVmLoader, XRFFLoader, etc. Those components can be found in the DataSource group and offer 12 different format input files. For our experiment we have chosen the arff format and we decided to use the Reuters dataset [6].

The Reuters dataset that is a collection of news published by Reuters agency into a XML format. All the preprocessing steps for transforming the dataset from plain text into an arff

format were done into a different java application. The arff format that must be given to WEKA contain a list with all attributes used into the dataset (defined with name and type) and after the "@data" directive a list of each sample (one on a line) with values for each attribute. We prefer that the last attribute is the class (if the document is contained into a specific class or not in).

The format for the arff file is:

```
@relation Reuters
@attribute 'A0' numeric
@attribute 'A1' numeric
….
@attribute 'A6998' numeric
@attribute 'A6999' numeric
@attribute 'class' {'yes','no'}
@data
2,2,1,1,1,…,0,0,0,0,0,no
0,0,1,0,1,…,0,2,0,1,0,yes
…
```

For this component, only the input file needs to be specified, as in Figure 2. The file needs to contain the entire dataset (both the training and the testing part).

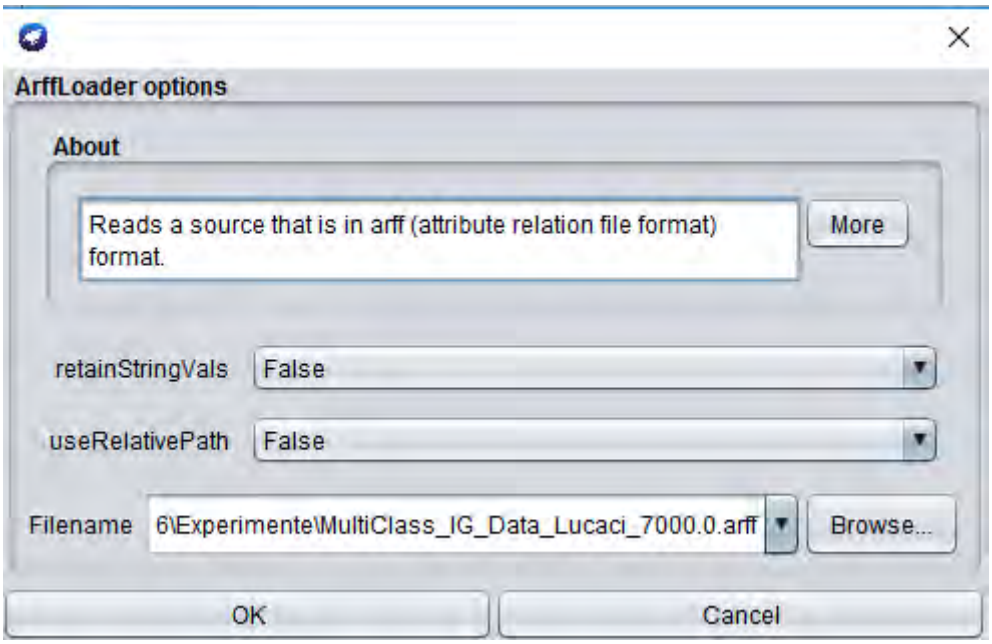

Figure 2. The ArffLoader Component

The two existing configurations are only for specify in what format the string arguments will be kept in memory (as string or not) and how the path for the input file is specified (useful when the experiment is moved in other part).

2.2 The ClassAssigner Component

After choosing the dataset, you have to use a component where to specify what attribute is considered as the desired class in the dataset (because we talk about a supervised learning)[5]. Such a component can be found into the "Evaluation" folder. The component is included in this folder because it is related to the evaluation algorithm and depends on how the evaluation will be done. The only configuration that needs to be done in this component is to select from a list the name of the attribute that is considered to be the class. In the presented experiments we have considered that our documents are into the class (and labeled with "yes") or not in the class (and labeled with "no"). Thus, we have considered a binary classification. The ArffLoader component permits you to specify a dataset with more classes, not only for binary classification.

In this framework, in order to establish what type of data will be transferred, a connection between components needs to be realized. Between ArffLoader and ClassAssigner components a "*dataSet*" connection should be used.

