Understanding Online Information Search Practices with Search Stories

Abstract (120 words)

The pursuit of truth in an overwhelming sea of online information is a contemporary challenge. Information Literacy (IL) is a key component of digital and media literacy and has been included in school programs all over the world. Although there are several prescriptive models for teaching IL, we know little about how young people search for information online, and about what actual learning paths exist in this domain. This paper presents an innovative and ecological data collection approach, which allowed to explore and understand young people's online search practices. Analyses conducted on a dataset of 595 search stories led to the identification of different search styles and of new avenues for effective and flexible reflective IL education.

Main text (2000 words)

Purpose: Information Literacy education as a key challenge in the digital society

In globally interconnected world information is an abundant, when not overwhelming, commodity. Our experience of the world is more often mediated than direct (Han, 2022), so that assessing the quality and reliability of the information we find or receive has become urgent, as the pandemic and related infodemic (Zarocostas, 2020) have taught us and as we continue to see every day in relation to local, national and international events.

Knowing how to find, select and share good information is crucial in a democratic society (White, 2016) and in our personal pursuit of truth, so that Information Literacy (IL) appears as a key competence (Ananiadou & Claro, 2009; Kurbanoglu, 2012). IL is integrated into most digital and media literacy models (Hobbs, 2010; Carretero, Vuorikari & Punie, 2017) as both an academic skill and for implementing effective life-long learning policies (Kurbanoglu, 2012).

Teaching IL is a challenging endeavor: it is a complex domain in which a broad set of skills and cognitive abilities interact, ranging from critical thinking to bookmarking tools. Moreover, it is confronted with an overwhelming and continuously growing mass of unverified and easily accessible digital information, which is changing the rules of the game and has made fact-checking a requirement (Brodsky et al., 2021). In addition, the technological landscape is constantly evolving. Thanks to the implementation of artificial intelligence algorithms and user profiling systems, search engines have become easier to use and at the same time opaquer, and social networks are rapidly becoming popular access points to the Web – and all of this requires more complex skills.

The development of effective IL programs requires a clear and possibly evidence-based understanding of how we – and young people – identify, access, and select information, which happens primarily online, on personal devices, and in relation to the various noble and trivial demands of daily life.

Theoretical framework: IL research and education models

IL research to date has been mainly based on (self-)assessment (Fraillon at al., 2002), self-efficacy (Kurbanoglu, Akkoyunlu & Umay, 2006), and analysis of online search behavior with monitoring tools such as URL timestamping (Gwizdka & Spence, 2006), eye-tracking (Aula, Majaranta & Räihä, 2005), or with think-aloud protocols (Brand-Gruwel, Wopereis & Walraven, 2009). Most research study focus on academic or job-related tasks.

These approaches still provide useful results but suffer from at least two major limitations. The former is the ecology of data collection: while searching online information is a daily task that happens as much on the public transport as in home and offices, these methods capture lab-conditioned behaviors in relation to primarily cognitively demanding academic tasks. The latter is validity: information skills are difficult to observe and evaluate, and they are often measured

through self-report instruments, which do not necessarily correspond to the actual effectiveness and behaviors (Mahmood, 2016).

On the educational side, we basically find two classes of instructional tools. On the one hand, prescriptive models that illustrate an ideal search process, structured in sequential or looping steps and based on the identification of a definite information need. Models like the 7 Pillars (SCONUL, 1999; SCONUL, 2011) or the Big 6 (Big6, n.d.; Wolf, Brush & Saye, 2003) fall in this category. The IPS-I model (Brand-Gruwel, Wopereis & Walraven, 2009) provides a further development, consolidating the key process element while emphasizing reflection-in-action and self-regulation along with conditional skills (reading, evaluating and computing skills).

On the other hand, there are models that provide guidance to identify relevant cues to assess the quality of documents and sources to support the selection process. Models in this category are based either on internal document features (vertical reading), for example CRAAP (Blakeslee, 2004), RADAR (Mandalios, 2013), or on finding information about the source (lateral reading), for example SIFT (Caulfield, n.d.)

However, we still know very little about how young people actually conduct online searches, i.e., we have a superficial understanding of our students' initial competencies so that, even with the support of reference models as the one just mentioned, it is difficult to identify learners' zones of proximal development (Levykh, 2008) and to differentiate learning, so to elude a simplistic one-size-fits-all approach.

