2017-06-12

Reflections on an Experiment, Evaluating the Impact of Spatialisation on Exploration

Clement Roux
Technological University Dublin, clemw198@hotmail.com

John McAuley
Technological University Dublin, john.mcauley@tudublin.ie

Eurographics Conference on Visualization (EuroVis) 2017
B. Kozlíková, T. Schreck, and T. Wischgoll
(Guest Editors)

Short Paper

Reflections on an experiment, evaluating the impact of spatialisation on exploration

C. Roux[1] and Dr. J. McAuley[1]

[1]Dublin Institute of Technology, Dublin, Ireland

Abstract

This paper reports on an experiment designed to evaluate whether visualising a digital library (using a spatialisation technique) can influence exploratory search behaviour. In the experiment we asked participants to complete a set of novel tasks using one of two interfaces - a visualisation interface, ExploViz, and its search-based equivalent, LibSearch. A set of measures were used to capture sensemaking and exploratory behaviour and to analyse cognitive load. As results were non-significant, we reflect upon the design of the experiment, consider possible issues and suggest how these could be addressed in future iterations.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [User Interfaces]: Evaluation/methodology—Graphical user interfaces (GUI)

1. Introduction

Search is a common activity for digital users today. An average user could interact with a wide range of search interfaces at any one time, from searching the Internet with Google, Bing or Baidu, to searching for movies with Netflix or Hulu, to searching for music with Spotify, Deezer or Apple, to searching for books on Amazon or EBay. However, the basic approach remains the same - a user inputs a set of keywords which returns a ranked list of items. This has proven effective, once the user can describe their needs succintly, however, if the user is unable to do so, the keyword-based approach can prove much less successful. In these cases, the user may want to browse the archive or explore areas of interest without knowing or wishing to express this interest a priori. Of course, visualisation, one would imagine, could support this exploration process by enabling the user to see connections and browse similar items while, at the same time, maintaining a local and global view of the data.

We present an experiment in which we sought to evaluate this proposition by examining the explorative affordances of visualisation or more precisely spatialisation - a popular technique used to cluster and present data based on item similarity. From the literature, we developed a visualisation system (ExploViz) using the state-of-the-art spatialisation design and an equivalent library interface (LibSearch) that mimics standard search interface design. We then evaluated the explorative affordances of ExploViz against LibSearch using a between group experimental design. We hypothesized that the visualisation system would encourage participants to explore more of the library however we found that both systems performed equally well. We discuss and critical appraise the design of our experiment and suggest how to improve upon potential design flaws. Although the results were not positive, we consider the approach - system, experiment and critical analysis - a contribution to the visualisation community, as it allows other researchers to avoid similar pitfalls and to ensure more robust and reproducible results. The paper is divided between a description of the system, using Muzner's nester design topology, an outline of the evaluation methodology and a critical analysis of the experimental design.

2. The Design of ExploViz

The ExploViz interface design follows Munzner's [BM13, Mun09] four-level nested model supporting the creation of different, often bespoke, visualisation systems. The methodology divides the design of a visualisation system into a number of nested levels, each layered on top of the last, and allows for the incorporation of feedback at different points during the design process. In the following section, we describe the design of the ExploViz interface, beginning with a problem, data and task abstraction and concluding with the encoding and interaction.

2.1. Problem, data and task abstraction

Problem. There is an argument in the literature on exploratory search that exploration is often under-supported with existing search systems [ABLS15,BISM14,Gar06,HK15,WF12,WKSS10]. A large majority of existing techniques rely on a search and browse interface with ranked results and similarity-based listings, which restricts or "filters" the user's access to content based on their previous activity [NHH*14, Par12]. However, exploratory search, by its very nature, seeks to encourage explorative behaviour.

