CRITERIA FOR A QUALITATIVE EVALUATION OF INFORMATION RESOURCES AVAILABLE ON THE INTERNET

Răzvan PETRE

Ministry of National Defense / Romania

Abstract:
The benefits of the Internet as information resource clearly outweigh its drawbacks, but an appropriate quality assessment method for data is necessary to draw valid conclusions. Applying evaluation criteria for assessing quality web resources is considered an appropriate method to ensure that electronic information be effectively selected. Different categorization methods are used to group evaluation criteria. Criteria are organized in broad items, such as scope, content, graphic and multimedia design, context, etc. Each of these items has its own specific indicators to be taken into account. This is confirmed by the literature on assessment of resources over the Internet, as well as by the various practices adopted all over the world in research and business environments.

Key words: Internet, search engine, quality assessment, information resources.

1. Introduction

The World Wide Web offers information and data from all over the world. Since so much information is available, and because of the fact that sometime information can appear to be "anonymous", it is required to build up skills to evaluate what you find. When someone uses an academic library, the books, journals and other resources have already been evaluated by publishers and librarians. Every resource you find has been evaluated in one way or another before you ever see it. Loads of information can be found online, but it's trickier to know what has been peer-reviewed online and what has not, as anyone can write a web page. Outstanding resources can reside beside the most doubtful.

2. The Internet

The **Internet** is the global system of interconnected computer networks that use the Internet protocol suite (TCP/IP) to link devices worldwide. It is a *network of networks* that consists of private, public, academic, business, and government networks of local to global scope, linked by a broad array of electronic, wireless, and optical networking technologies. The Internet carries an extensive range of information resources and services, such as the inter-linked hypertext documents and applications of the World Wide Web (WWW), electronic mail, telephony, and peer-to-peer networks for file sharing.

The linking of commercial networks and enterprises by the early 1990s marks the beginning of the transition to the modern Internet, and generated a sustained exponential growth as generations of institutional, personal, and mobile computers were connected to the network. Although the Internet was widely used by academia since the 1980s, the

commercialization incorporated its services and technologies into virtually every aspect of modern life

Internet use grew rapidly in the West from the mid-1990s and from the late 1990s in the developing world. In the two decades since then, Internet use has grown 100-times, measured for the period of one year, to over one third of the world population. Most traditional communications media, including telephony, radio, television, paper mail and newspapers are being reshaped or redefined by the Internet, giving birth to new services such as email, Internet telephony, Internet television, online music, digital newspapers, and video streaming websites. Newspaper, book, and other print publishing are adapting to website technology, or are reshaped into blogging, web feeds and online news aggregators. The Internet has enabled and accelerated new forms of personal interactions through instant messaging, Internet forums, and social networking [1].
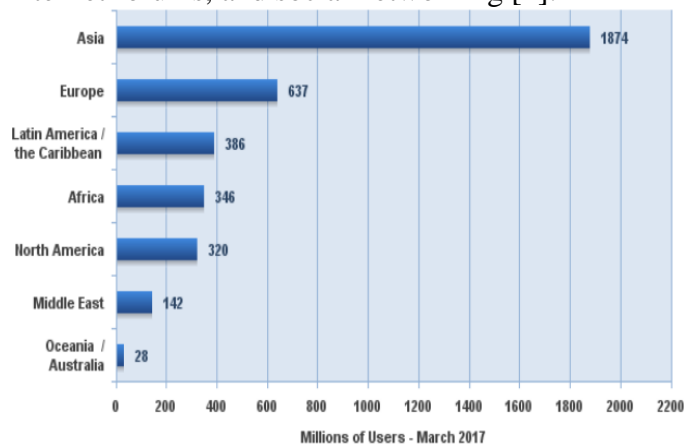

Fig. 1 - Internet users in the World by Geographic Regions – March 2017 [2]

2.1. Search engines

A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results often referred to as search engine results pages. The information may be a mix of web pages, images, and other types of files. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler.

The keyword search can be considered at least from 2 directions: finding the documents matching a user entered a keyword (or a phrase consisted of several keywords) or finding documents which are about the entered keyword. For the end user a search engine is a website with a text form, where he can type in his search keywords and receive a list of links to relevant pages or documents as a result. The result page also consists of descriptive textual information for each link and usually some relevant advertisement. Nowadays, search engines are often sophisticated software and hardware systems with closed and commercially protected inner structure. In short, each search engine includes robots (also known as crawlers or spiders) which travel through the Web following each link and mine documents to its database. Then data is extracted from the documents and processed using special algorithms. Each document and keyword are sorted and obtain some rating. In the end, a user optimally receives a list of web pages relevant to his query and sorted by relevance.