2.3 Attribute Selection Component

We decide to put this component in our experiment because we want to have the possibility to choose only a reduced number of attributes. We use the AttributeSelection component for selecting the best features. The component is in the Filters tab in the supervised area.

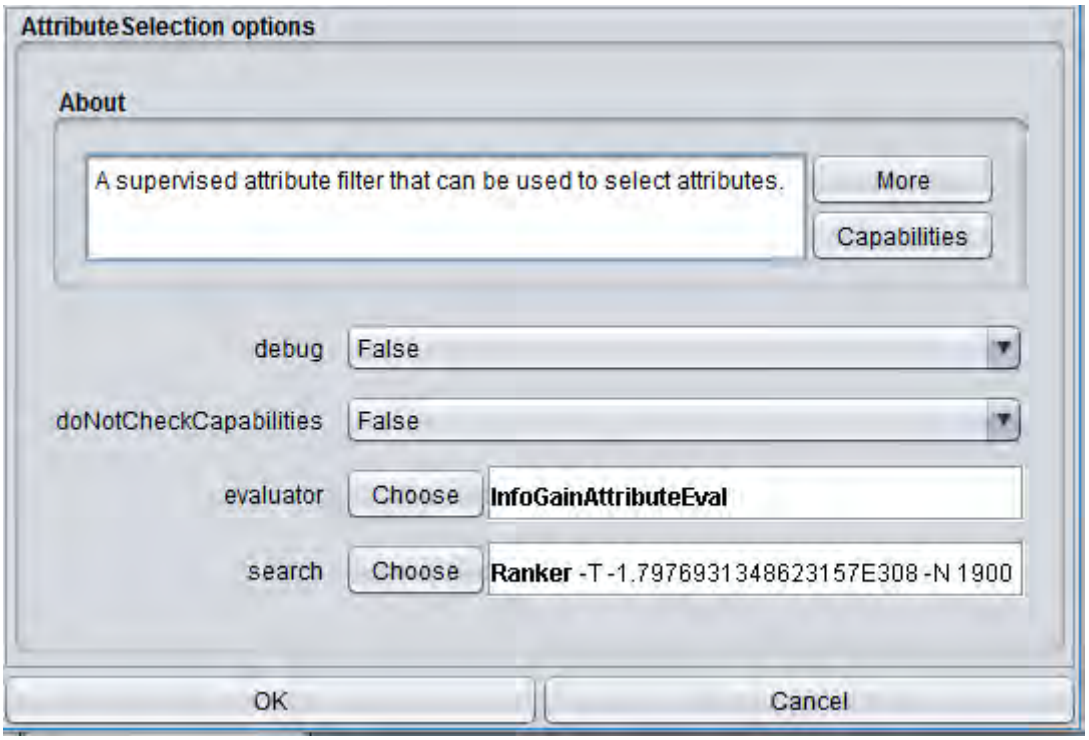

Figure 3. Attribute Selection Component

This component is a supervised attribute filter that can be used to select desired number of attributes. It is very flexible and allows various search and evaluation methods to be combined. In the evaluator property we can determine what method of attribute selection is used. The component supports attribute selection methods like: *InfoGainAttributeEval*, *GainRatioAttributeEval*, *OneRAttributeEval*, *PrincipalComponents* analyzer and more

others [2,3]. We have used the Information Gain Attribute Evaluation method. In the search property the method used in selecting the best attributes is specified (i.e. *Ranker*, *BestFirst* or *GreedyStepwise*). For the Ranker search method, that we have choose, the characteristic window as in figure 4 can to be configured. The Ranker component ranks attributes by their individual evaluations.