Data. The data, drawn from the Cyberlibris digital library, includes a corpus of business and finance books. We focused exclusively on the library's metadata (titles, authors, abstracts, categories and tables of content), which were parsed into word vectors [BNJ03] for the spatialisation process. We then applied Latent Dirichlet Allocation (LDA) on the document-term matrix from which the Euclidean distance between each pair of documents was computed. Next we used multidimensional scaling (MDS) to reduce the distance matrix into a two-dimensional space. MDS was chosen because of its ability to highlight high-level topics as three distinct clusters.

Task Abstraction. Munzner et al.'s original task description considers not only why a task is performed but also reflects on the level of specificity (high, mid, low) at which a task is performed [BM13]. The high-level aim of ExploViz system is to enable a user to browse, discover, access and consume information. The mid-level aim of ExploViz focuses on exploration, a process of information-seeking, in which neither the location nor the target are known a priori. Finally, the low-level aim of ExploViz is to enable the user to identify one or several books in the digital library. Possibly, the user can also compare these books and summarize information about the library's content.

2.2. Encoding and Interaction

The resulting spatialisation of business and finance 1363 books is represented in Figure 1. The process results in a spatial representation where each book is positioned by a point. The closer two books are in position, the more similar their content. In addition, three blue transparent convex-hulls, each representing the entire information space of one high-level topic or class, are used to capture the topics entire information space. In this image, the three high-level topics are business, finance and economics.

In keeping with the nested model, the point cloud and three convex-hulls can be defined as marks encoding each book and topic-level classification. In addition, three channels are used, namely the position on a common scale, the spatial region and the hue which are the three most effective visual channels as reviewed in [MM14].

The ExploViz spatialisation allows users to select books by clicking on the corresponding point and a user can navigate through the library's infromation space using semantic zoom. The semantic zoom allows the user to increase the scale of the visualisation which reflects the level of specificity - from overview to detail. As

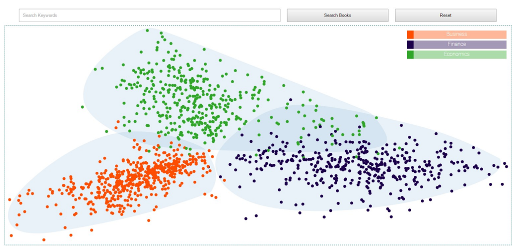

Figure 1: *The ExploViz visual search interface without zoom.*

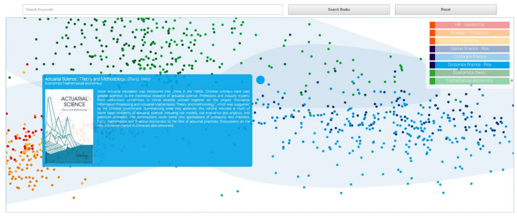

Figure 2: *The ExploViz in zoom mode, focusing on a book.*

can be seen in Figure 1, the high-level classification is encoded in the overview mode, whereas the low-level classification is encoded in the detail mode in Figure 2. The legend, which is situated in the upper right corner, adapts automatically. When the encoding changes, the convex-hulls remain unchanged so that the user is provided with an anchoring point within the library's infromation space. This ensures the user can identify the topic-level class in which they are currently navigating and reduces the potential for the user to become lost within the visualsiation.

Even when in detailed view, further information is required, such as a book's specific details. This is why an embedded card, containing the books' title, author and abstract, is provided when the user hovers over a book's glyph, as illustrated in Figure 2. The card's colour corresponds to either the high or the low-level classification depending on the level of specificity or zoom.

Additionally, a keyword search bar was provided at the top of the visualisation's frame. Similar to traditional digital library search tools, the search highlights relevant books within the spatialisation. However, the user's working context remains unmodified. Instead, it reduces the opacity of the points which are outside the search results in order to make the books of interest more visible. It is a focus + context tool and it differs from the traditional use of a search bar which provides a limited space of results.