Design and data collection: protecting the ecological framework of participants' search practices

In this paper we present the collection, visualization, and analysis of search stories as an innovative method for investigating online information research practices and to enhance IL education.

In the spring of 2021, 220 volunteer participants aged 16 to 20 years in Switzerland and Italy installed an ad hoc extension in their home browsers and agreed to solve four information-seeking tasks. The tasks were both open-ended and closed-ended and covered different non-academic subject areas and involved different cognitive processes, as in the following example:

> *Your friend Anna is worried because her younger sister has decided to become a vegan. Anna thinks that this could be a good choice from an ethical point of view, both for her body and the environment – but her sister is only 13 years old... Isn't it too soon? Besides, her sister already suffers from asthma and this health condition makes her weaker.*
>
> *Opinions are divided: some argue that vegan food is healthy, others do not. What is the truth? Would a vegan diet be healthy for a 13-year-old girl? What advice would you give Anna? And on what would you base it?*

Participants were also invited to complete a profiling questionnaire and a self-assessment of their digital and information skills. Once they had solved the tasks, they completed a survey with the Big5 scales (Hofstee, De Raad & Goldberg, 1992).

The whole idea behind this data collection process was to preserve the ecology of data collection, i.e., leaving participants outside research labs and have them interaction with search engines and online information on their own devices, whenever they want and in their own settings – be it at home, in the office or anywhere else (Author, 2021).

Overall, 152 participants completed the data collection process (Figure 1), and we were able to generate a set of 595 valid search stories. An additional benchmark set of 21 stories, based on the same tasks, was collected from 6 expert searchers (two librarians, two researchers, and two journalists).

Participants' browsing actions, recorded by the browser extension, were enriched with manual and automatic metadata based on content (e.g., whether the actions were on search engines pages or on other websites) and duration. Aggregate metadata describing globally each story and each user's behavior on different tasks were computed.

[FIGURE 1]

Exploring search stories

A search story is basically a data structure that captures the interaction of a user with the web while solving an information task.

Search stories are made by actions. An action is defined by an URL and a timestamp. Depending on the URL, each action can be classified as a *search action* (domain is a search engine) or *result action* (domain is another website). Depending on its duration and based on research-based thresholds, it is possible to infer the type of document processing, e.g., if a user was bouncing, skimming, scanning or reading a Search Engine Result (SER) page. Additional tagging was used to classify the visited websites (including e.g., social networks, online magazines, research websites, etc.) and to mark if the domain appears for the first time in the story or if it is a re-visit.

Actions within a search story are organized into search episodes. An episode always begins with a search action which comes after a non-search action. The number of episodes on the story indicates its *extension* and the average duration of the episodes its *intension.*

Finally, search stories can be visualized plotting both the sequence of actions and their duration. Colors were used to identify search and result actions, or domain types (Author, 2022a), as illustrated in Figure 2.

[FIGURE 2]

The large nonstandard database thus generated lends itself to a variety of lines of analyses that are focused on both visible behaviors during the search process and the interpretations assigned by the participants to the searches conducted.

This paper focuses on two key questions: is it possible to identify a particular research style in the way participants searched for information on the assigned tasks? And if there are styles, they be ranked in terms of effectiveness and efficiency?

To shed light on such questions, we identified different types of search stories, through a kind of “formal anatomy”. Considered dimensions were duration, intension/extension, use of queries, visits and re-visits to domains, type of document processing. This analysis was conducted using clustering and correlation analysis (Author, 2022c), as illustrated in Figure 3.

[FIGURE 3]

Results

The first results of the analysis of search stories clearly indicates that while most participants performed short searches with few queries and actions, more search styles exist. Also, some users seem to demonstrate a consistent search style across different tasks, but others adapt their search to the task and situation. The perceived importance of the topic at hand and how much the user knows (or thinks to know) about it also influence the search practice.

While no unique search style seems to lead to better results or to more efficient searches, the use of queries seems to be a reliable indicator of the quality of a web search: users that use different search queries are able to provide better solutions to information tasks than those doing one or few searches or repeating the same query string multiple times.

Overall, searching for information on the web appears more as a subtle art than an exact science, guided by principles rather than rules, and different people go about it in rather different ways.