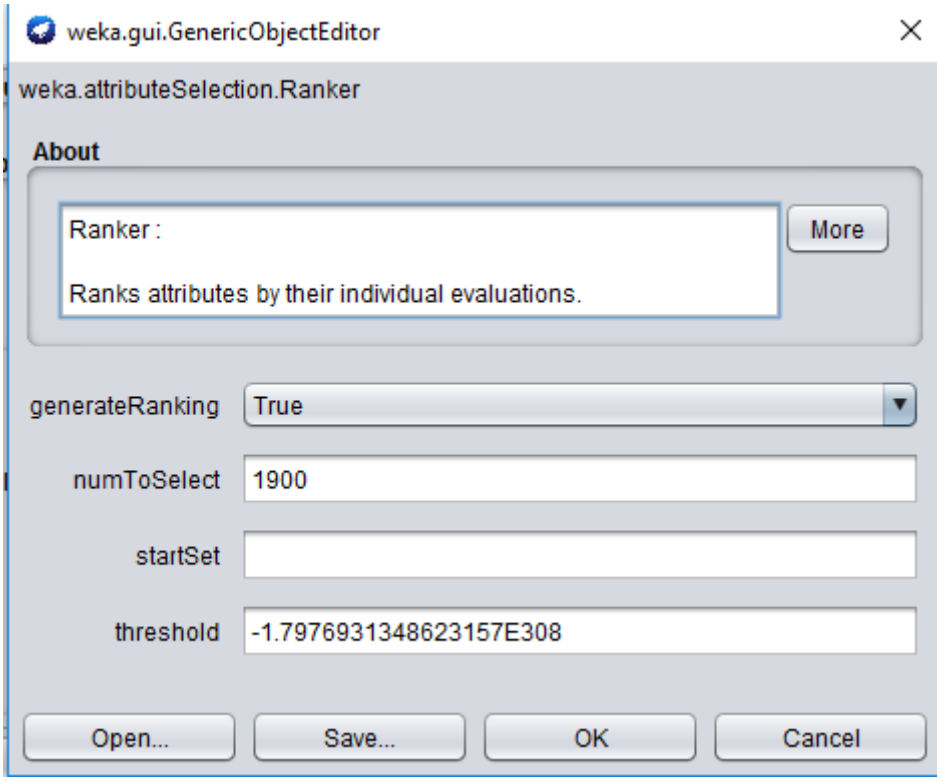

Figure 4. The Ranker properties

The *generateRanking* true is an imposed option. In the *numToSelect* field we can specify the number of attributes that are retained after selection. The default value (-1) indicates that all attributes are retained. In the field *startSet* we can specify a set of attributes that are overpassing the Ranker. When generating the ranking, Ranker will not evaluate the attributes that are in this list. In the last characteristic the *threshold* we can specify the threshold by which the attributes are discarded [2]. In our experiments we have used the default value for the threshold and have changed the numToSelect as desired in the range 200-5500 as presented in the experimental results section.

In the connections between the previous presented components we have selected as option to transfer form one component to another the *dataSet* option.

2.4 TrainTestSplitMarker

Until now we work with the entire dataset. Now we need a specific component that permits us to split randomly our dataset into a training part and a testing part. For this we can use *TrainTestSplitMaker* component that can be found in the Evaluation folder. This component is from *Evaluation* tab because it is also part of the evaluation process. This component permits at configuration to specify the percentage of training dataset and the random seed in the input file. This component generates two different outputs: one output

contains the training part form the dataset and the other contains the testing part. Therefor between this component and the next one we need to connect both outputs.

The Weka also permits to implement the CrossValidation idea, having a different component for this in the *Evaluation* tab, called *CrossValidationFoldMaker*. This component permits us to specify the number of experiments that are made with different splitting's of the dataset before showing the average result.

2.5 Classifier NaiveBayes

The WEKA framework contains a lot of learning algorithms, as classifier, clustering and association algorithms. The classifier algorithms have a specific tab with the *Classifier* name where a lot of algorithms from different categories (as *bayes*, *rules*, *trees*, *lazy* and more) can be found[3, 4]. WEKA has also a *Clusterers* tab with learning algorithms as *EM*, *Hierarchical*, *Simple KMeans* and more. For our experiments we use a classifier algorithm because we have a dataset that is already classified. We chose NaiveBayes classifier algorithm because it has a small number of characteristics that need to be specified.

