Contribution Details

1033

Finding visual patterns in information Search Stories

Luca Botturi[1] ✉ ⓘ, **Loredana Addimando**[1] ✉ ⓘ, **Silvia Giordano**[1] ✉ ⓘ, **Martin Hermida**[2] ✉, **Luca Luceri**[1] ✉, **Azza Bouleimen**[1] ✉

Organisation(s): 1: Scuola universitaria professionale della Svizzera italiana, Switzerland; 2: Pädagogische Hochschule Schwyz

Submitted by: Prof. Luca Botturi (Scuola universitaria professionale della Svizzera italiana)

Presenting Author: Botturi, Luca; Addimando, Loredana (luca.botturi@supsi.ch loredana.addimando@supsi.ch)

Format of Presentation: Paper

Proposal Information

We live today in an overwhelming and constantly growing sea of information. While information - from news to recipes to book reviews to health data - is not a scarce good any more, assessing its quality and reliability has become a pressing issue, as the pandemic and related infodemic have taught (Zarocostas, 2020). Searching and evaluating information is a core life activity, and it is fundamental in a democratic society (White, 2016). Information literacy (IL) is included among the century critical skills (Ananiadou & Claro, 2009), and key to pursue effective life-long learning (Kurbanoglu, 2012).

IL is also recognised as more and more central for today's education, and is included in most digital and media literacy models as an essential dimension, (e.g., DigComp 2.1; Carretero, Vuorikari, & Punie, 2017). The development of sound and effective IL education requires a clear and possibly evidence-based understanding of how we - and young people in particular - identify, access and select information in everyday life, which mostly happens online, on personal devices and in relation to everyday life needs. While existing IL models seem to converge on the key steps in searching information (e.g., SCONUL, 2011; Big 6, n.d.) and provide consistent prescriptive models, we still know very little about how young people actually search online, i.e., we have a shallow understanding of our learners' initial search expertise. The technology landscape is also undergoing continuous transformation: search engines have grown easier to use and at the same time more opaque, implementing AI algorithms and user profiling, and social networks are rapidly becoming popular access points to the web, with very questionable results - all of which is demanding more complex skills.

Research on IL so far has been mostly based on the (self-)assessment of IL skills (e.g., in the ICILS project; Fraillon et al., 2020), on measuring IL self-efficacy (Kurbanoglu, Akkoyunlu & Umay, 2006) and on capturing the online search behavior with monitoring tools, such as URL timestamping (Gwizdka & Spence, 2006) or eye-tracking (Jiang, 2014), often in relation to academic or job tasks. While all these approaches delivered and still deliver very useful results and insights, we developed a novel approach based on the collection and analysis of search stories, and designed to preserve the ecology of data collection in a complex and diversified environment (Botturi et al., 2021).

In this paper, a follow-up to our 2021 ECER contribution, we present the collection, visualization and algorithmic inspection of search stories as an innovative method to investigate online information search practices.

A search story is the collection of all the actions a user performed while solving an information search task. Actions are organized in search episodes, which determine the story structure. Actions are also tagged according to different features, related both to formal features (e.g., duration) and content features (ee.g., the web domain being visited). The resulting data structure can be then analyzed in two ways:

1. Through descriptive statistics and clustering, identifying similarities and variations across stories and drawing correlations with task and user profiles.
2. Through multiple static and interactive visualizations that allow researchers to inspect them and compare them with each other, eventually discovering visual patterns. Pattern combinations and frequencies were then analyzed thanks to a Machine-Learning algorithm.

Results so far allowed to identify several relevant patterns at different levels, and to start exploring their relationship with user profiles, task features and the quality of task solutions.

Methodology or Methods/ Research Instruments or Sources Used

The search story method was developed within the LOIS project. 220 participants in the age group 16-20 (selected on voluntary basis) were asked to install an ad hoc extension in their browser at home, and then to solve 4 pre-defined information search tasks. Tasks were both open and close and on different themes. In Spring 2021, 152 participants completed the study, and we could generate a dataset of 597 search stories. An additional benchmark set of 21 search stories, based on the same tasks, were collected from 6 expert searchers (librarians, researchers, journalists).