CRITERIA FOR A QUALITATIVE EVALUATION OF INFORMATION RESOURCES AVAILABLE ON THE INTERNET
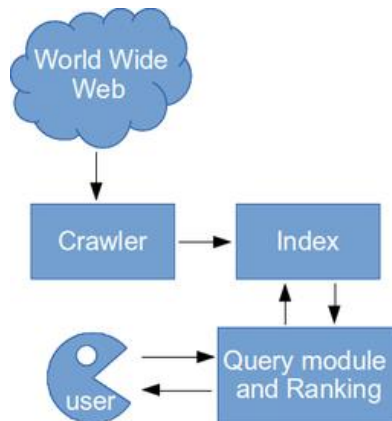

Fig.2 - High-level architecture of a Search Engine

Search (or document) indexing modules parse and store data about each document downloaded by the crawler. Search indexes are needed for increasing the speed of search and eliminating the need to low scan of each document in a database when a search query is received. Existence of the index increase the amount of necessary storage and consumption of computing resources (to fill in and update the index each time when new version of documents arrive), but it is a reasonable fee for the possibility of rapid search.

Query engine module is responsible for interaction with the end user – it receives a search query from a user, processes it addressing to the search index and gives back the results [1]. Modern search engines have a helper function called "auto-complete suggestions". When a user starts to enter a search query in a query form, a search engine will immediately extend already written word, phrase, or even a part of the world with relevant and frequent search queries, called "suggestions". That can dramatically increase overall usability and increase the speed of human-computer interactions.
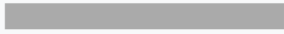
Ordering search results accordingly to relevance is another crucial task for every search engine. Huge variance of different factors can be considered for this purpose. All that metrics are stored in the search index, some of them are:

- number of each word occurrence in a document (called hits or term frequency);
- position of a particular word in a document, its font-size, surrounding tags;
- links to other pages and important information about these links (like text of the link) – it is very important for creating a model (or a map) of the World Wide Web, and to know precisely which page refers from and to.

Based on this information is calculated a quality ranking value for every indexed document, based on the link graph structure of the Web. The idea is that the importance of a particular web page can be measured in calculating the amount of other pages that link to it and their importance.

According to Net Market Share (as of March 2017) the global marketing share percentage, in terms of the use of Search Engines heavily favors Google, with over 80 %. This again reinforces that fact that Google are the market leaders; however it also highlights that the "Others" such as Yahoo, Bing and Baidu etc. still hold a large audience and it would be silly to simply ignore them. It's interesting to note that Google's large market share is still on the increase. Last year we saw a 68 % market share for Google, so Google has taken more than 10% of the market from its rivals in just the past 12 months.

CRITERIA FOR A QUALITATIVE EVALUATION OF INFORMATION RESOURCES AVAILABLE ON THE INTERNET

Search engine ⇔	Market share in March 2017 ▲	
Yahoo!	5.35%	■
Baidu	5.94%	■
Bing	6.92%	■
Google	80.52%	▬▬▬▬▬

Table 1 - The world's most popular search engines (with >1% market share) [3]

Although search engines are programmed to rank websites based on some combination of their popularity and relevancy, empirical studies indicate various political, economic, and social biases in the information they provide and the underlying assumptions about the technology. These biases can be a direct result of economic and commercial processes (e.g., companies that advertise with a search engine can become also more popular in its organic search results), and political processes (e.g., the removal of search results to comply with local laws). For example, Google will not surface certain neo-Nazi websites in France and Germany, where Holocaust denial is illegal.

Biases can also be a result of social processes, as search engine algorithms are frequently designed to exclude non-normative viewpoints in favor of more "popular" results. Indexing algorithms of major search engines skew towards coverage of U.S.-based sites, rather than websites from non-U.S. countries [4].

2.2. Advantages and weaknesses of the most popular source of information

In the information age few people can imagine the life without the Internet - a good and simple way of finding information. Most people are interested in reliable and direct ways to access, search and exchange information. The Internet as we all know is probably the most useful resource known to man in this day of age. While there is no arguing the positive impact that the Internet has had on society, there are certainly some drawbacks as well.

List of advantages of the Internet

1. Speed of information dissemination. The news and information industries have both seen massive changes in the speed at which they're able to share information with the general public thanks to the convenience and instant results offered by the Internet. This is particularly advantageous for instances that require the immediate dissemination of information such as during emergencies, health perils, safety threats, and other events that could lead to potential injuries and dangers.

2. Communication. One of the main goals of the Internet is to connect people. With social media, emails, chats, and other avenues you can instantly communicate and get to know people from all over the world. This has aided the world in beginning to bridge the culture gap that exists.

