MODERN MATHEMATICAL APPROACHES TO ASSESSING THE RELIABILITY OF INFORMATION FROM DIGITAL SOURCES IN SCIENTIFIC RESEARCH

Prokopovich-Tkachenko Dmitry
Candidate of Technical Sciences, Associate Professor, Head of the Department of Cybersecurity and Information Technology University of Customs and Finance
ORCID ID: 0000-0002-6590-3898

Zverev Vladymir
Candidate of Technical Sciences, Senior Researcher, Associate Professor Department of Engineering Software and cybersecurity State Trade and Economic University
ORCID ID: 0000-0002-0907-0705

Bushkov Victor
Postgraduate student of the Department of Software Engineering and Cybersecurity State Trade and Economic University
ORCID ID: 0009-0005-5097-2689

Abstract: The article presents a mathematical methodology for evaluating Internet sources, based on the analysis of three key criteria: source authority, peer review and relevance of information. The proposed methodology provides an objective assessment of the reliability of sources through a numerical scale, which simplifies the choice of resources for scientific activities. areas of research that cause irrational expenditure of financial, time and intellectual resources.

The study is extremely relevant in modern conditions, when the so-called "economic order of fakes" dominates, where the spread of false information can destabilize both academic and practical activities. The use of a reliable mathematical model allows you to reduce the risks of making erroneous decisions caused by unreliable data, as well as to support the efficiency of research processes.

The practical significance of the methodology is demonstrated through the evaluation of popular sources in the fields of business, cybersecurity and technical sciences. The article emphasizes the need to give preference to highly reliable sources and carefully work with those with an average level of reliability. The proposed approach will help minimize the negative consequences of the use of false information, conserving resources and promoting the development of scientific research based on verified data.

Keywords: Mathematical Methodology, Reliability Assessment, Internet Sources, Scientific Activity, Authority, Peer Review, Relevance.

Introduction

Modern scientific activity is increasingly based on information obtained from Internet sources, due to their accessibility, prompt dissemination and wide thematic coverage. However, the lack of universal criteria for assessing the reliability of such sources creates significant risks for the reliability of scientific conclusions. academic research and practical decision-making (Lazer et al., 2018).

Studies show that the application of machine learning algorithms allows for the effective identification of fake news and the assessment of the credibility of sources. For example, Shu et al. (2017) emphasize text analysis, source authority, and fact-checking through reliable references as key factors for assessing reliability. Another approach is proposed by Zhang et al. (2019), which explore the possibility of using blockchain technology to ensure transparency and authenticity of data, which is especially important for combating manipulative information.

Thus, the development of mathematically sound methods for evaluating Internet sources not only contributes to improving the quality of scientific research, but also helps to avoid risks associated with resource expenditure and decision-making based on unreliable data. These approaches become the basis for systematizing the verification of sources and increasing confidence in scientific results.

Ensuring the reliability of information from online sources is a critical task, given the importance of accurate data for scientific research. First of all, it is necessary to implement systematic methods of source verification. This includes the use of machine learning algorithms to detect fake news, analysis of the reputation of publications and authors, as well as fact-checking through cross-references to peer-reviewed publications (Shu et al., 2017). Technologies such as blockchain offer innovative opportunities to ensure data transparency and ensure its authenticity. This is important to prevent manipulation that can distort research results (Zhang et al., 2019).

It is also worth actively using authoritative scientific databases, where articles undergo a rigorous peer review process, which minimizes the risk of spreading false information. For example, resources such as PubMed, Scopus or Web of Science have quality standards that can be adapted to create new systems for assessing the reliability of Internet sources.

Scientific communities should also promote ethical norms for working with information, emphasizing the importance of data accuracy and transparency. In addition, critical thinking skills should be improved in researchers and students to analyze sources and identify potentially manipulative content (Lazer et al., 2018).

The integration of these approaches will not only minimize the risks of using false data in research, but also create a systematized platform to increase confidence in the results of science.