Figure 3 shows how the four-level nested model was applied during the design of the ExploViz search interface. It shows a "chained sequence of what - why - how analysis" inspired from the Glimmer visualisation tool which aims at discovering clusters of similar documents [MM14]. The ExploViz interface also allows for navigation in the library's metadata space, contrary to Glimmer which focuses on the resulting clusters from dimensionality reduction. This sequence of what - why - how can be considered as a design justification for ExploViz in the form of a design study.

3. Evaluation - ExploViz vs LibSearch

We conducted an online experiment to contrast the ability of Exploviz to encourage explorative search behaviour with a standard library interface called LibSearch, which we developed specifically for the experiment. LibSearch mimics the Amazon-type library interface commonly used in digital libraries (Figure 4). LibSearch enables the user to search for books and browse similar books (using the same similarity-based approach as used in ExploViz).

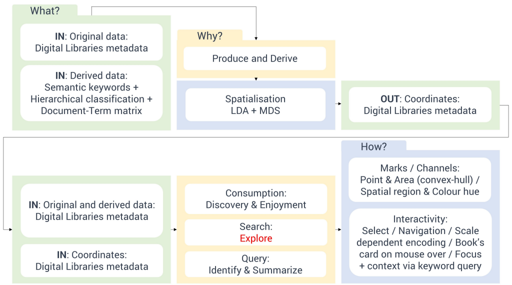

Figure 3: *The four-level nested model [Mun09] and the multi-level typology [BM13] adapted to the ExploViz visual search interface.*

3.1. Experimental design

To evaluate ExploViz, we adopted a between group experimental design to counter potential learning effect. Each participant in the study was provided with one interface, either ExploViz or LibSearch, and asked to complete a set of three exploratory search tasks. The difference between each group was measured with a set of quantitative metrics that measure exploration and sensemaking, while cognitive load was measured using an adapted Nasa-TLK as described in Liu et al. [LBF*13].

Data. We randomly selected a corpus of books from the library that focus on the subject of Business and Finance.

Participants. We recruited 61 participants from college mailing lists and online communities dedicated to data analytics.

Tasks and Metrics. Exploratory search is difficult to quantitatively evaluate. By its very nature, exploration relies on innate motivations such as an individual's curiosity or the relationship the user has to a particular dataset [KC09, WF12]. We devised a set of information-seeking tasks based on prior studies. It was essential for the integrity of the experiment that each task was achievable and manageable with each interface - ExploViz and LibSearch. For each task, we collected hover events (triggered at a 500 millisecond iunterval) and derived two metrics from the collected events - the number of hovered books and the area explored. Because each hover event has a position in Euclidean space, we can measure the total area explored within the information space by summing the distance between each hover event. Our approach is based on the assumption that each hover event with a threshold of 500 milliseconds is equivalent between both interfaces. In ExploViz, hovering over a book's glyph will trigger a hover event while in LibSearch, hovering over a books' title, covers or abstract, as displayed in the list-view, will also trigger a hover event. We also measured extraneous and germane cognitive load for each task using an adapted version of NASA TLX. The extraneous cognitive load explains the users' mental load allocated to understanding the search interface while the germane cognitive load explains the users' mental load allocated to learning new knowledge. We hypothesised that ExploViz would reduce extraneous load while fostering germane load.

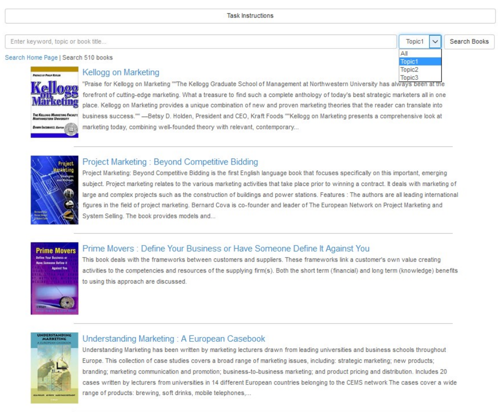

Figure 4: *The LibSearch interface*

Task 1 - Sensemaking Exploration. In this task the participant is asked to develop an appreciation of the information space by exploring the documents' metadata. The participant is then asked to enter a set of five representative keywords for each of the high-level topics in the library.