3. Unlimited access to information. Anything you could ever imagined can be found and learned about in detail on the Internet. There are millions upon millions of websites that are dedicated to all different types of topics. This wealth of information helps people to better understand things that they may have never thought or cared about before.

4. New economic methods are made. Another big thing the Internet has changed is business. It has become the biggest marketing tool, ever. People can also make purchases on the Internet, meaning that anyone can be a business owner.

5. It saves time and gets work done faster. Company owners can now transact business online and save time. With Internet connection, ordering to suppliers and communicating with clients and colleagues can be done through emails and fax.

6. Jobs and Business. People now also have the ability to search and apply for jobs completely through the Internet, which has opened up the world job market to many people with valuable skills to offer. More entrepreneurs can now open home-based businesses which can save them money because of reduced expenses

7. It is an effective platform for causes. There are a number of organizations who are fighting for a cause such as animal welfare, human rights and even a group of individuals who want to help the sick. By building websites and making use of the social media such as Facebook, Instagram and the like, millions of people are reached and touched.

8. It makes shopping and banking convenient for customers. Nowadays, the buying public can now shop for almost anything they want from the convenience of their homes, 24/7. Also, payments and bank transactions can now be done online, a convenience gladly embraced by more and more consumers every day.

<u>*List of weaknesses of the Internet*</u>

1. Exposure to false information. Because nearly every person on the face of the globe has access to the Internet, the likelihood of becoming the victim of misinformation is highly possible. There are very few laws and regulations that restrict the sharing of false information on the Internet.

2. A breeding ground for illegal activities. One of the unintended consequences of the Internet has been the immense amount of illegal and dangerous activity that it has harbored and perpetuated. The ease of information has given insight for burglaries, terrorist attacks, kidnappings, and many other types of crimes.

3. The world is becoming dependent. Most of our information is stored with the use of the Internet. This means that medical records, criminal records, educational records, and finance records are all completely dependent on the security of the sites that they reside on. Computer viruses are very common and are becoming more and more aggressive. They are transferred through the communication access of the Internet and if one would be designed well enough, it could shut down the economy.

4. Personal information could be compromised. People often put a wealth of their personal information on the Internet, without thinking of the consequences. Identity theft has dramatically risen since the Internet made its appearance in a mainstream setting.

5. Addiction. Social media, online games, and other time consuming things are beginning to take over the youth. They spend more time in their virtual life than they do in their real ones.

6. Too much dependency on the Internet can hamper business operations. Despite the convenience of smooth transactions from automation, using Internet and technology to run the business can also have a negative effect on the business once the computer systems break down. Operations will be put on hold if the system is down especially if the business utilizes the Internet to operate. Loss of Internet connection can slow down business processes and affect productivity.

3. Quality assessment for information resources evaluation

Quality of information resources is at the heart of the question of evaluation criteria. Both in the literature and in the present paper it has been engaged in terms of the criteria to assess it. The literature does not explicitly provide a definition of quality of web sites, but identifies evaluation criteria which serve to assess quality. In fact it appears that it is a difficult concept to grasp and describe, as shown by the focus groups' participants who, when explicitly asked for a definition, identified it with some indicators. Here some definitions are reported: "quality is synonymous of originality"; "quality means consistently organized and well structured resources"; "quality is a value related to content

CRITERIA FOR A QUALITATIVE EVALUATION OF INFORMATION RESOURCES AVAILABLE ON THE INTERNET

where data provided are not simply information, but knowledge"; quality is to be identified with fair reputation of authors and responsible providers".

Resource evaluation is mainly engaged in with reference to users, and recommendations on strategies and indicators to be followed are mostly addressed to researchers in their information selection process.

It is pointed out how credibility issues are related both to the content of the source and to the reader's purpose, so that personal and professional/educational use of information would imply a difference in evaluation of information. In general a variety of methods of criteria are to be considered, ranging from personal, informal methods to the more educational, formal techniques.

Different categorization methods are used to group evaluation criteria. Criteria are organized in broad items, such as scope, content, graphic and multimedia design, context, etc. Each of these items has its own specific indicators to be taken into account. Of the most relevant criteria from quality point of view are content and context.

For example, within the area of content, recommended features as relevant indicators are accuracy, authority, comprehensiveness, coverage, currency, objectivity, organization, substantiveness, uniqueness and validity [5].

3.1. Relevant quality criteria

Content quality criteria concern the evaluation of the information that the resource contains.

