INFORMATION QUALITY EVALUATION METHODOLOGY FOR WEB PAGE DIRECTORIES

MARCIN KRZESAJ

Faculty of Economics, Opole University

The article presents information quality evaluation methodology for websites directories. In the first part of the article the theoretical basis for website directories was presented (ways of information acquisition, classification and issues relating to quality of websites). The second part of the article was devoted to the research methodology for evaluating quality of information collected in website directories. The identification of quality characteristics carried out shall allow for evaluating quality of general information and of information in thematic directories on web pages. The article ends with a summary containing conclusions and guidelines for further research directions.

Keywords: information quality, quality in web page directories, information quality evaluation

1. Introduction

The importance of information in the modern world is growing rapidly. Information has become the next fundamental economic resource and production factor beside land, labor, capital and entrepreneurship. The level of the economic development is not mainly determined by amount of financial resources and condition of fixed assets but by information, which has characteristics of good quality such as, for example: topicality, completeness and credibility.

Information systems store very rich information resources. In particular, it applies to the information resources on the Web. The Internet along with the World

Wide Web is currently used as a global information system [1]. The information collected on the Internet is used for different purposes, e.g. consumption, taking specific decisions or increasing own information resources.

The ex ante evaluation of usefulness and quality of information regardless of the place of use is very difficult. The issue of the information quality is in the focus of interest of many researchers and has become a subject of numerous scientific publications [2, 3, 4, 5, 6, 7]. Most of these studies is focused on making overall evaluation of functionality of Internet services treated as a certain system, i.e. on evaluation of the quality of operations of such systems. This kind of research covers the studies on usability of websites developed by W. Chmielarz [8, 9, 10, 11]. However, there are very few publications relating to studies on quality of information on websites itself.

It is in many information systems operating on the Web (e.g. social networks, websites containing users' feedback on products, etc.) that there is lack of any control over quality of information posted and published there. This is one of the causes of the information chaos on the Internet and exponential increase of the share of the information of poor or very poor quality. Similar problems are faced by web page directories. As in other information systems, their frameworks and information quality models need to be developed. This argumentation is justified for many reasons. There is a need to work on the quality of the information on the Internet. Therefore, different classifications of quality dimensions and criteria need to be developed and then measures for individual quality criteria have to be designed. The resulting tools for evaluation of information quality should have characteristics of generality (possibility of adaptation to different environments), applicability (possibility of use by different entities; for example by an information administrator or a user of information) and flexibility (possibility of adaptation to the needs of users of different models).

The aim of this article is to propose a methodology for evaluating quality of information in the websites directories. The theoretical basis for website directories was presented in the first part of this article (ways of information acquisition, classification and issues relating to quality of websites). The second part of the article was devoted to the research methodology for evaluating quality of information collected in website directories. The identification of quality characteristics carried out shall allow for evaluating quality of general information and of information in thematic directories on web pages. The article ends with a summary containing conclusions and guidelines for further research directions.

2. Acquisition of information for directories on web pages

As it was noted by J. Unold, acquisition of information on the Web is conducted in two main ways: browsing and searching [1]. The first one is characterized by passivity and the second one by activity of the process of searching for information. Finding information requires quite more knowledge on the effective use of the available tools than the intuitive process of browsing.

One of the most common approaches to classification of the tools for finding information divide them into two categories: directories and search engines [12]. Other approach divides most of the tools for finding information on the Web into three categories: general directories, specialized directories and search engines [1]. Apart from them, there are specialized and hybrid tools, e.g. meta search engines, web crawlers, vertical portals.

Directories provide a contextual and structured content on the Internet. While search engines enable searching by given words and key phrases, but without context. According to J. Unold, browsing is more effective in the case of directories of web pages, while searching in the case of entering words or key phrases to search engines. The author mentioned above also draws attention to the one of the fundamental problems in searching for information on the Internet, which is the dichotomy. "On the one hand, indexes are very efficient in the search process, but they are not able to provide an automatic decision-making process based on proper evaluation of the information obtained. On the other hand, an individual user is admittedly able to take a correct decision but there is no physical possibility to analyze the information contained on millions of Web sites provided by a search engine" [1]. One of the possible solutions of this problem seems to be enhancement of the quality of the results obtained in the popular tools for acquisition of information.

It is in the literature that classifications of web page directories by their subject matter (e.g. general, specialist), by the way of conducting (e.g., moderated, unmoderated, paid, free of charge), or by configuration (e.g. different methods of registration of web pages) can be found. While analyzing available website directories, the following typology can be proposed:
- General directories (e.g. Onet.pl, Start24.pl).
- Thematic directories: industry, specialist, company directories (e.g. dirbud.pl).
- SEO directories (e.g. zbiorkatalogow.pl).