Implications for Information Literacy education and beyond

Implications for IL education are far-reaching. The evidence collected provides a better understanding of how young people search online and allows the identification of different learning journeys towards mastering IL competences. While traditional IL models set a compass, reflecting on search stories can promote the development of instructional approaches that identify and consider individual learners' skills and inclinations. For example, some learners might focus on how

they read SER pages, while others on the documents they select; some might benefit from learning to refine their query strings, while others from saving, revisiting and comparing documents.

The technology developed for capturing and visualizing search stories could also be used to develop innovative learning tools that empower reflective IL education, implementing a relational frame approach (Bruce, Edwards & Lupton, 2006). One of the limits in IL education is that it always focused on results and products (e.g., a selection of sources or written report) because the search process remained largely invisible. Search stories plots in fact make it visible allowing self-reflection, comparisons and peer-learning (Author, 2022).

The consequences reach out to teacher education, too. Preparing teachers for IL education should not be limited to learning and practicing IL models, possibly in a specific subject-matter area, but should include observing students to understand their individual practices, and scaffold them accordingly. Also, the definition of search tasks appears to be crucial. For example, people search differently and more effectively on topics they already know – so researching a familiar domain might be a good first online search task.

The implications of this work extend also beyond IL education. The data used to generate search stories are just a subset of the information regularly available for all user activities to search engine providers. Industry players might identify online search patterns and styles and might provide advice to their users to enhance their search practices, or even adapt their interface accordingly.

Finally, more research needs to be done in this domain, including comparing the search practices of different user groups (e.g., elderly people, or people with a diverse cultural background) or analyzing search practices across different device types, or different tasks. Shedding light on how we search and process information and finding solution to empower people to access high-quality information remains a challenge of paramount importance in our societies.

References

Zarocostas, J. (2020). How to fight an infodemic. *The Lancet, 395*(10225), 676.

Han, B. C. (2022). *Non-things: Upheaval in the lifeworld*. John Wiley & Sons.

White, R. W. (2016). *Interactions with Search Systems*. Cambridge University Press.

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

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*, UNESCO, Moscow, 1-42.

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

Hobbs, R. (2010). *Digital and Media Literacy. A Plan of Action.* Aspen Institute/Knight Foundation.

Brodsky, J. E., Brooks, P. J., Scimeca, D., Todorova, R., Galati, P., Batson, M., Grosso, R., Matthews, M., Miller, V., & Caulfield, M. (2021), *Improving college students' fact-checking strategies through lateral reading instruction in a general education civics course.* Cognitive research: principles and implications, 6(1), 23. https://doi.org/10.1186/s41235-021-00291-4

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.

Kurbanoglu, S., Akkoyunlu, B., & Umay, A. (2006). Developing the information literacy self-efficacy scale. *Journal of documentation, 62*(6), 730-743.

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.

Aula, A., Majaranta, P., Räihä, K. J. (2005). Eye-tracking reveals the personal styles for search result evaluation. In Costabile, M. F., Paternò, F. (eds.) *Proceedings of the IFIP conference on human-computer interaction 2005*, pp. 1058-1061, Springer, Berlin, Heidelberg.

Brand-Gruwel, S, Wopereis, I, & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education, 53*(4), 1207-1217, https://doi.org/10.1016/j.compedu.2009.06.004.

Mahmood, K. (2016). Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Communications in Information Literacy*, *10*(2), 3.

Wolf, Brush & Saye (2003). The Big Six Information Skills As a Metacognitive Scaffold: A Case Study. *School Library Media Research, 6,* 1-24.

SCONUL (1999). *Information skills in higher education: a SCONUL position paper.* SCONUL Working Group on Information Literacy.

SCONUL (2011). *The SCONUL Seven Pillars of Information Literacy.* SCONUL Working Group on Information Literacy.

Big 6 (n.d.). The Big Six. https://thebig6.org

Blakeslee, S. (2004). The CRAAP Test. *LOEX Quarterly*, *31*(3).

Caulfield, M. (n.d.). *Web Literacy for Students Fact-Checkers*. https://webliteracy.pressbooks.com/

Mandalios, J. (2013). RADAR: An approach for helping students evaluate Internet sources. *Journal of information science*, *39*(4), 470-478.

Levykh, M. G. (2008). The affective establishment and maintenance of Vygotsky's zone of proximal development. *Educational theory*, *58*(1), 83-101.

Author (2021). Search stories. Towards an ecological approach to explore online search behaviour. Paper presented at *ECER 2021* (online).

Hofstee, W. K., De Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of personality and social psychology*, *63*(1), 146.