No.	Content	Definition
1.	Validity	It depends on how well researched and trustworthy the content of the resource is
2.	Currency	It refers to the timeliness of information and maintenance of updating
3.	Accuracy	It depends on how correct (more formal) the information is
4.	Authority	It depends on the expertise, reputation and status of the author/producer
5.	Uniqueness	It depends on the existence of primary material that is contained in the resource and it is not available from other sources
6.	Substantiveness	It depends on substantive quantity of information, that is not information linked to in the resource, rather substantial content provided on the site
7.	Coverage	It refers to the depth and extent of the information. The resource should cover the subject adequately and extensively
8.	Organization	It depends on the organizational structure of information and if it is arranged logically and consistently
9.	Comprehen-siveness	It depends on the way information is provided as regards completion of the site (no incomplete information, no "under construction" message)

CRITERIA FOR A QUALITATIVE EVALUATION OF INFORMATION RESOURCES AVAILABLE ON THE INTERNET

No.	Content	Definition
10.	Objectivity	It refers to the presence of factual data and the lack of personal prejudice in the information presented

Table 2 Content criteria and definitions

The *validity* of a resource, for example, depends on how well researched, well substantiated and trustworthy the content of the resource is.

Anyone can publish on the Internet, so information has often not been through traditional quality "filters" such as publishers, editors or peer reviewers. Resources may not be what they appear to be or what they say they are, as on the Internet there is no guarantee that the resource is accurate or honest. Possible question criteria of validity are: a) the information is evaluated by a well known publisher, editor or reviewer, either an organization or an individual; b) the resource is available in another format, such as a printed publication or CD-ROM; c) the resource is authentic: it is what it says it is, and is actually coming from the person mentioned as author/publisher; d) the original source of the information is stated, this resource is known to be authentic and authoritative; e) in case of doubt the information is verified by the subject specialist; f) the information is based on thorough research; g) references are provided; h) a substantial bibliography is provided; i) any bias or partiality is stated or is of an acceptable level (no propaganda).

Closely related to validity the *accuracy* of the resource depends on how correct the information is. The lack of information filters on the Internet like peer reviewers, publishers and editors, means that mistakes are more likely to occur than in printed publications. Those may be accidental errors, but also deliberate deception.

Possible question criteria of accuracy are: a) has the information been checked by an independent individual or organization (for instance a publisher, peer reviewer or editor); b) does the resource contain a bibliography / references; c) are grammar and spelling accurate; d) are there many typo's; e) in case of doubt the information should be verified.

The *authority* of the resource depends on the expertise, reputation and status of the author/producer.

On the Internet the origin of a resource is not always clearly stated and its attribution cannot always be trusted. Sometimes some detective work is needed. In case of doubt the information will be verified by the subject specialist.

Possible question criteria of authority are: a) the author/information provider of the resource is explicitly stated; b) the author/information provider is authoritative; c) the origin of the information is explicitly stated and is authoritative; d) contact details and/or e-mail addresses of author /information provider are given.

Uniqueness depends on primary information that is contained in the resource and is not available from other sources. On the Internet a lot of resources contain relatively little primary information, but consist mainly of links to other, external resources, available from other servers, or that reproduce information available elsewhere.

A resource that provides information that is nowhere else (online) available, is in most cases of more value than resources that contain exclusively secondary information, unless there is substantial added value, either in evaluation, annotation, presentation, metadata or comprehensiveness of the information.

Possible question criteria of uniqueness are: a) the resource contains original work; b) the resource contains at least some primary information; c) the resource consists of more than just a list of links to external servers; d) in case of links to external resources there is

CRITERIA FOR A QUALITATIVE EVALUATION OF INFORMATION RESOURCES AVAILABLE ON THE INTERNET

some added value in terms of evaluation; annotation, presentation, metadata or comprehensiveness of the information.

The *substantiveness* of the resources is the quantity of information that is actually provided on the site.

On the web, where information is hyperlinked, it is not always clear where one resource ends and another begins. Many resources contain very little real information, but link almost exclusively to information that is produced and stored elsewhere. The information that is linked to should not be considered as part of the resource itself. A resource should be evaluated on the basis of the information it contains and not on the basis of the information that is linked to.

Possible question criteria of substantiveness are: a) the information is of substantive quantity; b) the resource contains more than just address information and/or links to other resources.

Comprehensiveness means that the information is provided in a fairly complete form.

On the web information is often provided that is not yet completed ("under construction"), or that is only partly made available online and refers to printed or off line versions for the complete resource.

Although it is a characteristic of Internet information that it is regularly changed and updated, a website should meet some minimum requirements, in terms of comprehensiveness of the information that is made available.

Possible question criteria of comprehensiveness are: a) the resource contains more than very basic information; b) the resource is completed, and not under construction; c) the resource is available as full text, not only the bibliographical details or abstract; d) the resource offers useful abstracts or summaries of the information; e) there are no dead links or empty files; f) there is no missing information; g) the information that is actually available is in accordance with the index or description of the content of the site.

