Dedicated to the memory of Academician Mitrofan M. Cioban (1942-2021)

Algorithm for assessing the credibility of online information

ADELA GOREA

Abstract. The present paper is devoted to an algorithm for evaluating the data credibility and analyzes a number of available tools for this. During the investigation, it was found that the most tools are used by journalists in doing this. Currently, there are attempts to automate the process of evaluating the information credibility, aimed not only at journalism professionals, but also at the regular user. In this sense, two tools were analyzed: NewsGuard and WOT. The analysis of these tools allows us to determine the mechanisms of collective online analysis (crowdsourcing) in order to evaluate the credibility of online information with the possible usage of computational intelligence without the mandatory experts' participation. The developed algorithm takes over the best functions of the studied tools and can be useful for evaluating the data credibility from various websites.

Keywords: credibility, crowdsourcing, algorithm, browser extension, NewsGuard, WOT.

Algoritm de evaluare a credibilităţii informaţiilor online

Rezumat. În lucrare este descris un algoritm pentru evaluarea credibilităţii datelor şi analizate câteva instrumente disponibile pentru acest scop. În cadrul cercetării, s-a constatat că cele mai multe instrumente pentru evaluarea credibilităţii informaţiei sunt utilizate de către jurnalişti. în prezent, există încercări de automatizare a procesului de evaluare a credibilităţii informaţiilor, destinate nu numai experţilor din domeniul jurnalismului, ci şi utilizatorului obişnuit. În acest sens, au fost analizate două instrumente: NewsGuard şi WOT. Analiza acestor instrumente ne permite să determinăm mecanismele de evaluare colectivă (crowdsourcing) pentru a determina credibilitatea informaţiilor online cu o posibilă utilizare a algoritmilor de inteligenţă computaţională fără participarea obligatorie a experţilor. Algoritmul dezvoltat preia cele mai bune funcţii ale instrumentelor studiate şi poate fi util pentru evaluarea credibilităţii datelor de pe diverse site-uri web.

Cuvinte cheie: credibilitate, evaluare colectivă, algoritm, extensie browser, NewsGuard, WOT.

1. Introduction

The Web has been never so big than now. It contains tremendous amount of information such as common web forms, videotapes, pictures, vlogs and social network publications and a lot of another entranceways. One explanation of this drastic increase is that it is not anymore formed by some of the professionals or dedicated people or foundations, but by every user who can access to it. Even though this has led to enormous information abundance, alternative perspectives and diversification nonetheless, it has as well brought new confrontations. It has discarded the traditional information suppliers from their major role and has forsook the public in the depths of web content with varying quality from reliable and true information to misinformation, facts that are not true [1].

Virtually many countries have data validation platforms, which consist of the groups of independent journalists and the association of independent journalists such as WebMii, FactCheck, Media Bias/Fact Check, International Fact Checking Network, StopFals, etc.

Although there is now a variety of credibility assessment studies for different areas, most previous evaluation models have relied on English content, labeled mainly by western users, which limits their applicability. The perception of credibility is subjective: different groups may have different opinions [2,3].

The purpose of the paper is to develop an algorithm that takes the best features of the existing web tools.

To achieve the purpose, the following objectives have been set: analysis of the existing tools to assess the credibility of data; developing an algorithm for evaluation the credibility of data; designing a database structure to implement the developed algorithm.

To achieve the goal, the article was structured in four parts. In the first part the existing tools for assessing the credibility of the data were researched. Then, based on the analysis of the existing tools, we developed an evaluation algorithm presented in written and schematic form. In the third part, all the points presented in the algorithm were described in detail. In the last part we illustrated an example of the database structure for the implementation of the developed algorithm.

2. Existent instruments for credibility evaluation

In order to investigate, there is a variety of tools that can help providing indicators of credibility on different levels.

Here we present the existing instruments that provide credibility-related assessments, focusing mainly on human-generated evaluations that come from experts in journalism. We can identify two main levels of evaluation: source-level and document-level.

