THE PUBLIC'S UNDERSTANDING OF "EVOLUTION" AS SEEN THROUGH ONLINE SPACES

Hyoung-Yong Park
Gyeongin National University of Education, Republic of Korea
E-mail: hypark@ginue.ac.kr

Hae-Ae Seo
Pusan National University, Republic of Korea
E-mail: haseo@pusan.ac.kr

Abstract

Evolution is a central concept that unifies all areas of life sciences. Despite longstanding scientific efforts in science education, the public's scientific awareness of evolution still needs to improve. Furthermore, teaching evolution is subject to recurring controversy. This study aimed to investigate the gap between public understanding of evolution seen through online spaces and contents in a school curriculum and explore its reasons. A content analysis was conducted using data mining on a major online portal in Korea. It examined the characteristics of creating and consuming content on evolution through the online portal service based on analyzing the number of posts related to biological evolution and active participants. It also discussed the feasibility of automatic document classification to distinguish between scientific understanding and non-scientific beliefs on the evolution and related online circulating contents. The results show that there are tactics for public exposure and dissemination of creationism through online discussions.
Keywords: *automated classification, machine learning, network analysis, public understanding of evolution*

Introduction

It has been widely acknowledged that no life phenomenon can be understood without an evolutionary perspective (Dobzhansky, 1973). For many scientists and science educators today, evolution is accepted as a unifying paradigm for the life sciences and a central idea that unifies many single concepts in biology. In line with this view, national curricula in many countries propose to cover evolution as the most important unifying concept in biology, and many studies have emphasized the importance of an integrative perspective based on the concept of evolution (AAAS, 1993; Fredrick et al., 1994; Rutledge & Warden, 2000; Scharmann & Harris, 1992).

Even though the scientific community in many countries around the world recognizes and supports evolution as a scientific theory that explains the history of life, public awareness of evolutionary theory remains low. Although evolution is a paradigm that unifies the life sciences, there is much resistance and controversy to the basic explanatory framework of evolution in education (Young & Strode, 2009).

https://doi.org/10.33225/BalticSTE/2023.173

Therefore, many students learn life science in a social context that hinders their scientific understanding of the history of nature (Kahan et al., 2011). It is reflected in the controversy over the revision of textbooks.

The Society for the Revision of Evolution Theory in Textbook (Gyojinchu; an anti-evolution group in Korea) made waves in the Korean science education community in 2011 by their petition. They are campaigning to remove content about “the evolution of humans” and “the adaptation of finch beaks based on habitat and mode of sustenance”, a reference to one of Darwin’s most famous observations (Park, 2012). As such, creationists have long attempted to change the public perception of evolution by stirring up controversies (Park, 2001).

On the other hand, with the development of information technology, learners increasingly rely on online media, such as searching for knowledge through the Internet, rather than traditional media. In particular, the influence of information on the Internet is expanding, such as online question/answer and encyclopedia services that pursue collective intelligence based on very high accessibility. However, because online content can be written and read by anyone, there are many concerns about whether publicly shared online information is scientifically correct or not. Moreover, non-scientific information and texts widely propagated online can be a reproduction tool that misleads students who need to be discerning. Therefore, it is necessary to have measures in place to monitor and discern the circulation of such information in a non-school context.

Research Aim and Research Questions

In this context, it is necessary to study how the public's understanding of "evolution" in online space differs from the content covered in life science education. Furthermore, based on the results, it is also necessary to draw educational implications for the correct understanding of the evolution of life. Therefore, according to the context and need for such a study, the research conducted in this study is as follows.

1) Analyzing aspects of online writing (question/answer) activities related to 'evolution'

2) Analyzing features of 'evolution' related posts registered in online knowledge services

3) Exploring the possibility of automated classification and filtering of 'evolution' related online posts

Research Methodology

General Background and Procedures

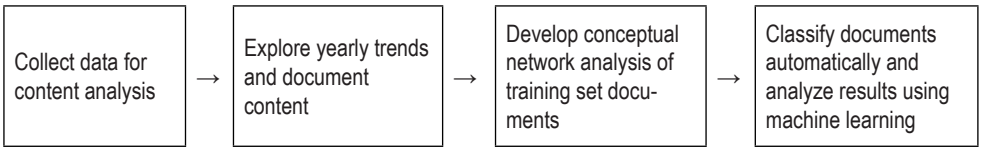
To explore the public’s understanding of evolution, the researchers targeted Jisik-iN (The same pronunciation as "intellectual" in Korean). As a representative online communication space, this service supports the exchange of information by asking and answering questions among the users in Korea (like Quora).