Coverage refers to depth and extent of the information. Possible question criteria of coverage are: a) the information has sufficient depth; b) the resource covers the subject adequately; c) there are no obvious omissions.

Context criteria apply to the origin of a site and its content, as well as to the suitability of a new resource to an existing collection. How a site meshes with and enhances existing or anticipated content within a collection will determine the quality of the collection over time.

Provenance - the origin or source of a site reveals and confirms much about its value and provides important information about its overall quality. Analyzing a site's URL often yields sufficient information to judge the affiliation of its creator and the reliability of its server.

Relationship to other resources. Like print resources, free Web resources can stand alone or in aggregate with other resources. As with any collection, however, the value of a collection of Web sites lies in the integrity of its individual components. The following principle applies: the greater the degree to which each site (component) within a collection relates to and enhances others within the same collection, the greater its value of the collection as a whole. Evaluations of free Web sites have shown that many of them contain links to the same sites. This redundant content will increase the size, but not the quality of the collection. Each site added to a collection must be viewed as an integral part of a larger mosaic. Redundant, superfluous, unrelated, or poorly suited pieces will not enhance the collection; they will only encumber it and ultimately discourage or confuse users.

3.2. Quality assessment process for data

An appropriate quality assessment method for data is necessary to draw valid conclusions. Determining the goals of data collection is the first step of the whole assessment process. The data users rationally choose the data to be used according to their strategic objectives or business requirements, such as operations, decision making, and planning. The data sources, types, volume, quality requirements, assessment criteria, and specifications as well as the expected goals need to be determined in advance.

In different business environments, the selection of data quality elements will differ. For example, for social media data, timeliness and accuracy are two important quality features. However, because it is difficult to directly judge accuracy, some additional information is needed to judge the raw data, and other data sources serve as supplements or evidence. Therefore, credibility has become an important quality dimension. However, due to the lack of uniform standards, data storage software and data formats vary widely. Thus, it is difficult to regard consistency as a quality dimension, and the needs of regarding timeliness and completeness as data quality dimensions are not high [6].

In order to do further quality assessment, is necessary to choose specific assessment indicators for every dimension. These require the data to comply with specific conditions or features. The formulation of assessment indicators also depends on the actual business environment.

Each quality dimension needs different measurement tools, techniques, and processes, which leads to differences in assessment times, costs, and human resources. In a clear understanding of the work required to assess each dimension, choosing those dimensions that meet the needs can well define a project's scope. The preliminary assessment results of data quality dimensions determine the baseline while the remaining assessment as a part of the business process is used for continuous detection and information improvement.

After the quality assessment preparation is completed, the process enters the data acquisition phase. There are many ways to collect data, including: data integration, search-download, web crawlers, agent methods, carrier monitors, etc. [7]. Nowadays data acquisition is relatively easy, but much of the data collected is not always good. We need to improve data quality as far as possible under these conditions without a large increase in acquisition cost.

Data sources are very wide and data structures are complex. The data received may have quality problems, such as data errors, missing information, inconsistencies, noise, etc. The purpose of data cleaning (data scrubbing) is to detect and remove errors and inconsistencies from data in order to improve their quality. Data cleaning can be divided into four patterns based on implementation methods and scopes: manual implementation, writing of special application programs, data cleaning unrelated to specific application fields, and solving the problem of a type of specific application domain. In these four approaches, the third has good practical value and can be applied successfully [8].

Then, the process enters into data quality assessment and monitoring phases. The core of data quality assessment is how to evaluate each dimension. The current method has two categories: qualitative and quantitative methods. The qualitative evaluation method is based on certain evaluation criteria and requirements, according to assessment purposes and user demands, from the perspective of qualitative analysis to describe and assess data resources. Qualitative analysis should be performed by subject experts or professionals. The quantitative method is a formal, objective, and systematic process in which numerical data are utilized to obtain information. Therefore, objectivity, generalizability, and

numbers are features often associated with this method, whose evaluation results are more intuitive and concrete.