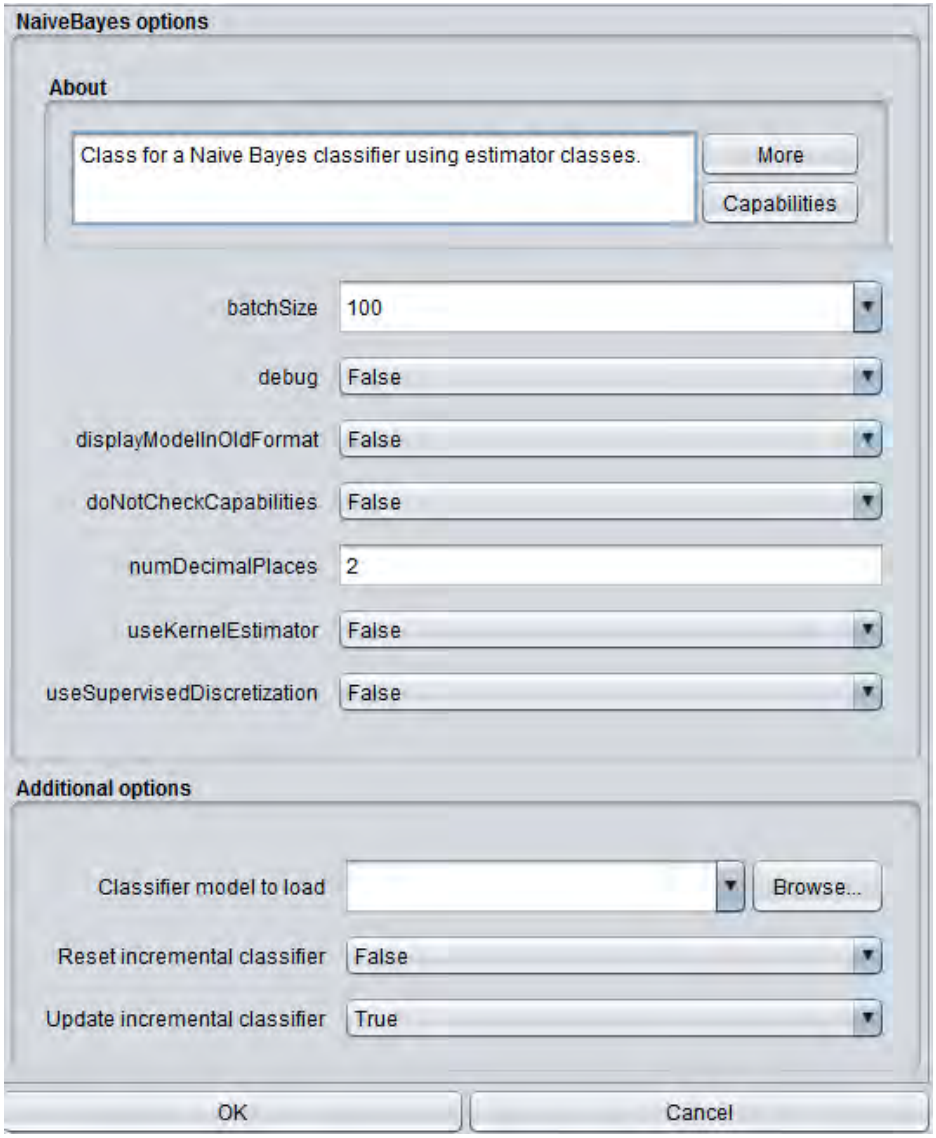

Figure 5. Naïve Bayes characteristics

This component is form *Classifiers* tab, in *bayes* region and is an implementation of the Naive Bayes classifier using class estimators. Precision of numeric estimators are computed based on the analysis of the training dataset. For all the parameters for this component (as *debug*, *displayModelInOldFormat*, *useKernelEstimator*, *useSupervisedDiscretization* and more) we have used the default values.

2.6 Classifier Evaluation

The previously presented component implements a learning algorithm. For evaluating the learning performance we need to use an evaluation component, from the *Evaluation* tab. We chose the *ClassifierPerformanceEvaluator,* that is designed to evaluate classifier algorithms. As configuration for this component we need to specify the evaluation metrics that has to be computed. The WEKA has a lot of evaluation metrics already implemented (as *accuracy*, *precision*, *recall*, *f-measure*, *TrueRate*, *NegativeRate* [9]). Default all 25 evaluation metrics already implemented in weka are selected. In our experiments we use only precision, recall accuracy and true rate.

