

A new paradigm for the scientific article

Anne-Katharina Weilenmann

BIBLINK.ch, Schaffhausen, Switzerland

E-mail: weilenmann@shinternet.ch

Abstract. Information overload is a great problem for the scientific community. To deal with the abundance of new scientific articles there are methods of sophisticated information retrieval tools and text mining tools. Little is known about the relationship between document structure and the structure of thought. This paper describes a project in this matter.

Keywords: Scientific article, deconstruction, reading behavior, new writing model

1. Introduction

Considering information overload and the abundance of new scientific articles,¹ which is becoming a greater problem for the scientific community, there has to be a new scientific discourse for finding innovative ways to cope with the growing amount of information. How can researchers of today manage this, how can they organize their personal information (over)load?

There are different solutions for optimizing the research process and literature review, including alerting services, personalization and smart technologies, sophisticated information retrieval and text mining tools. Text mining is not only an emerging trend, it is an indispensable tool to detect new patterns and findings in scientific texts.²

But is this enough to support the scholars and scientists of the future? There should be a way to get a step further and to find a new approach. It is no longer the search and retrieval process that should be the main focus of interest, but the granularity of the information, the scientific article itself.

A new pattern is emerging New forms and ideas are being developed to improve and to reorganize the scientific article. I have, therefore, designed an empirical study³ to address these issues by investigating two main questions:

1. What kind of cognitive processes can be observed while reading hypertexts and hyperlinks?
2. Is there another form to write a scientific article than the IMRD model (Introduction – Method – Results – Discussion)?

¹Ware, M. et al., *The STM report: An Overview of Scientific and Scholarly Journal Publishing*, 3rd. ed., International Association of Scientific, Technical and Medical Publishers, The Hague, 2012, available at: http://www.stm-assoc.org/2012_12_11_STM_Report_2012.pdf (08.07.2014).

²To reflect the importance of this topic special conferences are dedicated to this subject like “1st International Workshop on Mining Scientific Publications”. URL: <http://core-project.kmi.open.ac.uk/jcdl2012/> (08.07.2014).

³This study is at the same time my doctoral thesis.

2. Literature review

In the digital age the process of reading and writing has changed; there are just information snippets, fragments interlinked to each other, new forms of texts are created (Blogs, microtexts . . .), new formats are emerging. In 2009, Elsevier launched the “Article of the future”⁴ as a model for embedding source data; early in 2014 Wiley introduced an exciting new format, the “Anywhere Article”⁵ with interactive sidebars which offer easy and seamless access to citations.

On the one hand, you see the integration of supplementary materials to enhance and enrich articles to generate new outputs and to embed scientific texts. Bardi and Manghi [3] define five different forms to enhance publications: reference parts, embedded parts, structured-text parts, generated parts, executable parts, Shotton [16] determines the five stars of online journal articles (peer review, open access, enriched content, available datasets, machine-readable metadata). On the other hand, Nowakowski et al. [14] and Amado Alves [2] describe new methods of composing and decomposing scientific articles. Groth et al. [9] see the future of documents as nanopublications, for De Roure [1] the development of the article lies in the outcome of “research objects” as a collection of stuff (like artefacts, experiments). These scenarios require different distribution models. Priem/Hemminger [15] propose the idea of the “decoupled journal (DcJ)”: “The DcJ brings publishing out of its current seventeenth-century paradigm, and creates a Web-like environment of loosely joined pieces – a marketplace of tools”. Kircz [12] postulates the end of the traditional journal article: “The conclusion so far is that we face a transition after which the traditional journal article will cease to exist”.

There is very little research on the interaction between document structure and the structure of thought [4]. What does this mean for reading, especially for reading texts with hyperlinks? What effects can hyperlinks have on reading and understanding? Are the cognitive processes changing? To say with Nielsen [13]: “Hypertext presents several different options to the readers, and the individual reader determines which of them to follow at the time of reading the text The same is true of footnotes in traditional printed texts, since readers have to determine upon reaching the footnote marker whether to continue reading the primary stream of text or to branch off to pursue the footnote. Therefore hypertext is sometimes called the ‘generalized footnote’”.

3. Method

A suitable method for observing different cognitive processes is eye-tracking. Cole et al. [6] find out that “the cognitive effort due to reading experienced by a user during search may be expected to reflect, in part, their existing domain knowledge”. Further they recognize: “Of particular importance is the fact that eyes fixate until the meaning of the word(s) is acquired” [6]. In addition, this method allows to track the click patterns: “If we can gain access to the gaze data, it is possible to not only analyze their search scanpath and scanning strategy, but to also compare the search behavior to explicit data such as search speed and click patterns” [11]. (Figure 1 shows an illustration of an eye-tracking experiment.)

