CORRESPONDENCE

Let's publish full-text scientific articles in HTML, not just PDF

Libor Ansorge

TG Masaryk Water Research Institute, Podbabská 30/2582, 160 00 Prague, Czech Republic; libor.ansorge@vuv.cz, ORCID: 0000-0003-3963-8290

The dominant language of modern science is English, and some scientific journals previously published in other languages have switched to it.[1] However, only about 20% of the world's population speak English. Our digital age has brought three tools to work around this limitation. The first is the Internet, thanks to which we can read information from all over the world, including scientific articles, from virtually anywhere. The second tool is the editorial system of scientific journals. Most scientific journals use a content management system. Even small publishers who publish a diamond/platinum open-access journal and cover the cost of the entire journal from their own resources can use platforms such as the Open Journal System (OJS). OJS and similar platforms allow them to efficiently manage the workflow of a scientific journal with minimal operating costs. The third tool is a machine translation. In recent years, applications such as Google Translate have made remarkable improvements, and if the source text is written concisely and clearly, then machine translation is "a tool with great potential for overcoming language barriers".[2] Some publishers even integrate a machine translator directly into the journal's website, such as the *Korean Environmental Engineering Research* journal.[3]

So what's the problem? Machine translation cannot handle PDF text placed on a web page, as it works on the principle of overwriting the content of a page, and PDF does not allow it. In other words, Google Translate can handle almost every article information given in HTML, except for the PDF article.

Yet solutions are there. For example, OJS makes it possible to create an HTML version of the article, but this requires some extra work from editors, which is probably why many scientific journals provide full text articles in PDF format alone. The *European Science Editing* journal also belongs to this majority, even though the Arpha[4] publishing platform it uses enables publishing full texts in HTML, as evidenced by the *Biodiversity Data Journal*[5] published on the same platform.

In addition, text-to-speech programs used by visually impaired readers have a similar problem with PDF files as machine translation has. For the test, I installed the Read Aloud: A Text to Speech Voice Reader extension and read one article from the *European Science Editing*. The extension could read only parts of the text in PDF. At the same time, it could not distinguish between sentences, so the result would be confusing for the blind.

To conclude, current technologies make it possible for editors and publishers to remove barriers to the accessibility of full-text scientific articles to all readers, regardless of their language skills or visual impairment. So why not make use of them?

Conflicts of interest
None to declare

References

1 Kaplan RB. English — the Accidental Language of Science. In: Ammon U (ed) *The Dominance of English as a Language of Science: Effects on Other Languages and Language Communities*. Berlin, New York: De Gruyter Mouton, 2011:3-26. doi:10.1515/9783110869484.3

2 Rivera-Trigueros I. Machine translation systems and quality assessment: a systematic review. *Lang Resources & Evaluation*. Published online April 10, 2021. doi:10.1007/s10579-021-09537-5

3 Korean Society of Environmental Engineers. Environmental Engineering Research. https://www.eeer.org/ (Accessed September 27, 2021)

4 Pensoft. ARPHA. http://www.arphahub.com/ (Accessed September 27, 2021)

5 Biodiversity Data Journal. https://bdj.pensoft.net/ (Accessed September 27, 2021)