2.7 TextViewer

For showing the results the WEKA propose a lot components grouped in the *Visualization* tab. Some components show the results graphically, while others write all the results in the text files. We chose for our application to write the results into a text file.

3. Conclusions

We have studied and present in this paper some of the most important visual components that are available in the WEKA framework for the previously presented steps. These components are: "Arff Loader", "Attribute Selection", "Normalize", "Train Test Split Maker", a lot of classifier algorithms, "Performance Evaluator" and "Text Viewer". In order to prove the functionality of the visual framework in text document classification we have made and present some experiments. The most important advantage of the visual WEKA framework is the possibility to test different approaches without programming abilities.

4. References

[1] [1] Ian H. Witten, Eibe F., Hall, M. A., Pal C.J., Data Mining – Practical Machine Learning Tools and Techniques with Java Implementation, Morgan Koufmann Press, 2000

[2] [2] Han, J., Kamber, M., - Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2001;

[3] [3] Manning, C., - An Introduction to Information Retrieval, Cambridge University Press, 2009;

[4] [4] Tom. M. Mitchell, Machine Learning, The McGrow-Hill Companies, 1997

[5] [5] Mitkov R., The Oxford Handbook of Computational Linguistics, Oxford University Press, 2005;

[6] [6] Misha Wolf and Charles Wicksteed – Reuters Corpus: http://trec.nist.gov/data/reuters/reuters.html, accessed in 03.2016
[7] [7] http://www.cs.waikato.ac.nz/ml/weka/, accessed in 03.2016
[8] [8] http://www.cs.waikato.ac.nz/ml/weka/documentation.html, accessed 03.2016
[9] [9] https://en.wikipedia.org/wiki/Precision_and_recall, accessed 03.2016

Evaluation of Classification in More Than Two Classes

Daniel Volovici

1.03.2016

Abstract

Machine Learning is the most important part of Artificial Intelligence in the same sense as we cannot speak about intelligence without the capacity of learning. One of the basics type of learning is to learn to classify objects or putting labels on objects. If you are able to recognize that an object have the attributes of a class C or not (meaning that it is part of class non C), than you will be able to classify in more than one classes: with the strategy one-vs-all or with the strategy one-vs-one. Classification as a learning task imply training with examples of objects a priori labeled with the class which they belong. But if in data we do not have definitions of classes, splitting data into groups has the name of clustering. The idea behind clustering is that probably the data are produced by different processes or that they belong naturally to different groups. So, the best way to evaluate the quality of the clustering is to try to cluster data generated to be part of different classes.

The most used way for evaluation of classification and clustering methods is the confusion matrix defined for two classes. Starting from this matrix it is obtained the measures of Precision, Recall and the Fmeasure. Exist a generalization to n classes using a nxn matrix. But for the situation where exist a different number of clusters than the number of original classes we must use a nxm contingency matrix also named association matrix. And because the degree of association is measured by the dominance of the principal diagonal it is very important to use time efficient methods of manipulation of the lines and columns of matrixes.

Keywords list: classification, clustering, contingency matrix, association, matrix, precision, recall.

1 Introduction

A method used to centralize experimental data is to write them into a table, a two-dimensional array. An important decision is to choose what will be describe on lines and what on columns. When a test is performed repeatedly

a number of times or on many individuals it is preferred that each individual to be represented on a line of the table and the results of measurements for that individual to be put on the cells of that line according with the column designated for every measure (for each attribute).

Table 1: Centralization of experimental data

	Attribute 1	Attribute 2	...	Attribute j	...	Attribute m
Individual 1	a_{11}	a_{12}	...	a_{1j}	...	a_{1m}
Individual 2	a_{21}	a_{22}	...	a_{2j}	...	a_{2m}
...		.....				
Individual i	a_{i1}	a_{i2}	...	a_{ij}	...	a_{im}
...		.....				
Individual n	a_{n1}	a_{n2}	...	a_{nj}	...	a_{nm}

An important influence on this type of tables came from medical tests for diagnosis of different diseases. The set of measurements is related with a collection of diagnosis tests designed to signal the presence or the absence of a desease. The actual paradigm is to apply the test to a number of subjects and, after that, to centralize these data counting the number of those with positive and of that with negative results. Decision about the result of the test (to be consider positive or negative) will be taken accordingly the values of attributes. If it is used only the value of a single attribute to decide if the test is positive or negative, then usually we calculate a threshold to be compared with the attribute's value. In the attribute's value is lower than the threshold we decide that the test is on one side, for example positive (or vice versa); if the value is higher we decide that the test results is on the other side (in our example, negative).