Eye-tracking can visualize the different passages of a text already known. In this way it should be possible to delete these passages, to deconstruct and re-construct a text, to generate new and shorter forms

⁴Elsevier Introduces Article of the Future Project. URL: <http://newsbreaks.infotoday.com/Digest/Elsevier-Introduces-Article-of-the-Future-Project-55322.asp> (08.07.2014).

⁵The “Anywhere Article” Launched Across Wiley Online Library. URL: <http://eu.wiley.com/WileyCDA/PressRelease/pressReleaseId-110325.html> (08.07.2014).

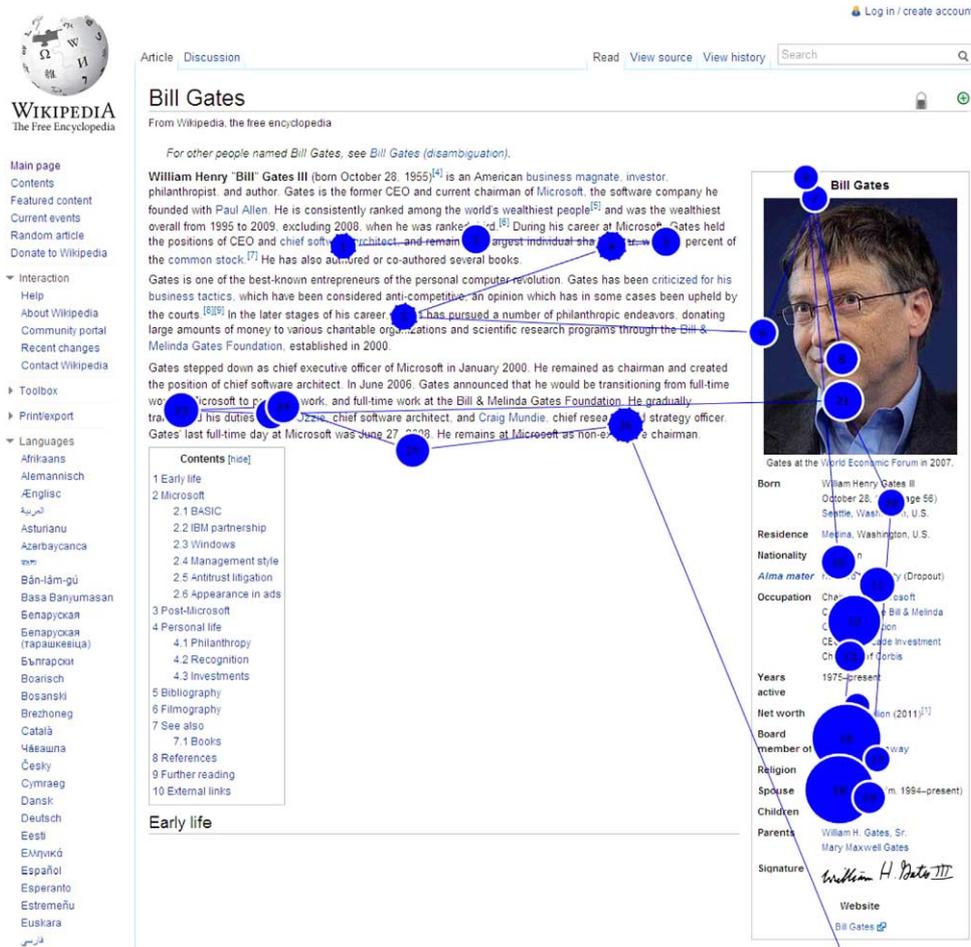


Fig. 1. Snippet of Wikipedia article: Bill gates [sic] biographical article being scanned (gaze plots i.e. the larger it is the longer duration in purple) from Clark et al. [5]. (Colors are visible in the online version of the article; <http://dx.doi.org/10.3233/ISU-140753>.)

of it. In the end you can generate a document which fits the right knowledge level of the researcher, a text which contains only the very essence of the content. How long is the smallest unit for understanding? Could this be a blog or just a tweet?

3.1. Study design

The empirical study that I have designed contains determination of the test persons, which means the number of scientists and their reading patterns, as well as the papers/articles they have read in a defined time scale. Tenopir/Volentine [18] show that there are different reading patterns for each academic discipline and that medical/health scientists have to deal with the largest amount of papers. Therefore this study concentrates on life sciences. To get a sample of scientist's reading and writing behaviors within a month, a questionnaire will be sent to faculty staff. Ferebee and Davis [8] have demonstrated that young people (born after 1980) show quite different attitudes to deal with information (overload). They distinguish between two groups: technologically structured individuals (TSIs) and nontechnologically

structured individuals (NTSIs). The eye-tracking experiment will be administered to each of them to investigate diverse needs.