NewsGuard focused on the source-level assessments, has manually and methodically reviewed thousands of news sources which are mostly based in the United States of America. NewsGuard can be used as a Chrome extension as well. When the user browses the news information in different sources, it can show related news sources or it can appear as a result of some web searches. These criterion may be splitted into two arrays: *credibility* (it does not constantly post fake content, constantly corrects or modifies failures, collects and provides trustworthy information, keeps away from misleading headlines, manages the difference between news and point of view responsibly) and *transparency* (the website exposes ownership and financing, uncovers who's responsible, certainly designates advertising, the site gives the names of content makers, beside either contact details, or biographic data). [4]. The websites also come into possession of an overall *evaluation* (total sum of points given for every feature) and a *label*, that can be *Trustworthy*, *Unfavourable*, *Satire* or *Platform* (journals, records, postcasts, contents created for the user or social media). The total sum of credibility and transparency scores is 100 at maximum. A news website is accepted as "safe" in case it has gathered at least 60 points.

Next tool, *My Web of Trust* (WOT), is a crowdfunded prominent service that maintains assessments of world wide web documents through a browser expansion. Two constituent parts with regard to total sum of points and certainty criterion, are *trustworthiness* ("How much do you believe the given internet site?") and *safety* ("How suitable can be this website for children?") [5]. Users can see the overall scores and remarks originating from the community and can supply with their own rating. The user can as well grade it and post a comment with his/her retrospect based on their subjective impress. WOT consists of two mechanisms. These are protection in real time and manual override. Protection in real-time automatically notifies the user of online threats.

If the user comes along a site that has not received a reputation score yet, he can ask the WOT community to grade it. For ranking and reconsiderations, WOT applies intelligent algorithms and manual checking to discover and take off false reviews, and likewise we can also utilize blog and social media (Facebook, Twitter, Google+) for verification [6, 7, 8].

From our point of view in the NewsGuard we can highlight the following strengths: easy to use, can be used by anyone and plugin shows which sites are evaluated and which are not. However, we can mention several weaknesses of this plugin: sites from the other languages can be sent for the analysis, but it is not clear when they will be evaluated; users cannot contribute an opinion on a site; there are no computational intelligence algorithms which evaluate the news [9].

The second instrument, WOT crowdsourced reputation service is a more open tool for assessing credibility by enabling users to assess the credibility of information [10]. This Web tool allows the user to add a review. These reviews will be evaluated by the groups of experts as well as by intelligent algorithms [11].

Voting and/or leaving a user review is an efficient method for establishing credibility at the user level, where we can rank users/accounts by their credibility scores for a given piece of information [12, 13].

3.   GENERALIZED ALGORITHM

Taking into account all the advantages of existing credibility assessment tools described above, we propose the following algorithm that relies on most of the strengths of the previous two instruments in online credibility assessment, such as: easy to use – we can see the already evaluated sites and the possibility for any users to contribute to the evaluation or review, which also would be evaluated by the experts and/or some other intelligent algorithms. We consider important to use the possibilities of intelligent algorithms in order to simplify the necessity of crowdsourcing technics that can be in some cases subjective.

Input: *Browser with information;*

Output: *Browser with the information about credibility score for information.*

(1) *Display of the existing evaluation of the information;*

(2) *Opening of an interface for evaluating information;*

(3) *Data recording of the user who evaluates the information;*

(4) *Data recording of the information in the browser:*

    (a) *Link to information;*

    (b) *Current date;*

    (c) *The language in which the material is written;*

    (d) *The actual information;*

(5) *Data-sheet for assessing the credibility of the information;*

(6) *Initial saving of the form in the database;*

(7) *Assessment of the user form;*

(8) *If the evaluation in the form is accepted, it is permanently entered in the credibility evaluation database;*

(9) *Recalculation of the score for the credibility of the post.*

The algorithm can be applied to different types of information, not only textual, but also multimedia (video, audio and graphic). The obtained algorithm can be applied

not only to news sites, as presented in the NewsGuard tool, but also to any other field (scientific or educational), as well as other types of multimedia information. Algorithm implementation requires the development of user-friendly interfaces for the assessment to be accessible to the reader. The algorithm offers the possibility for any user to express their opinion regarding the credibility of the online information accessed by themselves, as this possibility is missing in NewsGuard and WOT, where the assessment of credibility is done only by expert journalists.