Web page directories occupy the leading place in the process of acquiring information on the Internet. A web page directory contains a set of links to web pages, sorted by topic into categories by the editorial staff of the service. The most common directories have a tree structure containing a list of the main categories. It is in the categories that web pages are described which are subjected to

evaluation carried out by the editorial staff of the given directories. Individual thematic categories have lower hierarchy levels narrowing the scope of search.

3. The research methodology for evaluation of quality of information

The aim of the study was to establish a set of criteria helping to evaluate quality of information in directories of web pages. The execution of the task required tests of cognitive nature. The analysis of the literature concerning the research methodology for evaluation of information quality allowed for defining the research problem and proposing the appropriate set of criteria for the examined issue. The issue of information quality evaluation covers several consecutive stages:

1. Conceptualization of information quality.
2. Defining the criteria for information quality.
3. Defining of quality measures for the given information objects.
4. Measurement of information quality.

Ref. 1. While defining the examined issue of information quality, one should refer to its definition in the literature. R. Y. Wang and D. M. Strong believe that quality of information is a multidimensional category and an analysis of its characteristics and criteria should be carried out in the widest possible context [13]. Different authors propose different definitions of the information quality [1, 2, 15]. The proposed definitions correspond to different types of quality: technical, perfect, recipient-oriented and manufacturer-oriented one. However, it is emphasized in the literature that there is no universal and general understanding of quality of information as well as constituent characteristics forming it.

For the needs of this study, the following definition by J. Ruževičius and A. Gedminaitė was adopted, which is the broadest definition [14]: "Quality of information is a sum of properties and dimensions of the information which allows for meeting or exceeding expectations and requirements as well as expressed and undisclosed needs of a user - a knowledge worker". This definition of the quality of information has an interdisciplinary character and allows for defining it in the broadest context.

The conceptualization of the issue being examined was started with a discussion of the specifics of web page directories (general and thematic ones) in terms of their advantages and disadvantages. It became a starting point to determine problems concerning quality of information and allowed for the widest possible coverage of the issue. Table 1 shows advantages and disadvantages of web page directories.

Table 1. Advantages and disadvantages of general web page directories

Advantages	Disadvantages
Selectivity	Limited database size Relatively small content
Categorization (hierarchical structure)	Secret and sometimes unclear editorial policy
Contextuality of browsing	Difficulties by keeping topics up-to-date
Built-in information search tool	Uneven level of studies
Relevance for general questions	Payment for placement may impair quality of the content
Ease of browsing	Crawls only the main page of the website
	Lack of transparency concerning the privacy policy and protection of personal data
	Differentiated quality and consistency

Source: own study based on: J. Unold, *Teoretyczno-metodologiczne podstawy przetwarzania informacji w cyberprzestrzeni*, UE Wrocław, Wrocław 2011, pp. 249-250

Description of the information quality characteristics identified on the basis of advantages and disadvantages of web page directories:

- **Completeness/sufficiency**: the limited size of the database both for general and thematic web page directories results in inability to provide all the information on a specific topic. The web page directories accept only unique web addresses (without subpages), which may cause acquisition of too little information.
- **Topicality**: non-compliance of the information with the actual state of web pages indexed in directories (updating procedures are executed too rarely).
- **Consistency**: descriptions of web pages do not always allow for their use.
- **Reliability:** description of the content of the web page in the directory depends on the editors of the website, which may affect their uneven level. It happens very often that less popular web page directories do not contain information on the purpose of the website (these can be SEO directories) or on sources of its financing (sponsored web directories).
- **Reliability**: no information on qualifications of the persons evaluating directories.
- **Clarity:** publication policy contained in the regulations often does not contain explicit language and precisely defined terms.
- **Information security:** lack of clarity as to the privacy policy and the protection of personal data.
- **Structural compliance**: web page directories contain thematically sorted collections of links to web pages located in a structured way, which facilitates contextual browsing of the content.

- **Accuracy:** specialization and proximity of the content placed in the given category of the directory.
- **Usability:** as a search result of web pages in a directory with use of an integrated search tool, a user receives a title and description prepared by an editor and his subjective evaluation of its content

The advantages and disadvantages of directories presented above highlight the specificity of the examined problem and allow for identification of the information quality characteristics. Most of the presented characteristics is of a qualitative nature relating directly or indirectly to the information quality being evaluated.

Ref. 2. Defining the criteria for information quality.