The participants' navigation actions, recorded and time-coded by the browser extension, were enriched with both manual and automatic metadata. Actions were tagged according to content (e.g., if they were search actions or result actions, or based on a classification of web domains) and duration (based on thresholds taken from the literature). Aggregated metadata describing each story were calculated. The stories were visualized into plots of different types:

In box plots each action is represented by a colored square. The color can be used to indicate the type of action (e.g., a SER page, or a SER page being re-visited), the class of domain (e.g., newspaper, research database, social network), or its duration (e.g., suggesting if the user is reading of just scanning a SER page)

Bar plots are like box plots, but each action is a rectangle, its width being proportional to the action's duration.

Moreover, story plots have two formats: graphic plots are simple PNG images, while Interactive plots are interactive HTML pages that allow the exploration of queries and URLs.

Recent studies widely reviewed visual techniques to support the exploratory analysis of temporal graph data (Erten, et al 2006). Kerracher et al.'s task taxonomy (2015), for example, considers the range of possible tasks involved in exploring graph data. Visualizations allow researchers to inspect the huge amount of collected data and to identify visual patterns at different levels, for example: sequential patterns occurring in a single episode (e.g., search and then refine query); patterns involving multiple episodes (e.g., reviewing key results at the end of the search story); or complex patterns (e.g., regular variations in the time spent on SER pages).

The hypotheses drawn visually were then formalized and fed into a machine learning algorithm to check consistency, combinations and

variations across the dataset. The results were then fed back into the statistical analysis.

Conclusions, Expected Outcomes or Findings

One of the challenges of the LOIS project was to make sense and formulate hypotheses in a large dataset in a new and unconventional format. The visual exploration of search stories provided extremely interesting insights, which opened up intriguing research venues, pursuing some educationally relevant issues, for example:
Are there different information search styles? How can they be described?
To what extent does the nature of the task influence a user's search style?
Are there successful search patterns?
In general, the analysis of search stories reveal online information search more as a subtle art than an exact science, guided by principles more than rules. Such insights might support the development of a more responsive and flexible IL teaching approach.
Also, visualizations themselves can be used as instructional materials, as they visualize the search process, offering an opportunity to shift the focus of IL instruction from the product to the process.

References

Ananiadou, K., and Claro, M. (2009). 21st Century Skills and Competences for New Millennium Learners in OECD Countries. OECD Education Working Papers, 41, OECD Publishing. DOI: 10.1787/218525261154
Big 6 (n.d.). The Big Six. https://thebig6.org
Botturi, L., Hermida, M., Cardoso, F., Galloni, M., Luceri, L. & Giordano, S. (2021). Search stories. Towards an ecological approach to explore online search behaviour. Paper presented at ECER 2021 (online).
Carretero, S., Vuorikari, R. and Punie, Y. (2017). DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use, EUR 28558 EN, DOI:10.2760/38842
Erten, C., Kobourov, S.G., Le, V., and Navabi, A.(2006). Simultaneous graph drawing: Layout algorithms and visualization schemes. In: Symposium on Graph Drawing, 437– 449.
Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2020). Preparing for Life in a Digital World. IEA International Computer and Information Literacy Study 2018 International Report. Springer Open.
Gwizdka, J., & Spence, I. (2006). What can searching behavior tell us about the difficulty of information tasks. A study of Web navigation. Proceedings of the American Society for Information Science and Technology, 43(1), 1-22.
Jiang, T. (2014). A clickstream data analysis of users' information seeking modes in social tagging systems. In Proceedings of the 9th iConference, 314–328.
Kerracher N., Kennedy J. , Chalmers K, and Graham M. (2015). Visual Techniques to Support Exploratory Analysis of Temporal Graph Data. Eurographics Conference on Visualization (EuroVision). http://dx.doi.org/10.2312/eurovisshort.20151133
Kurbanoglu, S. (2012). An analysis of the concept of information literacy. Proceedings of the International Conference of the Media and Information Literacy for Knowledge Society (June 24-28), 1-42.
Kurbanoglu, S., Akkoyunlu, B., & Umay, A. (2006). Developing the information literacy self-efficacy scale. Journal of documentation, 62(6), 730-743.
SCONUL (1999). Information Skills in Higher Education: A SCONUL Position Paper. London: Society of College, National and University Libraries.
White, R. W. (2016). Interactions with Search Systems. Cambridge University Press.
Zarocostas, J. (2020). How to fight an infodemic. The Lancet, 395(10225), 676.

Remarks / Messages

Remark/Message from the Authors:
Additional co-authors to this research are Marzia Galloni (SUPSI-DFA), Chiara Beretta (SUPSI-DFA) and Felipe Cardoso (SUPSI-DTI). Could they also be added?