The schematic representation of the algorithm is presented in Figure 1.
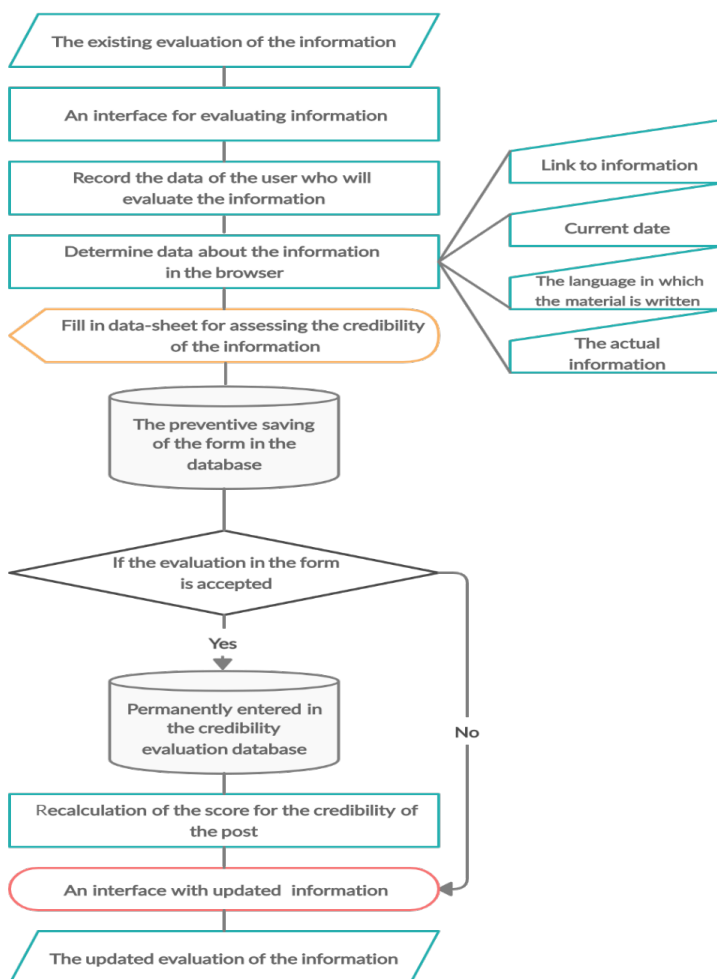
Figure 1. Graphical representation of the algorithm.

In the following we will describe in more detail the algorithm presented above.

4. DETAILS REGARDING THE ALGORITHM

We will specify that the above algorithm is designed for web applications, especially for news sites. If the site or online article has already been evaluated, we consider that the evaluation results would be useful to be visible for the user as a score or comments from other users. This would show that the site has been reviewed by several users before.

The algorithm supposes a mechanism that will allow the user to contribute to the evaluation of the credibility of the web page in this sense. Specific interfaces will be developed that will allow the registration of the user's opinion regarding the online information credibility.

The algorithm offers the possibility that each user could evaluate a lot of news. In this regard, it is useful to record user's data.

Based on how a user evaluates multiple sites, we could calculate a user rating that is given to a particular page. For example, an article has been evaluated by a user. If the score provided by this user has a very similar value to those of the previous evaluators, then this assessment no longer requires additional verification.

Initially, to directly evaluate the credibility of a website, we need some more additional data. Among them we will consider: link to information (automatic), current date (automatic), language in which the material is written (semi-automatic) and actual information (depending on the type of information: text, image, audio, video) is not provided for all types of information. This data deserves to be kept in the database for further processing [14].

When the data about the online article have been recorded, we can register the user opinion about its credibility. This assessment will be done based on a form. The criteria that we consider important based on different studies [15] are presented in Table 1. Each criterion is assessed with a score.