There have been various attempts to systematize quality criteria for information in the literature. B. K. Kahn, D. M. Strong, R.Y. Wang define 16 characteristics, which they assign to four dimensions of information quality: integrity, reliability, usefulness and applicability [15]. Another attempt to organize the issue of quality was taken by L. English, who distinguished two dimensions relating to primal and pragmatic characteristics of information quality [2]. A different systematics was proposed by L. Floridi [3]. He grouped twenty-seven information quality criteria in four categories of information quality characteristics. While discussing different approaches to determination of information quality criteria several Polish authors, including B. Stefanowicz [16], J. Kisielnicki [5], M. Niedźwiedziński [6], should be mentioned. The most important information quality characteristics include: relevancy, accuracy, timeliness, completeness, consistency, form suitability, availability, clarity, credibility, communicativeness, reliability, flexibility, usability, redundancy, complexity, naturalness, semantic coherence, structural compliance, security, verifiability, variability and reputation [17]. The examples of the taxonomy for evaluation of information quality mentioned above derive from different premises; therefore, it is difficult to select a set of criteria for information quality evaluation in web page directories. Most of the studies on evaluation of Internet services is concentrated on evaluation of performance of such services, that is on their usefulness. While publications relating to studies on information quality on websites itself are very rare (e.g. evaluation of information in libraries, evaluation of quality of medical information).

Due to the interdisciplinary nature of evaluation and the broad context, a set of general methodologies for examining quality of medical information on the Internet was used to determine the criteria. The combination of common criteria for evaluating quality of information was based on Tools/methods:
- Evaluation criteria contained in EU Directives,
- The HON Code of Conduct,
- Silberg's Method,
- Hogne Sandvik Method
- HITI Method.

As a result of the analysis of the selected methods for evaluation of information quality, a set of ten criteria was created: references/competencies, disclosure of information, timeliness, attribution, availability, content transparency, privacy policy and data protection, advertising policy, interactivity and technical criteria (navigation, links, proper operation) [18].

An attempt was made to assign criteria for evaluating information quality to the identified characteristics of information quality. It is in the table 2 that there is a list of the identified characteristics of information and the criteria for information quality evaluation assigned to them.

Table 2. The list of the identified characteristics of information and the criteria for information quality evaluation assigned to them

Order No.	A recognized information quality characteristic	Evaluation criteria	Group of criteria
1	Reliability	References/ Competencies	Qualitative
2	Reliability	Information disclosure	
3	Topicality	Timeliness of information	
4	Reliability	Attribution	
5	Completeness/Sufficiency	Availability	
6	Clarity	Content transparency	
7	Safety	Privacy Policy and data protection	
8	Safety/Credibility	Advertising policy	
9	Structural compliance	Technical criteria (navigation, links, correct operation)	non -qualitative
10	Usability	Interactivity	

It is on the list that there are two groups of criteria visible. The first one refers to evaluation of the characteristics directly affecting the quality of information. These criteria include: references/competencies, information disclosure, information timeliness, attribution, availability, content transparency, privacy policy, data protection and advertising policy. The second group of criteria is used to evaluate characteristics of non-qualitative nature (interactivity, technical criteria) and plays a supporting role by evaluating information quality. Improvement of the information quality in web page directories can be affected by internal search engines and personalization of search phrases. The identified characteristics of information quality, accuracy and consistency were not reflected in the criteria, which might result from the specificity of the examined objects. The next stage of the evaluation is determination of measures for the presented criteria.

Ref. 3. It is in the literature on the subject that the following tools are suggested for determination of measures for given objects [17]:

- − Functions and operators (e.g. minimum and maximum operators)
- − Synthetic indicators (e.g. a weighted arithmetic, geometric or harmonic mean of partial indicators),
- − Point method, GVP method,
- − Surveys (e.g. DISCERN questionnaire).

It was for evaluation of information quality that a questionnaire survey was proposed. The use of surveys to evaluate information quality criteria is quite popular and relatively easy to perform. A tool in a form of a questionnaire survey containing thirteen research questions was created (table 3).

Table 3. The list of criteria for information quality evaluation and research questions assigned to them

Criteria	Research Questions
References/ Competencies	1. Were references/qualifications of the reviewer/owner of the given web page directory included?
Information disclosure	2. Was the information about the owner of the web page directory published? 3. Was the purpose of the web page directory clearly defined? 4. Was the recipient of the content of the service clearly defined? 5. Was the information about the source of financing of the web page directory published?
Timeliness of information	6. Are there data about update of information in the web page directory?
Attribution	7. Was the source of the information and the date of its publication posted?
Availability	8. Does the web page directory contain meta information about content of the directory?
Content transparency	9. Was a clearly defined publication policy published?
Privacy Policy and data protection	10. Was any information on the privacy policy and the protection of personal data published?
Advertising policy	11. Was the information about policy regarding advertising published?
Interactivity	12. Has the web page directory got a built-in search tool?
Technical criteria (navigation, links, proper operation)	13. Are the collections of links to web pages located in a structured way?