This service was started in 2002 by N company, which has the highest share of Internet search engine users in Korea at about 55%. Since it has the largest number of users, much information has been accumulated. However, unlike Wikipedia, the viewer

cannot modify it, so incorrect knowledge is often left unattended, and this is also where the problems of knowledge search services are most prominent.

The study employed descriptive content analysis, text network analysis, and AI-based document classification techniques to analyze data collected from a specific online space over eighteen years, from 2002 to 2019. The data was gathered and analyzed following the research procedure depicted in Figure 1.

Figure 1
Procedure of the Study



Sample

The researchers collected questions and answers through data mining on Q&A services of the major search portals selected for analysis. In the data collection process, the categories were limited to 'biology' and 'life science', and the keyword 'evolution' was used to search for questions/answers, open bases, and posts (documents). Through the data collection process, 12,130 answers to 4,051 online questions and 438 open-encyclopedia articles were collected for content analysis.

Data Analysis

To analyze the trend of 'evolution' related online writing activities, a frequency analysis was conducted to explore trends and document contents by year. Then for the automatic classification process of the collected document data, documents corresponding to 10% (1,278) of the full documents were randomly selected and used as a training data set for automatic classification. Through the researchers' review of the documents in the training data set, the documents were classified into 'Scientific (SC)', 'Non-scientific (NS)', and 'Other (OT)', and representative documents were selected centering on the posts of authors with high activity. The classification of the training data showed that 61 documents contained scientific (SC) ideas about evolution, 68 documents contained non-scientific (NS) ideas, and the remaining 1,149 documents fell into the other (OT) category.

Next, a conceptual network analysis was conducted for the Korean national curriculum documents (Ministry of Education, 2015) and documents representing SC and NS groups. The features and meanings of the networks' relationship were extracted by analyzing the conceptual networks. Then, based on the network analysis results, machine learning (ML) features were extracted for an artificial intelligence system that can automatically classify online documents on evolution into SC and NS groups.

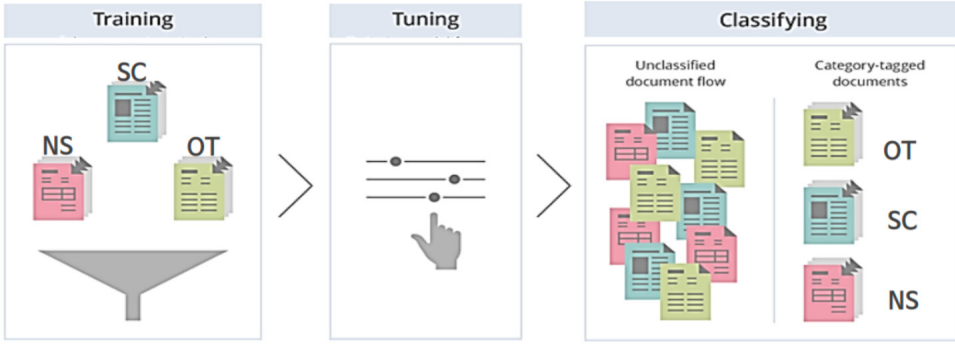
Finally, a supervised machine-learning approach was employed for each document class using the training set to classify the collected documents. This process involved TF-IDF-based automatic classification of all the documents. Principal Component Analysis

https://doi.org/10.33225/BalticSTE/2023.173

(PCA) was used to visualize and interpret the results of the document classification, grouping the documents into distinct categories.

Figure 2

Autonomative Classification Process of Online Documents Related to "Evolution"