Task 2 - Directed Exploration. The second task is a open-ended scenario-based task that asks the participant to select two books whereby it is suggested that the participant is preparing for a new job as a manager of a financial instruments team in a multinational bank.

Task 3 - Free Exploration. The third task is a free form exploratory search task in which participants were asked to select a single book of their choice.

4. Results

In the remainder of the paper, we report the results of the experiment and critically analyse the approach. Table 5 summarises all p-values resulting from the Wilcoxon rank sum and student tests used to compare the ExploViz and LibSearch distributions for each task. The asterisk outlines the significant difference supporting each hypothesis, with a confidence level of 5%. The fields without asterisk express a non-significant difference in the partcipants' behaviour or cognitive load. The number of hover events and the area explored measured in the tasks 1, 2 and 3 are not significantly different between both groups. Similarly, the distance between the two selected books in task 2 and the exploration time in task 3 are not significantly different. Both sensemaking accuracy measures for topic 1, 2 and 3 captured in task 1 are not significantly different either.

However, the first task's results show a significantly higher level of germane cognitive load which expresses the participants' perception of their success and satisfaction while performing task 1. Moreover, the higher the germane cognitive load the more supportive an exploratory search tool as argued by [LBF*13]. The cognitive load measured in tasks 1, 2 and 3 are not significantly different.

Most results show a non-significant difference in exploratory search behaviour, cognitive load and sensemaking accuracy between both groups. In other words, there is no strong evidence to identify whether visualising a digital library using a spatilisation technique impacts exploratory search behaviour when compared to a traditional search and ranked-list interface.

Metrics	Ta(sk) or To(pic)	Wilcoxon or Student results
Number of hover events	Ta 1	$W = 431.5$, $p = .6641$, $Mdn = 64.5, 69$
	Ta 2	$W = 341$, $p = .08104$, $Mdn = 15.5, 36$
	Ta 3	$W = 321.5$, $p = .1549$, $Mdn = 11, 19$
Area explored	Ta 1	$W = 3278.5$, $p = .07646$, $Mdn = .0058, .0067$
	Ta 2	$W = 2351$, $p = .7749$, $Mdn = .0018, .0026$
	Ta 3	$W = 1862$, $p = .9525$, $Mdn = .0015, .0010$
Distance btw 2 books	Ta 2	$W = 448.5$, $p = .8508$, $Mdn = .12, .099$
Exploration time	Ta 3	$W = 463.5$, $p = .9885$, $Mdn = 145.5, 154$
Extraneous cognitive load	Ta 1	$W = 2057$, $p = .2656$, $Mdn = 3, 3$
	Ta 2	$W = 1883$, $p = .8533$, $Mdn = 3, 3$
	Ta 3	$W = 1711.5$, $p = .4698$, $Mdn = 2, 3$
Germane cognitive load	Ta 1 *	$W = 1484.5$, $p = .04702$, $Mdn = 3, 4$
	Ta 2	$W = 1782$, $p = .7269$, $Mdn = 3, 3$
	Ta 3	$W = 1821.5$, $p = .8895$, $Mdn = 4, 4$
Sensemaking Syntactic proxy	To 1	$W = 572$, $p = .113$, $Mdn = 10560.5, 9371$
	To 2	$W = 571$, $p = .1168$, $Mdn = 9607, 7569$
	To 3	$W = 360$, $p = .1427$, $Mdn = 3496.5, 3966$
Sensemaking Semantic proxy	To 1	$t = 0.73$, $df = 59$, $p = .466$, $M = .20, .19$
	To 2	$t = -1.17$, $df = 59$, $p = .2483$, $M = .22, .23$
	To 3	$t = -0.21$, $df = 59$, $p = .8314$, $M = .17, .17$