While using a survey questionnaire some problems appear, which are related to reaching and obtaining answers from an appropriate group of respondents e.g. users, which are able to evaluate individual criteria.

It was for the evaluation of the information quality that an information quality evaluation indicator was suggested. It is on the basis of the collected answers (*Yes* or *No* for 13 questions asked) that an indicator of information quality in web page directories will be created. The value of the indicator is within the range of 0 up to 13. The zero value of the indicator means that all the questions were answered with *No*. All *Yes* answers set the maximum value of the indicator equal to 13. It is in order to determine the importance of the criteria used for information quality in web page directories that appropriate weights have to be applied. This issue requires further in-depth studies.

4. Summary

- It is in the article that a methodology for evaluating information quality in web page directories was proposed.
- A survey questionnaire consisting of thirteen questions was proposed for evaluating information quality.
- The created tool is addressed mainly to users of web page directories.
- One of the limitations of the presented tool is the subjective nature of such an evaluation due to the possibility of different interpretations of the questions contained in the survey by various respondents.
- The next stage of the research will be the evaluation of the information quality in selected web page directories (general and thematic ones).
- While summing up, it should be remembered that there is a field specificity by evaluating information quality.

REFERENCES

[1] Unold J. (2011) *Teoretyczno-metodologiczne podstawy przetwarzania informacji w cyberprzestrzeni*, UE Wrocław, s. 165.

[2] English L. (1999) *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*, John Wiley & Sons Inc., New York.

[3] Floridi L. *Information Ethics: On the Philosophical Foundations of Computer Ethics*, http://www.wolfson.ox.ac.uk/~floridi/ie.htm.

[4] Stvilia B., Gasser L., Twidale M. B., Smith L. C. (2007) *A Framework for Information Quality Assessment*, Journal of the American Society for Information Science and Technology, pp. 1723-1724.

[5] Kisielnicki J. (1982) *Metody badania zapotrzebowania na informacje*, w: *Informacyjne problemy planowania*, PWE, Warszawa.

[6] Niedźwiedziński M. (1987) *Cechy informacji – próba systematyzacji*, w: *Jakość danych w systemach informacyjnych*, Oleński J. [ed.]: Systemy informatyczne nr 1, Wydawnictwo OBSR, Warszawa, pp. 360-370.

[7] Stefanowicz B. (2004) *Informacja*, Wyd. SGH, Warszawa.

[8] Chmielarz W. (2008) *Metodyczne problemy oceny witryn bankowości elektronicznej, w: Systemy wspomagania organizacji. Informatyka ekonomiczna jako dziedzina nauki i dydaktyki*, Prace Naukowe AE w Katowicach, Katowice.

[9] Chmielarz W. (2011) *Analiza i ocena wybranych internetowych sklepów spożywczych,* Problemy zarządzania nr 1, s. 30-40.

[10] Chmielarz W. (2007) *Analiza metodyk porównania witryn internetowych na przykładzie branży odzieżowo-obuwniczej,* w: Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą, nr 10, Januszewski A. [ed.], PSZW, Bydgoszcz, s. 35-43.

[11] Chmielarz W. (2007) *Analiza porównawcza witryn internetowych w branży kosmetycznej*, http://www.swo.ae.katowice.pl/_pdf/338.pdf, SWO.

[12] Maciejowski T. (2004) *Firma w Internecie,* Oficyna Ekonomiczna, Kraków, s. 124 i 129.

[13] Wang R.Y., Strong D.M. (1996) *Beyond accuracy: What data quality means to data consumers*, Journal of Management Information Systems, nr 12, (4), pp. 5–34.

[14] Ruževičius J., Gedminaitė A. (2007) *Business Information Quality and its Assessment*, Engineering Economics, nr 2 (52), p. 19.

[15] Kahn B.K., Strong D.M., Wang R.Y. (2002) *Information Quality Benchmarks: Product and Service Performance*, Communications of the ACM, Vol. 45, No 4,pp. 186-187.

[16] Stefanowicz B. (2004) *Informacja*, SGH, Warszawa, s. 101-107.

[17] Czerwiński A., Krzesaj M. (2014) *Wybrane zagadnienia oceny jakości systemu informacyjnego w sieci WWW*, Studia i monografie nr 501, Wydawnictwo UO, Opole, s. 49-51.

[18] Krzesaj M. (2016) *Wybrane narzędzia i metody ewaluacji jakości informacji na stronach WWW*, in printing.