The development of critical thinking is key in the fight against fake information and manipulation on the Internet. For scientists and students, this means the ability to assess the reliability of sources based on a number of criteria: authoritative authority, reputation of the publication, the availability of peer-reviewed articles and factual data confirmed by other studies (Shu et al., 2017). The use of curricula that develop analytical thinking can significantly improve the ability to distinguish between reliable information and manipulative information.

In addition, practical tasks aimed at analysing different types of sources should be implemented, including finding primary sources and fact-checking using authoritative platforms such as Google Scholar, FactCheck.org, or Snopes. Using fake news cases as educational examples will help you learn how to identify false information in real-world situations (Lazer et al., 2018).

The application of fact-checking technologies is also promising. Automated data analysis services and blockchain can guarantee the transparency and accuracy of information, ensuring the authenticity of content (Zhang et al., 2019). Finally, the development of critical thinking should be accompanied by constant support for ethical norms in working with information to avoid the spread of manipulation or unreliable data.

1. In the modern information environment, Internet sources become the basis for many scientific studies due to their availability and efficiency. However, the lack of uniform standards for assessing the reliability of information creates risks for the quality of scientific conclusions, increasing the danger of spreading manipulative or false information. The use of such data can lead to distortion of academic discourses and significant expenditure of resources. An analysis of existing approaches demonstrates that the use of machine learning algorithms, blockchain technology, and authoritative scientific databases are promising tools for improving the process of assessing the reliability of sources. In addition, the development of critical thinking and ethical standards for working with information among scientists and students provides an additional level of data verification.

Purpose of the study

To develop a mathematically grounded methodology for evaluating Internet sources, which would integrate modern data analysis technologies (machine learning algorithms, blockchain) and consider ethical and critical approaches to assessing the reliability of information. The main task is to create a universal tool that minimizes the risks of using false data, increases confidence in scientific results and ensures optimal use of resources in the decision-making process.

This research lays the foundation for systematizing source verification, increasing the transparency of scientific processes and promoting adaptation to the challenges of the "fake economy".

Methods

Modern innovative methods, such as machine learning algorithms and blockchain, are promising for solving the problem of assessing the reliability of Internet sources. Machine learning can analyse large volumes of texts to identify fake information using neural network models, classification algorithms, and textual characteristic analysis (Shu et al., 2017). Blockchain, in turn, ensures transparency and immutability of data through decentralized fixation of information (Zhang et al., 2019). However, these approaches have certain limitations: complexity in implementation, the need for significant resources, and a potentially high degree of digital hallucinations in machine learning, which can affect the accuracy of the results. In view of these challenges, a simplified and understandable mathematical method is proposed to assess the reliability of sources. This method can be based on multivariate analysis, which uses criteria such as the reputation of the source, which takes into account its history and scientific evaluation (for example, the presence of peer-reviewed publications or cooperation with reputable organizations); crosschecking, which compares information with other sources on similar topics to identify discrepancies or corroborate; frequency of citations, which analyzes how often the source is mentioned in scientific databases (PubMed, Scopus, etc.); and structural analysis of the text, which reveals potential manipulations through an assessment of the grammar, style and logic of the presentation of the data. This technique is less technically complex but allows you to get an initial assessment of the reliability of the source. It can be easily implemented in educational and scientific environments, providing basic transparency to the analysis process. Thus, the use of simple mathematical approaches combined with innovative methods can provide a flexible and reliable solution to the problem, creating a basis for adaptation to a rapidly changing information environment.