Research Results

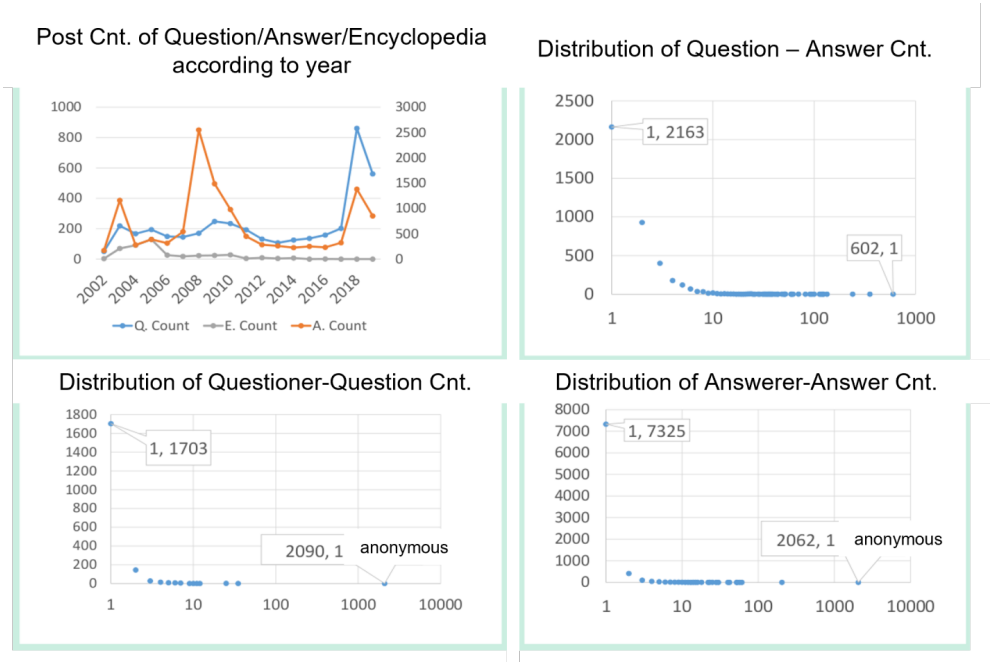
Trends in Online Authoring (Question/Answer) Activity Related to "Evolution"

The frequency analysis results on 'evolution' related to online writing activities was shown in Figure 3. Evolution-related online question/answer activity has been cyclical and volatile, with a recent upward trend. It is thought that online question/answer activity tends to increase around periods of heightened public interest in evolution, such as curriculum revisions and *the petition of the Society for the Revision of Evolution Theory in Textbook (Gyojinchu)* controversy.

Over 75% of the questions received two or fewer replies, and less than 3% received ten or more. Excluding anonymous posters, less than 1% of users have written six or more questions or answers about "evolution", and less than 1% of users have written more than 5% of total questions and 10% of total answers. This result shows that some users are highly active. Therefore, it is crucial to focus on the documents created by these users to determine whether they reflect scientific knowledge about evolution or contain non-scientific content such as religious beliefs. The trends in creating online articles about evolution suggest that online knowledge about evolution is likely to be heavily influenced by a small number of highly active users and anonymous authors.

https://doi.org/10.33225/BalticSTE/2023.173

Figure 3
Trend of "Evolution" Related Online Q&A Activities



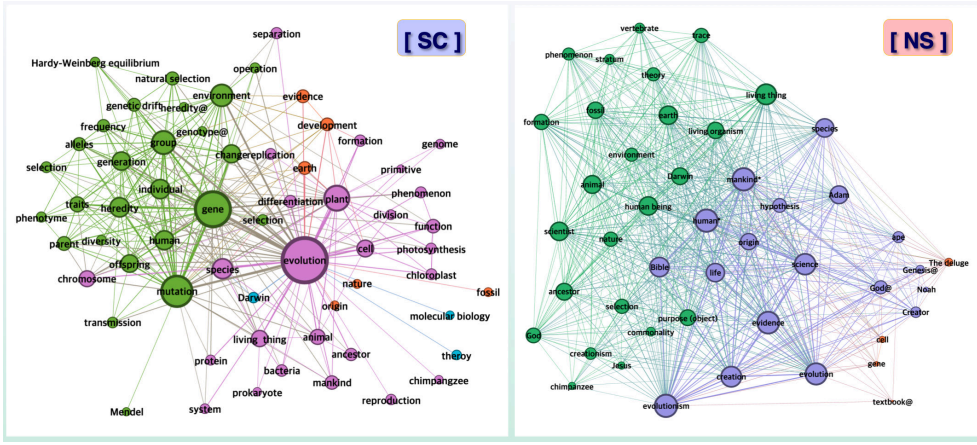
Conceptual Network of Online Answer Threads Related to "Evolution"

Figure 4 shows a text network of two groups' evolution related documents in the online space. Among the online responses, documents containing scientific knowledge (SC) about evolution showed a high centrality of concepts necessary to explain how life evolves by natural selection, such as "genes", "mutations", "populations", "alleles", and changes in the gene pool of a population, such as the "Hardy-Weinberg equilibrium". It is clear that the concept of "evolution" is a crucial concept that integrates several concepts related to the continuity and diversity of life. On the other hand, the concept relationship network for documents containing non-scientific knowledge (NS) showed a high centrality of concepts related to religious beliefs, such as "Bible", "God", and "Genesis". It formed a dense relationship network around these concepts. Contrasts such as 'evolution' and 'creation' were identified, as well as relationships indicating an objectivist worldview based on 'human' thinking. The appearance of concepts such as 'textbook' suggests that these documents are related to creationist arguments.