The number of scientists and papers is still subject to definition.

4. Conclusion and further research

The results of the eye-tracking sample enable to see the scientific article in a different and innovative way: to create an adaptive tool which can recognize a scientist's level of knowledge and to find the smallest unit for writing a scientific article.

With the growing production of data, the data paper is a new type created for collecting and publishing metadata about data [17]. According to this evolving format one could perhaps imagine a sort of citation paper, where you would find uniquely a list of references/citations. So there would exist one paper with the very essence of the content (essence paper) and one citation paper. If a scientist reads first the citation paper and recognizes that most of the references are well known, he or she will decide whether it is worthwhile reading the essence paper or not. Or perhaps these developments would result in articles consisting only of tweetations, to cite Eysenbach: "I call a citation in a tweet (mentioning a journal article URL) a 'tweetation'" [7].

It is a great challenge, however, to leave old pathways and to go beyond boundaries

Further research is needed concerning the granularity of a text: how far can deconstruction go to write an understandable text, can this go to the word level? Given the smart technologies that we have at our disposal, there are a lot of possibilities to develop a new and innovative ecosystem for scientific articles, which perhaps would allow us to reach the status of Kelly's vision: "In the new world of books, every bit informs another; every page reads all the other pages." [10].

References

- [1] K. Allen, The future of scholarly communications: the 'social machine'? Speaking at UKSG, David De Roure surveys the rapidly evolving scholarly communication ecosystem, *Information Today Europe*, April 21, 2014.
- [2] M. Amado Alves, The shattered document approach to adaptive hypertext: Design and evaluation. in: *Mining the Digital Information Networks, Proceedings of the 17th International Conference on Electronic Publishing*, N. Lavesson et al., eds, IOS Press, Amsterdam, 2013, pp. 3–12.
- [3] A. Bardi and P. Manghi, Enhanced publications: Data models and information systems, *Liber Quarterly* **23**(4) (2014), 240–273.
- [4] A.P. Bishop, Document structure and digital libraries: How researchers mobilize information in journal articles, *Information Processing & Management* **35** (1999), 255–279.
- [5] M. Clark, I. Ruthven, P. O'Briant Holt and D. Song, Looking for genre: The use of structural features during search tasks with wikipedia, in: *Proceedings of the 4th Information Interaction in Context Symposium*, ACM, New York, NY, USA, 2012, pp. 145–154.
- [6] M.J. Cole, J. Gwizdka, C. Liu, N.J. Belkin and X. Zhang, Inferring user knowledge level from eye movement patterns, *Information Processing & Management* **49** (2013), 1075–1091.
- [7] G. Eysenbach, Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact, *Journal of Medical Internet Research* **13** (2011), 4, e123.
- [8] S. Ferebee and J.W. Davis, Persuading library use in technologically structured individuals, *Library Hi Tech* **29**(4) (2011), 586–604.
- [9] P. Groth, A. Gibson and J. Velterop, The anatomy of a nanopublication, *Information Services & Use* **30** (2010), 51–56.
- [10] K. Kelly, Scan this book! *The New York Times Magazine*, May 14, 2006.
- [11] J. Kim, P. Thomas, R. Sankaranarayana, T. Gedeon and H.-J. Yoon, Eye-tracking analysis of user behavior and performance in web search on large and small screens, *Journal of the Association for Information Science and Technology* **2014** (2014), doi: 10.1002/asi.23187.

- [12] J.G. Kircz, New practices for electronic publishing 2: New forms of the scientific paper, *Learned Publishing* **15**(1) (2002), 27–32.
- [13] J. Nielsen, *Multimedia and Hypertext: the Internet and Beyond*, AP Professional, Boston, 1995.
- [14] P. Nowakowski, E. Ciepela, D. Harężlak, J. Kocot, M. Kasztelnik, T. Bartyński, J. Meizner, G. Dyk and M. Malawski, The collage authoring environment, *Procedia Computer Science* **4** (2011), 608–617.
- [15] J. Priem and B.M. Hemminger, Decoupling the scholarly journal, *Frontiers in Computational Neuroscience* **6** (2012), 1–13.
- [16] D. Shotton, The five stars of online journal articles – A framework for article evaluation, *DLib Magazine* **18** (2012), 1/2.
- [17] Y.M. Socha (ed.), Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data, *Data Science Journal* **12** (2013), 1–75.
- [18] C. Tenopir and R. Volentine, UK scholarly reading and the value of library resources: Summary results of the study conducted, Spring 2011, Center for Information and Communication Studies, University of Tennessee, USA, 2012.