* $p < .05$

Figure 5: *P-values resulting from the tests comparing the LibSearch and the ExploViz distributions of results.*

5. Discussion and Future Work

We consider this experiment as an initial attempt to understand how visualisation, and specifically spatial visualisation, impacts upon exploratory behaviour. The approach was quite ambitious and requires further refinement before we can draw a claim. The aim of our approach was to move beyond the laboratory, recruit a large number and more diverse set of participants, yet apply an experimental methodology - an approach which has some success in similar domains [BDF15, FDPH17]. The difficulty is that, at the same time, we sought to measure exploration, which is in itself, a challenge. If we were to apply a similar methodology in the lab with a refined application, for example, our approach may yield a more positive outcome. Nevertheless, we view this study as a contribution as it enables researchers to reflect on some of the difficulties that we encountered and to possibly take from these to develop a more robust experimental framework. In the remainder of the paper, we reflect on the experimental design, consider some of the potential issues and discuss future work.

Data. The subject we selected for the experiment, business and finance, may not have been of interest to the participants in the study. Exploration is driven by innate curiosity and data related to individuals or to a particular cohort will result in different explorative behaviour. The size of the cohort, however, will dictate whether an experimental or observational approach is applied.

Event Capture. It is difficult to capture and compare behaviour across interfaces particularly if the interfaces provide a fundamentally different experience. We could not, for example, capture click events on ExploViz as the users do not need to click to use the visualisation system and explore the books. Using hover events provides some level of indication but lacks the level of specificity for accurate comparison. Event capture needs to accurately reflect this challenge. We encourage researchers to build a taxonomy of possible actions that can be undertaken by participants. This way, each measure will be more tightly controlled supporting a more comparative analysis. Reflecting this guideline, we would suggest comparing visual interfaces as a more controllable experimental approach - one using spatialisation for instance and the other using a standard visual framework for digital libraries.

Tasks and Measures. The use of NASA-TLX also requires revision. There are several variables that impact cognitive load in the current design of ExploViz - the search, the book selection, etc. are all design elements that influence cognitive load. We removed time-related questions from the NASA-TLX because they lacked relevance for the current study. Although other researchers have taken a similar approach, this calls into question the validity of the instrument when applied in this way. If the system has several design elements and the study has no time-bound tasks, then NASA-TLX is probably not suitable. Maybe a usability test, like a System Usability Score (SUS), can provide an indication of system's coherence instead. In addition, while distance is a useful measure and can be applied quite readily using similarity based techniques, it's application was rather bluntly applied because our analysis ignored relationships between books, asking questions such as did the visualisation encourage participants to bounce between subjects?

Participants. The results are, of course, dependent on participant's initial level of knowledge and general acceptation about technology and visualisation. Even though this kind of interface is not new for experts in visualisation, it was entirely new to the majority of participants. This novelty effect may have been a significant factor in the results. Similarly, list views represent the most common interface for searching and presenting information in use today. Participants may have developed or adapted a set of techniques to perform exploratory search tasks using the traditional search interfaces.

Future Work. Our aim to build on this work in three unifying directions. First, we need to establish a common task taxonomy and a set of measures that can be used to adequately compare exploratory behaviour across interfaces. Although there are taxonomies available, we are explicitly interested in exploration and this requires additional research. Second we need to marry observational analysis with experimental lab-based design in a mixed methods approach to support the creation of a more scientifically robust set of results. Third we will move away from comparing visual and non-visual interfaces, until at least we have established a way to address event capture (using eye tracking for example), and focus on assessing the exploratory affordances of different visual interfaces.

6. Summary

In this paper we presented the results from an initial experiment designed to assess the ability of spatial visualisation to encourage exploration. The study was ambitious in scope but requires further refinement. We considered how to address these issues and present directions for future work.