Compared to SC documents, NS documents were characterized by a higher density and relatively low modularity, suggesting that NS documents tend to have a higher degree of thematic cohesion. Therefore, the differences in the structural features of the relationship network and conceptual organization of the two types of documents can be used as good features for automatic document classification.

Figure 4

"Evolution" Related Online Post Contents' Conceptual Network



* *Notes*. Visualization of only the top 15% based on the frequency of relationships, which indicates the degree of each concept appeared together in the documents

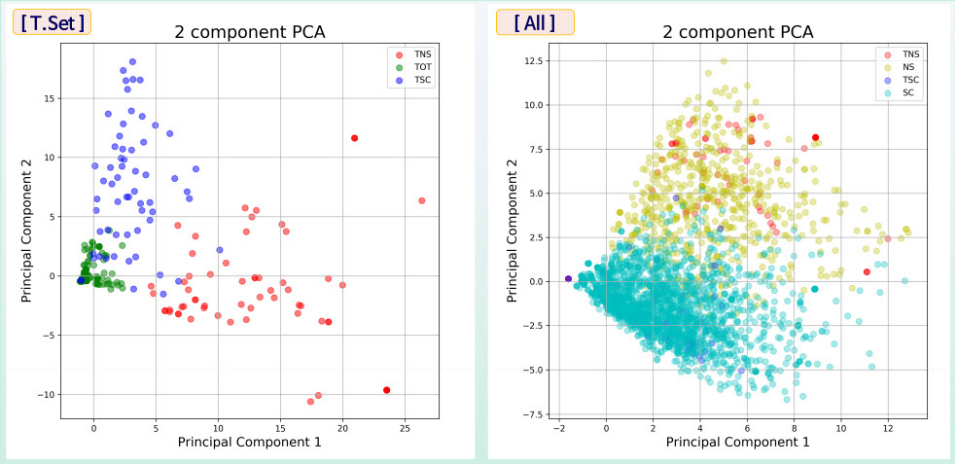
Automated Classification of "Evolution" Related Online Posts Using Machine Learning

Finally, to explore the possibility of automated classification and filtering of 'evolution' related online posts, the researchers selected training data through the analysis of highly active users' answer posts and frequently answered questions. As a result, 200 keywords were identified through concept network analysis. Then, TF-IDF values of the online documents were used as features to vectorize the documents. As a result, a supervised machine learning model for automated classification – Scientific, None-Scientific, Others - was created using the vectorized document data.

In Figure 5, documents are distributed in as many dimensions as the number of features is reduced to two dimensions through principal component analysis (PCA) and visualized. It can be seen that the classification results of the training documents form unique groups by type. The trained model was used to classify the entire online answer posts. The PCA analysis showed that documents containing scientific knowledge (SC) and documents containing non-scientific content (NS) formed separate groups around the respective training data set. Therefore, the automatic classification of online documents on evolution can reduce the public's unprotected exposure to non-scientific content online.

Furthermore, the results of the document classification showed that the number of documents over time increased and decreased, with 5-10% of online responses classified as containing non-scientific explanations of evolution across almost all periods. It is necessary to refine the automatic document classification model through further analysis of the documents classified as Other (OT). It may be possible to distinguish between scientific and non-scientific documents using unsupervised learning methods.

Figure 5
"Evolution" Related Online Post Contents' Conceptual Network



Discussion

As a result of this study, the information about evolution shared online contains non-scientific information. It also reveals that a few active users play a central role in producing and distributing such non-scientific information. Online media has risks because communication in such an online public space is exposed to the unspecified majority without filtering and refinement process. It is also dangerous because it can cause the illusion that a small group's non-professional thoughts are those of the majority.

So, it is necessary to understand the nature of the 'double-edged sword' that online collective intelligence services such as Wikipedia and Quora, which are operating as platforms for effective knowledge sharing today, can have (Wang et al., 2013). Moreover, many new technologies, such as data mining and artificial intelligence, can be effectively utilized (Shu et al., 2017). The research needed to identify and filter non-scientific contents and users who abuse the open attributes of online communities should be continued.

In addition, the cyclical volatility of evolution-related discussions in the online space suggests that an attempt is being made to give equal status to creationism and evolutionism through online space concerning revising the national curriculum. However, online campaigns that reproduce such non-scientific viewpoints are a kind of media manipulation (Fitzpatrick, 2018) that exploits the open nature of Internet media. In the long run, it will become a significant obstacle to the public's scientific understanding of evolution.