It is possible to reorder the rows of the table in descending (or ascending) order of the attributes values so that be easier to visualize the distribution of those with pozitive/negative test result (Fig. 1). Going on and centralizing even more we will obtain the exact number of those with positive test and those with negative result (Fig. 1).

The most important problem impacting in the field of medical diagnosis is the revelation that tests are never perfect and, because of this situation, it is possible to appear *false positive* and *false negative* results.

The same situation appear in all the fields where we decide to classify or to cluster an object to a group. When the characteristics of the groups are known before testing the procedure is consider **classification** and when we do not know a priori the groups and sometimes not even the number of them, we speak about **clustering**.

In the literature dedicated to epidemiology the data related to the results

	Attribute
Individual 1	a_1
Individual 2	a_2
...	...
Individual P-1	a_{P-1}
Individual P	a_P = Threshold
Individual P+1	a_{P+1}
...	...
Individual $P+N$	a_{P+N}

	Number of experimental results
Individuals tested positive	P
Individuals tested negative	N

Figure 1: Centralization of tests with a single attribute

of a test is usually [FFF14] represented in the form of a table as the one in Table 2. In this area researchers have access to results of the test and usually they do not know for sure if the humans being tested are really healthy or have the disease. They need a so called *gold standard* to find the truth. This is also denominated as *criterion standard* or *reference standard* and could imply expensive and sometimes dangerous additional testings. Because of the fact that we have access only to the experimental results of the test it is natural to represent the possible outcomes in rows and let the columns to be assigned to the estimation of the presence of the disease.

Table 2: The results of a test for diagnosing a desease

		DISEASE	
		Present	Absent
TEST	Positive	tp	fp
	Negative	fn	tn

In many other situation, researchers have access to the true values of the items analysed and in these situations it will be more natural another arrangement of the data in the table. We may think of transposing the matrix from Table 2. One of the most important is the case of simulations when we intend to test different methods on data known to be part of a class. In all these cases it will be useful the confusion matrix.

2 Confusion matrix

The performance of a learning algorithm is visualised using *confusion matrix* (also named *error matrix*) which is a table where each row represents the instances in an actual class and the columns represents the instances in a

predicted class according with the algorithm. This representation is natural in classification, but also must be considered for testing the performances of clustering methods: we could generate data so that every point to be a realization of a class and to observe how well the clustering algorithm group the data in clusters more or less similar with the real (true but unknown) starting generation process. In the community of statisticians working in clustering the confusion matrix has the name *contingency matrix*.

Table 3: Confusion matrix for binary classification (for 2 classes)

		Estimated Cluster Predicted Class		
		Examples estimated as *positive*	**Examples estimated as*negative***	
True Class (real examples to be observed)	Positive examples	***tp***	***fn***	the number of positive examples ***tp+fn***
	Negative examples	***fp***	***tn***	the number of negative examples ***tn+fp***

In the subfield of Machine Learning specialised in problems with classification (supervised and unsupervised) are very important two measures of quality: *Precision* and *Recall*. In Information Retrieval [Rij79] and especially in Text Retrieval [CM12] the evaluation measures have a meaning easy to understand:

- $-Recall$, proportion of all true members of class retrieved by the algorithm;
- $-Precision$, proportion true members of class from the number of those considered positive.

$$Precision = \frac{tp}{tp + fp} \tag{1}$$

$$Recall = \frac{tp}{tp + fn} \tag{2}$$

For unsupervised classification (clustering) the problem is a little bit different: exist two different classes and we assume they have the same importance. This is the reason to consider more important the other two evaluation measures: *Accuracy* (also named *Success Rate*) and *Error Rate*.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{3}$$

$$Error\,Rate = \frac{fp + fn}{tp + tn + fp + fn} \tag{4}$$

$$Accuracy = 1 - Error\,Rate$$

Accuracy is a measure [VBCM10] related with the association between the true sharing of examples in the two true classes on one side and the distribution of them in the two estimated clusters on the other side. And because statistics offers more powerful tools [Fle81] for estimating the degree of association [And73] we suggest to transform the problem of evaluation in one of association. For this purpose we will transform notations from Table 4 in those in Table 5. Starting from here it is possible to generalize to n classes like in Table 6.