The maximum value that can be accumulated by an article following the evaluation is 25 (100%) points. I agree with the suggestion from article [16], that if an online article obtains at least 15 (60%) points, it would correspond to a credible article or site.

Preliminary data of the evaluation should be saved (for example in XML or JSON format). Based on the preliminary assessment, we will find out if the user evaluation is accepted. If the expert considers the assessment correct, it will be saved in our database, otherwise it will be neglected. After introducing the opinion in the database, the score given to the page will be recalculated based on a new evaluation.

Table 1. Credibility assessment form.

CRITERION	MIN Score	MAX Score
Novelty (up-to-date information in the article)	0	5
Correspondence of the content with the topic of the article	0	5
Originality (the article belongs to an author or is taken from another site)	0	5
Clarity of the content of the article	0	5
Grammar level	0	5
Score	0	25

5.   CREDIBILITY ASSESSMENT APPLICATION DATABASE

Based on the described algorithm above, we propose the following structure of the database presented in Figure 2. We notice that the database consists of three main parts: the first part stores user data, the second part contains data about the article and the third part has the evaluation data in which we keep who and what article evaluated. Taking
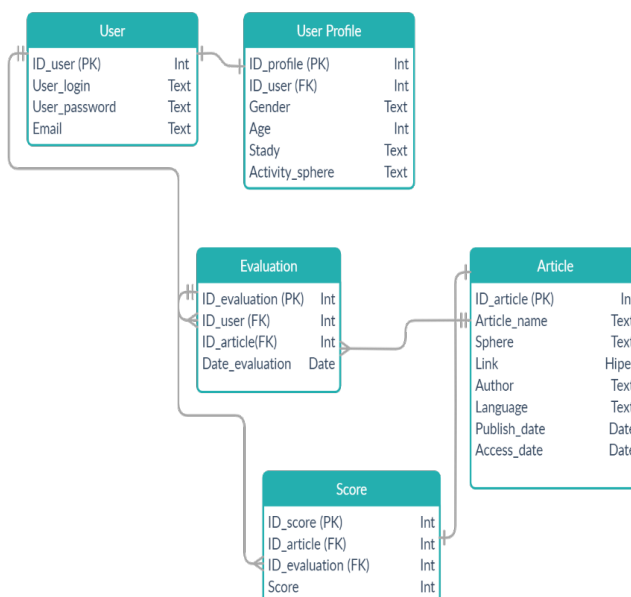
Figure 2. Database structure.

all this into account, the structure reflects these needs of the algorithm presented above.

Probably the volume of information can be exaggerated in volume, and we will study this problem further with the application of distributed data processing technologies that will allow minimizing time and information processing.

6. Conclusions

In this article there were analyzed 5 data verification platforms that consist of independent journalists' groups and the Association of Independent Journalists, such as: WebMii, FactCheck, Media Bias/Fact Check, International Fact Checking Network, StopFals and which are intended exclusively for fact-checking specialists who comment on the link between sources and misinformation.

Having analysed these two tools, NewsGuard and WOT, there were determined mechanisms of collective online analysis (crowdsourcing) for evaluating the credibility of online information with the possible usage of computational intelligence without the mandatory experts' participation.

There was developed an algorithm consisting of 9 steps, based on the analyzed web tools. The algorithm allows any user to express their opinion about the credibility of online information. The algorithm defines 5 criteria for evaluating the credibility of information, each of which can take values from 0 to 5. The score for each criteria corresponds to the confidence level of this article. We contend that 15 points will correspond to credibility of article or website.

The implementation of the algorithm requires the use of a database, the structure of which is stated in the article on the basis of 5 tables, corresponds to the goals of the algorithm in the present article.

References

[1] Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., Nakov, P. Predicting factuality of reporting and bias of news media sources. In: *arXiv preprint arXiv:1810.01765*, 2018, 3528–3539. DOI: 10.18653/v1/D18-1389.

[2] Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., Choi, Y. Truth of varying shades: Analyzing language in fake news and political fact-checking. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, 2931–2937. DOI: 10.18653/v1/D17-1317.