Park (2001) already argues that creationists use debates to disseminate their ideas and create the impression that they are on equal footing with the scientific community. Creationists can gain attention and legitimacy by participating in public debates, even if their arguments lack scientific evidence. Additionally, debates can be used to sow confusion and doubt among the general public, ultimately hindering the acceptance of scientific theories. Thus, scientists and educators should recognize this tactic and take steps to counter it through effective communication and education.

https://doi.org/10.33225/BalticSTE/2023.173

Conclusions and Implications

This study explored the public's understanding of evolution and the potential for filtering out non-scientific information by analyzing the texts generated and communicated in online spaces. The conclusions from this study can be summarized as follows.

First, the number of online posts related to evolution has recently shown some fluctuations, with 5% and 10% of posts containing non-scientific beliefs, depending on the period. Given that a small number of highly active or anonymous users can significantly influence public perceptions of evolution through the question/answer process. It seems necessary to continue monitoring the generation of relevant knowledge online.

Second, the conceptual network of documents related to evolution was visualized and analyzed to compare those containing non-scientific knowledge based on religious beliefs with those containing scientific explanations. The analysis revealed significant differences in the structure and conceptual organization of the two networks. Based on these findings, replacing concepts that form the non-scientific understanding of evolution and developing educational measures to promote a correct understanding of the topic will be necessary.

Third, this study explored the possibility that information processing technologies such as data mining, natural language processing, and machine learning can be effectively used to classify knowledge (documents) in specific science-related areas. This outcome can be a robust tool for filtering learning materials to provide learners with reliable scientific knowledge and for building artificial intelligence (AI) systems that can continuously and automatically assess learners' understanding. So, further research should be conducted in this area.

Declaration of Interest

The authors declare no competing interest.

References

AAAS. (1993). *Benchmarks for science literacy: A project 2061 report*. Oxford University Press.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher, 35*(3), 125–129. http://www.jstor.org/stable/4444260

Fitzpatrick, N. (2018). Media manipulation 2.0: The impact of social media on news, competition, and accuracy. *Athens Journal of Mass Media and Communications, 4*(1), 45-62. https://doi.org/10.30958/ajmmc.4.1.3

Frederick, A., Swarts, O., Roger A., & Frank J. (1994). Evolution in secondary school biology textbook of the PRC, the USA, and the latter stages of the USSR. *Journal of Research in Science Teaching, 31*(5), 475-505. https://doi.org/10.1002/tea.3660310505

Gess-Newsome, J., & Lederman, N. (1993). Pre-service biology teachers' knowledge structures as a function of professional teacher education: A year-long assessment. *Science Education, 77*(1), 24-45. https://eric.ed.gov/?id=EJ458314

Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research, 14*, 147–174. https://doi.org/10.1080/13669877.2010.511246

Ministry of Education. (2015). General overview of the 2015 revised curriculum. Ministry of Education Notice No. 2017-74 [Annex 1].

Park, H. J. (2001). The creation-evolution debate: Carving creationism in the public mind. *Public Understanding of Science, 10*(2), 173. https://doi.org/10.1088/0963-6625/10/2/302

Park, S. B. (2012). South Korea surrenders to creationist demands: Publishers set to remove examples of evolution from high-school textbooks. *Nature, 486*(7401), 14-15.

Rutledge, M. L., & Warden, M. A. (2000). Evolutionary theory, the nature of science & high school biology teachers: Critical relationships. *The American Biology Teacher, 62*(1), 23-31. https://doi.org/10.2307/4450822

Scharmann, L. C., & Harris, W. M. (1992). Teaching evolution: Understanding and applying the nature of science. *Journal of Research in Science Teaching, 29*(4), 375-388. https://doi.org/10.1002/tea.3660290406

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.

Wang, G., Gill, K., Mohanlal, M., Zheng, H., & Zhao, B. Y. (2013, May). Wisdom in the social crowd: an analysis of Quora. *In Proceedings of the 22nd international conference on World Wide Web* (pp. 1341-1352). http://dx.doi.org/10.1145/2488388.2488506

Young, M., & Strode, P. (2009). *Why evolution works (and creationism fails).* Rutgers University Press.

Received: *April 16, 2023* Accepted: *May 14, 2023*

Cite as: Park, H.-Y, & Seo, H.-A. (2023). The public's understanding of "Evolution" as seen through online spaces. In V. Lamanauskas (Ed.), *Science and technology education: New developments and Innovations. Proceedings of the 5th International Baltic Symposium on Science and Technology Education (BalticSTE2023)* (pp. 173-181). Scientia Socialis Press. https://doi.org/10.33225/BalticSTE/2023.173