3 Contingency matrix

The term of *contingency matrix* is used especially in statistics for representing in form of a tableau the frequency distribution of variables (very used for multivariate variables). It is also named *cross tabulation* it was introduced by Karl Pearson. In multivariate statistics it is vary important to discover dependencies between different variables. If some dependency exist we could determine a degree of association between variables.

In the present context, that of determination of the quality of clustering/classification, we are interested to measure the degree of association between real (true) classes and estimated clusters. So we will consider as variables the membership to classes/clusters. The membership will be consider a variable with m (number of classes) nominal values; on every line it will represent a class and on every column a cluster (or the estimated class). The contingency matrix could be consider also as an *association matrix* between real/true classes and clusters(estimated classes).

Table 4: The problem of splitting examples in one estimated class ***yes*** and ***no*** membership

		Estimated Class	
		yes	**no**
True Class	Positive examples	tp	fn
	Negative examples	fp	tn

In an ideal situation, the clustering method put examples in exact one correct class with no mistakes, no false positive and no false negative examples like in the example on Table 7. In this type of situation is no problem

Table 5: Transformation of the problem of one class into one with two classes

		Estimated Class	
		K_1	K_2
True Class	C_1	$a_{11} = tp$	$a_{12} = fn$
	C_2	$a_{21} = fp$	$a_{22} = tn$

Table 6: Generalization to n classes

		Estimated Class					
		K_1	K_2	$\cdots$	K_j	$\cdots$	K_n
True Class	C_1	a_{11}	a_{12}		a_{1j}		a_{1n}
	C_2	a_{21}	a_{22}		a_{2j}		a_{2n}
	$\cdots$						
	C_i	a_{i1}	a_{i2}		a_{ij}		a_{in}
	$\cdots$						
	C_n	a_{n1}	a_{n2}		a_{nj}		a_{nn}

to identify what cluster correspond with every class and to reorder lines and columns for obtaining a matrix with all nonzero elements on the principal diagonal.

Table 7: Example of an ideal clustering/classification

		Estimated Class (Cluster)			
		K_1	K_2	K_3	K_4
True Class	C_1	80	0	0	0
	C_2	0	50	0	0
	C_3	0	0	30	0
	C_4	0	0	0	20

For a possible situation more close to real situations like that from Table 8 we have some examles assigned to different other groups and we can consider them as false positive or false negative. It is important to be aware that are different types of false positives, one type for every class other than the true one; and different types of false negatives for every other cluster than the one assigned with the associated class.

Because we consider valid that association that maximize the *Accuracy* we want to maximize the sum of the cells corresponding to found associations. So we will rearrange the lines and the columns so that the sum of the cells on the principal diagonal to be maximum and we will obtain the correspondence: $C_1 - K_1$, $C_3 - K_2$, $C_2 - K_3$ and $C_4 - K_4$.

Table 8: Example of a not ideal clustering/classification

		Estimated Class (Cluster)			
		K_1	K_2	K_3	K_4
True Class	C_1	76	1	1	2
	C_2	0	0	30	0
	C_3	1	47	1	1
	C_4	3	2	0	15

4 Evaluation method

Table 9: A complex example of clustering

		Estimated Cluster			
		K_1	K_2	K_3	K_4
True Class	C_1	94	27	70	44
	C_2	69	56	10	4
	C_3	21	53	35	19
	C_4	0	33	1	3

Because in this more complex distribution of examples the great values in some of the cells are not very significant because it is possible to have many examples in one class and/or one cluster. To establish the importance of the value in one cell we could compare it with an uniform random distribution of the examples. The method used [WFH11] for this goal is to summarize the values on every line l_i and on every column n_j. The total number of examples is Sum.