After assessment, the data can be compared with the baseline for the data quality assessment established above. If the data quality accords with the baseline standard, a follow-up data analysis phase can be entered, and a data quality report will be generated. Otherwise, if the data quality fails to satisfy the baseline standard, it is necessary to acquire new data.
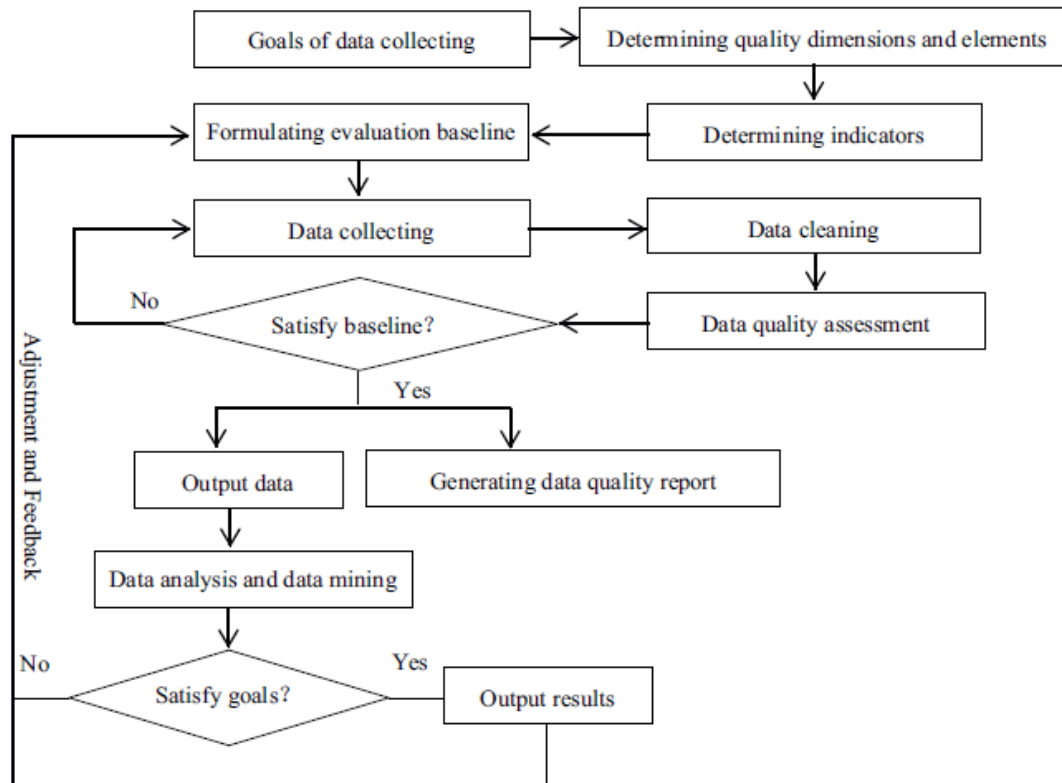

Fig. 3 - Quality assessment process for data

If the analysis results meet the goal, then the results are outputted and fed back to the quality assessment system so as to provide better support for the next round of assessment. If results do not reach the goal, the data quality assessment baseline may not be reasonable, and we need to adjust it in a timely fashion in order to obtain results in line with our goals.

3.3. Evaluating sources on the Internet for credibility

The development of the Internet and its delivery of online information have created a new information environment that can affect perceptions of the credibility of that information. Online information can be created by amateur sources (e.g., blogs and citizen journalism) posted on personal sites or on news websites (e.g., comments on news stories, CNN's iReport). Seeking to satiate the insatiable online audience with "instant" news can result in inaccuracies and premature conclusions. Online "polls" intended to allow the audience to express opinions are insufficient in that they are neither random nor representative samples. The freedom for anyone to publish anything and the convergence of information genres (i.e., infotainment) affect editorial review and credibility.

A study of the impact of the Internet on value orientations in mainland China by Jonathan Zhu and Zhou He (University of Hong Kong) revealed that credibility of the Internet was far more important than conventional media to Chinese citizens. This may be

because conventional media has a historical record of being credible whereas Internet sources, especially user-generated content, can be questionable [9].

Websites are often the first or only interaction consumers have with a company. Also, these interactions are usually short- in the magnitude of a few seconds to a few minutes. Source credibility can be established online in a number of ways to include logo design, website sophistication, and source citation. Lowry, Wilson and Haig (2013) showed that well designed logos that are synergistic with the company's product/service can trigger positive credibility judgments about the company's website and results in greater trust and willingness to interact with the company. These effects are magnified when the website design extends and complements the logo design [10]. Additional source credibility theory research has shown that color schemes and other visual elements can predict perceptions of credibility [11].

On social media sites, source credibility can be ascribed based on the number of followers and the ratio between followers and follows a user has. Research made by Westerman, Spence and Van Der Heide (2011) shows that there is a curvilinear effect for the number of followers, "such that having too many or too few connections results in lower judgments of expertise and trustworthiness. A narrow gap between followers and follows may also result in a higher perception of competence (i.e., if a user has many followers but does not follow many others, that person may be regarded as less of an expert). Further research showed that frequency of updates led to higher perception of credibility.

Klout.com is a social media analytic website and app that ranks users according to online social influence and assigns a "Klout Score". Research by Westerman et al. has shown that a higher "Klout Score" results in higher perceived competence and character but has no bearing on the caring (i.e., sociability) dimension of credibility.