To assess the reliability of each source, various criteria can be used, such as the authority of the source, peer review, relevance of information, etc. Suppose that we evaluate each source on a scale from 1 to 10 according to the following criteria:

1. Authority of the source (A)
2. Peer Review (R)
3. Relevance of information (U)

Then the overall confidence score (T) will be the average of these three criteria:

$$T = \frac{A + R + U}{3}$$

Here is a table with ratings for each source	**Authoritative (A) General**	**Review (R)**	**Current Articles (U)**	**General**
Gartner	9	8	9	9+
McKinsey	Company	9	8	9
IDC	8	7	8	8+i
Cisco	9	8	9	9+
Accenture	8	7	8	8+i
Harvard Business Review	9	9	9	9+!
Forbes	8	7	8	8+i

Thus, the general estimates of reliability for each source are as follows:

Here is a list of organizations with their official websites:

1. **Gartner**: www.gartner.com
2. **McKinsey & Company**: www.mckinsey.com
3. **IDC**: www.idc.com
4. **Cisco**: www.cisco.com

5. **Accenture**: www.accenture.com
6. **Harvard Business Review**: hbr.org
7. **Forbes**: www.forbes.com

To develop cybersecurity strategies, it is important to use sources with high reliability, as this ensures the reliability and relevance of information. Based on the reliability assessments, the following conclusions can be drawn:

8. Sources with a score of 8.67 and above:

1. Gartner: 8.67
2. McKinsey & Company: 8.67
3. Cisco: 8.67
4. Harvard Business Review: 9.00

These sources are highly reliable and can be used to develop cybersecurity strategies. They provide reliable and up-to-date information, which is critical for this area.

9. Sources with a score of 7.67:

1. IDC: 7.67
2. Accenture: 7.67
3. Forbes: 7.67

These sources can also be used, but with some caution. They have average reliability and can be useful as additional sources of information. It is recommended to check information from these sources with the help of other, more authoritative sources. Three main criteria were chosen to assess the reliability of Internet sources: authority of the source (A), peer review (R) and relevance of information (U). Source authority assesses an organization's reputation in the professional or scientific community, considering its history, impact on the industry, and recognized expertise. For example, organizations such as Gartner or McKinsey & Company have high credibility due to their years of experience in analysis and consulting in business and technology. Peer review reflects the degree of scientific verification of the information published by the source.(R) For example, Harvard Business Review regularly publishes materials that are thoroughly reviewed by the editorial team, while IDC and Accenture articles are supported by empirical evidence. Sources without rigorous peer review, such as Forbes, receive a lower score on this criterion. Information relevance (U) assesses the relevance of the data provided to current conditions and trends, considering the frequency of updates, responsiveness to new challenges, and relevance of content. Cisco regularly updates cybersecurity data, and Gartner publishes annual analytical forecasts that are widely used in the industry.

Scores (from 1 to 10) were assigned to each source for each of the criteria based on their well-known characteristics. The overall confidence score (T) was calculated as an average using the formula:

$$T = \frac{A + R + U}{3}$$

For example, for Harvard Business Review, authority (A) is rated 9 due to its high reputation among the academic and business communities, peer review (R) is also rated 9 due to rigorous review of materials, and relevance (U) is rated 9 due to regular data updates. Thus, the overall estimate is:

$$T = \frac{9 + 9 + 9}{3} = 9$$

In the case of Forbes, credibility (A) was rated at 8, since it is a well-known business publication, but without a narrow specialization; peer review (R) received a 7 due to the lack of rigorous scientific verification, and relevance (U) was rated 8 for prompt news coverage. Overall rating for Forbes:

$$T = \frac{8 + 7 + 8}{3} = 7.67$$

This approach allows you to identify sources with a high level of reliability that can be used for strategically important research, for example, in the field of cybersecurity. Sources with a score of 8.67 and above, such as Gartner, McKinsey & Company, Cisco, and Harvard Business Review, can be seen as the most reliable. Sources with a score of 7.67, like IDC, Accenture, and Forbes, are recommended to be used as additional, with the obligatory verification of their data using more reputable resources.

Let's conduct a comparative analysis of a reliable and unreliable source using the same methodology. To do this, let's take two sources:

1. Reliable source: Scientific article from a peer-reviewed journal.
2. Unreliable source: A blog or personal website without review.

We evaluate these sources according to three criteria: authority (A), peer review (R) and relevance (U).