$$l_i = \sum_{j=1}^{n} a_{ij}$$

$$n_j = \sum_{i=1}^{n} a_{ij}$$

$$Sum = \sum_{i=1}^{n} l_i = \sum_{j=1}^{n} n_j$$

Table 10: Working on the matrix

		Estimated Cluster				
		K_1	K_2	K_3	K_4	
True Class	C_1	94	27	70	44	$l_1 = 235$
	C_2	69	56	10	4	$l_2 = 139$
	C_3	21	53	35	19	$l_3 = 128$
	C_4	0	33	1	3	$l_4 = 37$
		$n_1 = 184$	$n_2 = 169$	$n_3 = 116$	$n_4 = 70$	$Sum = 539$

If all the examples were uniform random distributed according with the numbers belonging to classes and clusters a computed proportionally:

$$f_{ij} = \frac{l_i \cdot n_j}{Sum} \tag{5}$$

Table 11: Uniform random distribution of examples

		Estimated Cluster			
		K_1	K_2	K_3	K_4
True Class	C_1	80	74	51	31
	C_2	47	44	30	18
	C_3	44	40	28	17
	C_4	13	12	8	5

The majority of textbooks related with contingency matrix uses the square of the difference between the value a_{ij} and the uniform randomized value f_{ij} because they try to determine if it is true the hypothesis that exist an association between variables [JS11], [Fle81]. The squared values are used form obtaining χ^2 evaluation value or the *kappa* criterion. We use the differences normalized, but not squared, Δ_{ij} because we intend to find also the best possible association between classes and clusters according with the three criteria above mentioned.

$$\Delta_{ij} = \frac{a_{ij} - f_{ij}}{\sqrt{f_{ij}}} \tag{6}$$

For obtaining the most adequate assignment of clusters to true classes I propose three criteria:

1. each cluster is associated with a class and only with one;
2. assignment of the cluster with the class is better if the number of the related cell is greater;
3. maximizing the total sum of normalized differences Δ_{ij} on the cells of the association (rearranged on the principal diagonal of the matrix).

Table 12: Assignment of clusters to classes

		Estimated Cluster			
		K_1	K_2	K_3	K_4
True Class	C_1	1.53	-5.43	2.73	$2.44_{(3)}$
	C_2	$3.12_{(2)}$	1.88	-3.64	-3.3
	C_3	-3.43	2.03	$1.41_{(4)}$	0.58
	C_4	-3.55	$6.28_{(1)}$	-2.46	-0.82

5 Conclusions

In conclusion, we propose a method fos assigning the association (the correspondence) between the true class and the estimated cluster. Reordering the lines and the columns we obtain a matrix with the principal diagonal representing the values of guessed classes. Using that form of the matrix could calculate the *Acurracy* of the group. In the future we will try to make the procedure feasible in real time for great number of classes and to try to use for establishing the optimal number of clusters.

References

[And73] M. R. Anderberg. *Cluster Analysis for Applications*. Academic Press, 1973.

[CM12] Radu Creţulescu and Daniel Morariu. *Text Mining: tehnici de clasificare şi clustering al documentelor*. Editura Albastră, Cluj-Napoca, 2012.

[FFF14] Robert H. Fletcher, Suzanne W. Fletcher, and Grant S. Fletcher. *Clinical epidemiology: the essentials*. Wolters Kluwer/Lippincott Williams & Wilkins Health, 5th ed edition, 2014.

[Fle81] J.L. Fleiss. *Statistical Methods for Rates and Proportions*. Wiley Series in Probability and Statistics. Wiley, 1981.

[JS11] Nathalie Japkowicz and Mohak Shah. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press, New York, NY, USA, 2011.

[Rij79] C. J. Van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, Newton, MA, USA, 2nd edition, 1979.

[VBCM10] Daniel Volovici, Macarie Breazu, Gabriel Dacian Curea, and Daniel Ionel Morariu. Statistical methods for performance evaluation of web document classification. *Studies in Informatics and Control*, 19(2):169, 2010.

[WFH11] Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition, 2011.

International Journal of Advanced Statistics and IT&C
for
Economics and Life Sciences

VOLUME 6 Issue 2 CONTENTS 2016

ISSN 2067 – 354X