[3] Alarifi, A., Alsaleh, M. Web Spam: A Study of the Page Language Effect on the Spam Detection Features. In Machine Learning and Applications. In: *Machine Learning and Applications (ICMLA)*, 11th International Conference, 2012, vol. 2, 216–221. DOI: 10.1109/ICMLA.2012.229.

[4] Wineburg, S., Mcgrew, S., Breakstone, J., Ortega, T. Evaluating information: The cornerstone of civic online reasoning. In: *Stanford Digital Repository*, 2016. Retrieved January, 8:2018.

[5] ALRUBAIAN, M., AL-QURISHI, M., ALAMRI, A., AL-RAKHAMI, M., MEHEDI HASSAN, M., FORTINO, G. Credibility in Online Social Networks. In: *IEEE Access*, 2019, vol. 7, 2828–2855. DOI: 10.1109/AC-CESS.2018.2886314.

[6] VIVIANI, M., PASI, G. Credibility in social media: Opinions, news, and health information. A survey. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. Hoboken, NJ, USA: Wiley, 2017, vol. 7. DOI:10.1002/widm.1209.

[7] LIA, R., SUH, A. Factors Influencing Information credibility on Social Media Platforms: Evidence from Facebook Pages. In: *Procedia Computer Science*, 2015, vol. 72, 314–328. DOI: 10.1016/j.procs.2015.12.146.

[8] ZUBIAGA, A., AKER, A., BONTCHEVA, K., LIAKATA, M., PROCTER, R. Detection and resolution of rumours in social media. In: *ACM Computing Surveys (CSUR)*, 2018, vol. 51(2):32. DOI: 10.1145/3161603.

[9] PROCHAZKA, F., WOLFGANG, S. How to measure generalized trust in news media? An adaptation and test of scales. In: *Communication Methods and Measures*, 2019, vol. 13, no. 1, 26–42. DOI: 10.1080/19312458.2018.1506021.

[10] KUMAR, R.V., SHERIFF, I. Topical Categorization of Credible Microblog Content. In: *IJSRD - International Journal for Scientific Research & Development*, 2014, vol. 2, 656–658.

[11] GO, E., YOU, K.H., JUNG, E., SHIM, H. Why do we use different types of websites and assign them different levels of credibility? Structural relations among users' motives, types of websites, information credibility, and trust in the press. In: *Comput. Hum. Behav.*, 2016, vol. 54, 231–239. DOI: 10.1016/j.chb.2015.07.046.

[12] ALRUBAIAN, M., AL-QURISHI, M., AL- RAKHAMI, M., ALAMRI, A. A credibility assessment model for online social network content. In: *Social Data Mining and Analysis to Prediction and Community Detection*. Cham, Switzerland: Springer, 2017, 61–77. DOI:10.1007/978.3.319.51367.6.3.

[13] MCKELVEY, K. R., AND MENCZER, F. Truthy: enabling the study of online social networks. In: *Proceedings of the 2013 conference on Computer supported cooperative work companion. ACM*, 2013, 23–26. DOI:10.1145/2441955.2441962.

[14] LIA, R., SUH, A. Factors Influencing Information credibility on Social Media Platforms: Evidence from Facebook Pages. In: *Procedia Computer Science*, 2015, vol. 72, 314–328. DOI: 10.1016/j.procs.2015.12.146.

[15] ZUBIAGA, A., AKER, A., BONTCHEVA, K., LIAKATA, M., PROCTER, R. Detection and resolution of rumours in social media. In: *ACM Computing Surveys (CSUR)*, 2018, vol. 51(2):32. DOI: 10.1145/3161603.

[16] AKER, A., KEVIN, V., BONTCHEVA, K. Predicting News Source Credibility. In: *Conference for Truth and Trust Online*, 2019. DOI:10.36370/tto.2019.5.

(Gorea Adela) ALECU RUSSO STATE UNIVERSITY, STR. PUŞKIN 38, MD-3121, BĂLŢI, REPUBLIC OF MOLDOVA
E-mail address: `adela.gorea@usarb.md`