Reliable source: Scientific article from a peer-reviewed journal

Criterion and	These are the
Authoritative (A)	10
Peer Review (R)	10
Current Articles (U)	9
Total Eye (T)	$\frac{10 + 10 + 9}{3} = 9.67$

Unreliable source: Blog or personal website

Criterion and	These are the
Authoritative (A)	4
Review (R)	2
Current Articles (U)	5
Total Eye (T)	$\frac{4+2+5}{3} = 3.67$

Conclusion

1. A reliable source (a scientific article from a peer-reviewed journal) has a high overall reliability rating (9.67) and is reliable for use in the development of cybersecurity strategies.

2. An unreliable source (blog or personal website) has a low overall credibility score (3.67) and is not recommended for use in serious research or development of cybersecurity strategies.

Thus, sources with a high reliability score, such as scientific articles from peer-reviewed journals, should be used to develop cybersecurity strategies, and unreliable sources, such as blogs or personal websites without peer review, should be avoided.

Of course, let's do a comparative analysis with real sources, including a specific blog. Let's take two sources:

1. Reliable source: Scientific article from the journal IEEE Transactions on Information Forensics and Security.

2. Unreliable source: A blog on a personal website, such as John's Tech Blog.

We evaluate these sources according to three criteria: authority (A), peer review (R) and relevance (U).

Reliable source: Scientific article from the journal IEEE Transactions on Information Forensics and Security

Criterion and	These are the
Authoritative Articles (A)	10
Peer Review (R)	10
Current Status (U)	9
Common Eye (T)	$\frac{10+10+9}{3} = 9.67$

Unreliable source: Blog on John's Tech Blog

Criterion and	These are the
Autistic Articles (A)	3
Review (R)	2
Current Articles (U)	4
Total Eye (T)	$\frac{3+2+4}{3} = 3.00$

Conclusion by example:

1. A reliable source (a scientific article from the journal IEEE Transactions on Information Forensics and Security) has a high overall reliability rating (9.67) and is reliable for use in the development of cybersecurity strategies.

2. An unreliable source (a blog on John's Tech Blog) has a low overall credibility score (3.00) and is not recommended for use in serious research or developing cybersecurity strategies.

Thus, sources with a high credibility score, such as scientific articles from peer-reviewed journals, should be used to develop cybersecurity strategies, and unreliable sources, such as blogs or personal websites without peer-reviewed, should be avoided

Discussion

The proposed mathematical methodology for assessing the reliability of Internet sources has several significant advantages. Its ease of use makes it easy to adapt the technique to different areas of scientific activity. The universality of the approach contributes to the analysis of both scientific articles and publications on websites, evaluating them according to common criteria. The technique provides a structured approach to analysis, which increases the accuracy of assessing the reliability of sources. An important advantage is the possibility of extending the method by including additional criteria, such as citation index or compliance with interdisciplinary standards. However, the technique also has certain limitations. The main thing is the subjectivity of the assessment of criteria, because different researchers can interpret the level of authority or relevance of the source in different ways. There is also a need to unify assessment scales to ensure maximum objectivity. The complexity of automating the assessment process, especially when working with large amounts of data, is another challenge. In addition, the technique may prove to be less effective for analysing sources in specific industries, which have their own unique requirements for assessing reliability.

Conclusions

The mathematical methodology for assessing reliability is an effective tool for systematizing the analysis of Internet sources in scientific activities. It helps to structure the assessment process, improves the quality of research and reduces the risks of using inaccurate data. The proposed approach can be adapted to different areas due to its versatility and expandability. Further research can be aimed at automating the assessment, expanding the criteria and adapting the methodology to specific areas of knowledge. Despite certain limitations, the use of this technique contributes to the creation of a basis for increasing the transparency and reliability of scientific processes.

REFERENCES

1. Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094–1096. https://doi.org/10.1126/science.aao2998
2. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36. https://doi.org/10.1145/3137597.3137600
3. Zhang, X., Xu, J., & Liu, J. (2019). Blockchain-based trust mechanism for IoT systems. Future Generation Computer Systems, 92, 33–45. https://doi.org/10.1016/j.future.2018.09.027