Focus groups have also shown that user's evaluation criteria of blog credibility slightly differed from that of traditional media credibility. Participants indicated that the source of the blog was the first component evaluated and factors such as the source's knowledge, passion, transparency, reliability, and influence impacted credibility judgments. Participants also indicated that the blog's message/content needed to be authentic, insightful, informative, consistent, accurate, timely, popular, fair, and focused [12].

4. Information resources about research infrastructure in Romania – ERRIS case study

Platform *Engage in the Romanian Research Infrastructures System* (ERRIS) was created to support the coordinators of research infrastructures of public / private in Romania and those who want to benefit from the services offered by these facilities, fostering collaboration and networking with national and profile international scientific community in Romania.

Developed as a social networking platform ERRIS is a "facebook of things" where research infrastructures are "actors", having the opportunity to make known their services in a market as large and diverse.

ERRIS platform creation fulfills the necessary actions, internationalization of the national research, development and innovation, in line with the National Strategy for Research, Development and Innovation 2014-2020. ERRIS platform can be accessed at www.erris.gov.ro.

ERRIS is a platform developed and hosted by the Executive Agency for Higher Education, Research, Development and Innovation Funding (UEFISCDI), within the

framework of the "Improving the efficiency of the electronic data monitoring of R&D activities" European funded project.

ERRIS is the first Romanian online platform which connects the research infrastructure owners with potential clients (researchers and company representatives). The portal is a booking gate for services provided by the public and private Romanian research infrastructures. Its aim is to promote the infrastructures at the national and international level and to increase their services and equipments visibility. Moreover, ERRIS is a client oriented platform which responds to all users' needs.
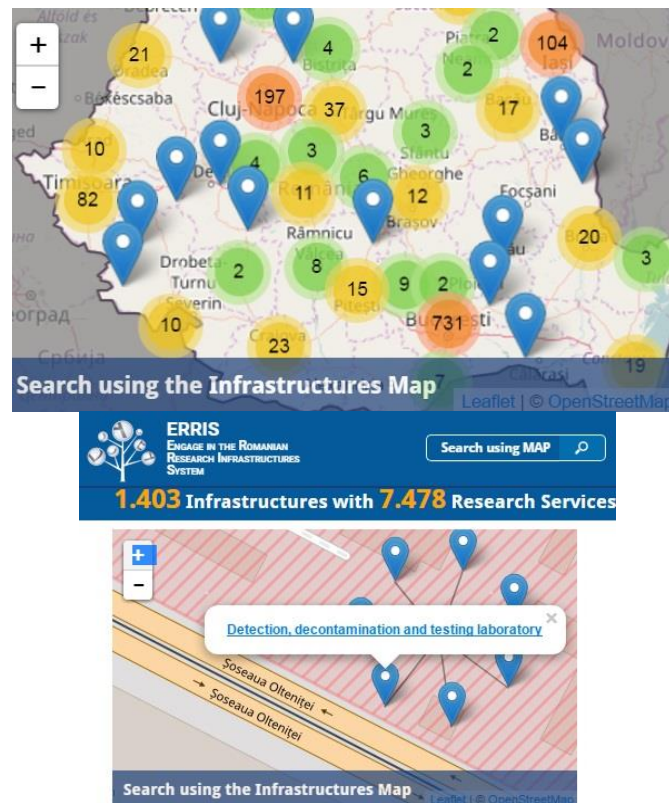

Fig. 4 – Search function of ERRIS platform

ERRIS is the first Romanian research infrastructure registry that brings together the majority of the public and private research infrastructures. That is way ERRIS provide a wealth of information to suite experiences of variety of users. Here one can find useful information about a profile of an institution and its research infrastructure. The services provided by the infrastructure are open for everyone interested in using them. All services are listed in such a manner to easily find specific areas of interest. That is way every infrastructure display a list of the members of the scientific and technical team. The rating of users is visible for every service which has been used. User's reviews represent a reputation system for the institutions, aiming not only to become a quality benchmark but also to stimulate continuous improvements of the services. The data base gives the users opportunity to find any Romanian infrastructure relevant to their specific interest. ERRIS represents a reliable and complex source of information with a very user friendly interface.

The platform facilitates easy and effortless access to services associated to the registered infrastructures. The reputation system provides the clients the opportunity to evaluate the research infrastructures, hence leading both to an informed choice for the interested beneficiaries, and to an efficient promotion of the infrastructures.

Regarding the search function of the platform, every keyword is searched within a specific item lists (institution, infrastructure, equipment, service, domains, persons). The result of

the Search function will be represented by those infrastructures which contain the respective keyword. In June 2017 the platform Search function found 1.406 Infrastructures with
7.491 Research Services, 61 Technological Services and 19.827 Equipments.