Author (2022a). Online search stories visualizations for enhancing self-reflective Information Literacy practices. *Georgia International Conference on Information Literacy* (March 31-April 1).

Author (2022b). Ricercare informazioni e costruire opinioni: un'analisi sulle modalità di ricerca dei giovani in rete*. 14° Giornate Nazionali di Psicologia Positiva*, Università Federico II, Napoli, 10-11 giugno 2022.

Bruce, C., Edwards, S., & Lupton, M. (2006). Six Frames for Information literacy Education: a conceptual framework for interpreting the relationships between theory and practice. *Innovation in Teaching and Learning in Information and Computer Sciences*, *5*(1), 1-18. https://www.tandfonline.com/doi/full/10.11120/ital.2006.05010002

Author (2022c). Understanding Information Literacy with Search Stories. *SSRE 2022*, Lausanne.

Figures

[FIGURE 1]

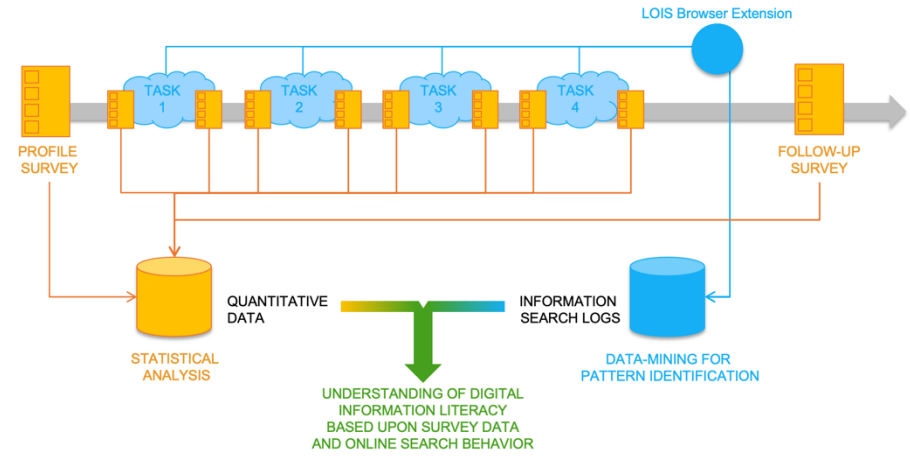

Fig 1. The data collection system of research histories in the LOIS project.

[FIGURE 2]

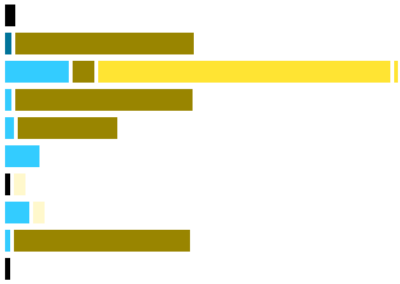

Fig. 2. Visualization of a search story. Each segment represents an action, where its length is proportional to duration. The blue segments represent actions on search engines (dark blue: new query; light blue: repeated query), the ochre segments on website pages (ochre: new website; yellow: re-visit to the same domain). Black segments are system actions.

[FIGURE 3]

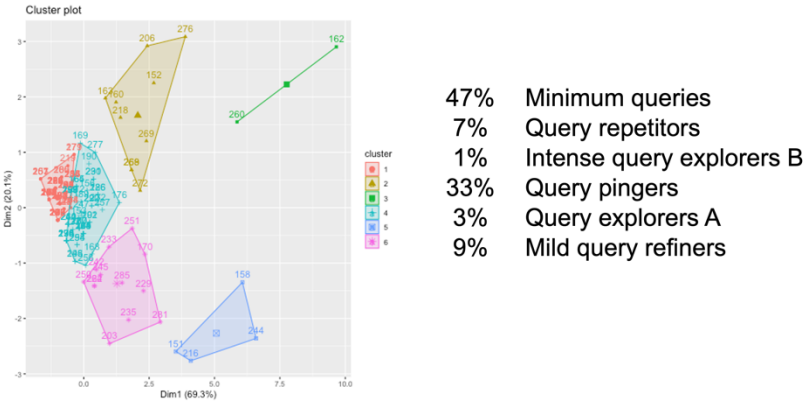

Fig. 3. An example of clustering on the query dimension and a specific task. On the right, descriptors of the cluster features are indicated, along with the percentage of individuals that belong to each cluster