Fig. 5 – Available resources on ERRIS platform

The quality benchmark for the information presented on ERRIS platform it is given by user's reviews and their evaluation for the following criteria:
- ❖ The level of precision/ accuracy/ similarities between the services/ equipment which have been requested and what had been offered on site.
- ❖ The service accessibility regarding the administrative procedure.
- ❖ The services/ equipment condition and performance.
- ❖ The quality of communication with the staff.
- ❖ The desired services /equipment availability.

Fig. 6 –ERRIS platform quality benchmark

5. Conclusion

The benefits of the Internet clearly outweigh its drawbacks but it does not necessarily mean it is good for the majority. The key here is to have a closer look not just on the number of advantages it offers but also on the impact it has on people as a whole. It is indeed an effective tool but can also bring harm if used improperly.

The information resources available on Internet present explosive growth. How to ensure data quality and how to analyze and extract information and knowledge hidden

CRITERIA FOR A QUALITATIVE EVALUATION OF INFORMATION RESOURCES AVAILABLE ON THE INTERNET

behind the data become major issues for industry and academia. Poor data quality will lead to low data utilization efficiency and even bring serious decision-making mistakes.

I analyzed the Internet as the source of information, along with its advantages and weaknesses. I have presented how a search engine works and its high-level architecture. Likewise, I have presented a data quality assessment for information resource evaluation. The effective and most commonly chosen criteria for the evaluation of information resources available on the Internet are those relating to content, such as accuracy, authority, comprehensiveness, coverage, currency, objectivity and validity. That reflects their role, who is mainly concerned on substantial content to be found for research and professional activities. This is in line with what is emphasized in the literature and confirms the appropriateness and relevance of well established criteria which are commonly adopted in the print world.

Nowadays applying evaluation criteria for assessing quality web resources is considered an appropriate method to ensure that electronic information be effectively selected. This is confirmed by the literature on assessment of resources over the Internet, as well as by the various practices adopted all over the world in research and business environments.

Due to the rapid and increasing provision of resources with ever changing features and to the growing number of people and organisations putting material on the Internet, here will always be problems that will require further analysis.

The experience of the last years shows us that there is a strong need for criteria to be used in the evaluation of web-based information resources, and that such criteria may vary in relation to the discipline and domain to which these resources belong, as well as to their format. Moreover, difficulties are encountered in interpreting and applying criteria, even if guidelines and recommendations have been provided in various settings.

References:

[1] Ian Peter, *"So, who really did invent the Internet?",* The Internet History Project, 2014.

[2] http://www.Internetworldstats.com/stats.htm.

[3] https://www.netmarketshare.com.

[4] Vaughan Liwen, Mike Thelwall, *Search engine coverage bias: evidence and possible causes.* Information Processing & Management, 2004, 40 (4), pp. 693–707.

[5] Ginevra Peruginelli, *Evaluation criteria of legal electronic resources. A case study of Italian legal web resources' selection.* "Informatica e diritto", 2006, Vol. XV, n. 2, pp. 211-243.

[6] Shankaranarayanan G., Ziad M., Wang, R. Y. *Preliminary Study on Data Quality Assessment for Socialized Media,* China Science and Technology Resources, 2012, 44(2), pp. 72–79.

[7] Zhu Y. Y., Xiong Y., *Datology and Data Science*, Shanghai: Fudan University Press, 2009

[8] Wang Y. F., Zhang C. Z., Zhang B. B., *A Survey of Data Cleaning.* New Technology of Library and Information Service, 2007, 12, pp. 50–56.

[9] Zhu Jonathan, He Zhou, *Information Accessibility, User Sophistication, and Source Credibility: The Impact of the Internet on Value Orientations in Mainland China,* Journal of Computer-Mediated Communication, 2002, 7 (2).

[10] Lowry Paul Benjamin, Wilson David, Haig William, *A Picture is Worth a Thousand Words: Source Credibility Theory Applied to Logo and Website Design for Heightened Credibility and Consumer Trust,* International Journal of Human-Computer Interaction, 2013, 30 (1), pp. 63–93.

CRITERIA FOR A QUALITATIVE EVALUATION OF INFORMATION RESOURCES AVAILABLE ON THE INTERNET

[11] Gianluigi Guido, Prete Irene, Sammarco Luigi, *Prototypes Versus Examples: A New Model of Online Credibility for Commercial Websites*. Journal of Targeting, Measurement & Analysis for Marketing, 2010, 18 (1), pp. 3–16.

[12] Edwards Chad, Spence Patric, Gentile Christina, Edwards America, *How Much Klout Do You Have... A Test of System Generated Cues on Source Credibility*. Computers in Human Behavior, 2013, 29: A12–A16.