References

[ABLS15] ALETRAS N., BALDWIN T., LAU J. H., STEVENSON M.: Evaluating topic representations for exploring document collections. *Journal of the Association for Information Science and Technology* (Aug. 2015), 1–14. URL: http://onlinelibrary.wiley.com/doi/10.1002/asi.23574/abstract, doi:10.1002/asi.23574. 1

[BDF15] BOY J., DETIENNE F., FEKETE J.-D.: Storytelling in Information Visualizations: Does it Engage Users to Explore Data? ACM Press, pp. 1449–1458. URL: http://dl.acm.org/citation.cfm?doid=2702123.2702452, doi:10.1145/2702123.2702452. 4

[BISM14] BREHMER M., INGRAM S., STRAY J., MUNZNER T.: Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool for Investigative Journalists. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 2271–2280. doi:10.1109/TVCG.2014.2346431. 1

[BM13] BREHMER M., MUNZNER T.: A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2376–2385. doi:10.1109/TVCG.2013.124. 1, 2, 3

[BNJ03] BLEI D. M., NG A. Y., JORDAN M. I.: Latent dirichlet allocation. *Journal of machine Learning research 3*, Jan (2003), 993–1022. URL: http://www.jmlr.org/papers/v3/blei03a.html. 2

[FDPH17] FENG M., DENG C., PECK E. M., HARRISON L.: HindSight: Encouraging Exploration through Direct Encoding of Personal Interaction History. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 351–360. URL: http://ieeexplore.ieee.org/abstract/document/7539616/. 4

[Gar06] GARY MARCHIONINI: Exploratory search: from finding to understanding. *Communications of the ACM - Supporting exploratory search 49*, 4 (Apr. 2006), 41–46. URL: https://www.ischool.utexas.edu/~i385t-sw/readings/Marchionini-2006-Exploratory_Search.pdf, doi:10.1145/1121949.1121979. 1

[HK15] HOEBER O., KHAZAEI T.: Evaluating citation visualization and exploration methods for supporting academic search tasks. *Online Information Review 39*, 2 (Apr. 2015), 229–254. URL: http://www.emeraldinsight.com/doi/10.1108/OIR-10-2014-0259, doi:10.1108/OIR-10-2014-0259. 1

[KC09] KULES B., CAPRA R.: Designing exploratory search tasks for user studies of information seeking support systems. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (2009), ACM, pp. 419–420. URL: http://dl.acm.org/citation.cfm?id=1555492. 3

[LBF*13] LIU Y., BARLOWE S., FENG Y., YANG J., JIANG M.: Evaluating exploratory visualization systems: A user study on how clustering-based visualization systems support information seeking from large document collections. *Information Visualization 12*, 1 (Jan. 2013), 25 – 43. doi:10.1177/1473871612459995. 3

[MM14] MUNZNER T., MAGUIRE E.: *Visualization analysis and design*, crc press ed. AK Peters Visualization Series. A K Peters, Dec. 2014. 2

[Mun09] MUNZNER T.: A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (Nov. 2009), 921–928. doi:10.1109/TVCG.2009.111. 1, 3

[NHH*14] NGUYEN T. T., HUI P.-M., HARPER F. M., TERVEEN L., KONSTAN J. A.: Exploring the filter bubble: the effect of using recommender systems on content diversity. ACM Press, pp. 677–686. URL: http://dl.acm.org/citation.cfm?doid=2566486.2568012, doi:10.1145/2566486.2568012. 1

[Par12] PARISER E.: *The Filter Bubble: What The Internet Is Hiding From You*. Penguin, London, Mar. 2012. 1

[WF12] WILDEMUTH B. M., FREUND L.: Assigning search tasks designed to elicit exploratory search behaviors. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval* (2012), ACM, pp. 1–10. URL: http://dl.acm.org/citation.cfm?id=2391228. 1, 3

[WKSS10] WILSON M. L., KULES B., SCHRAEFEL M. C., SHNEIDERMAN B.: From Keyword Search to Exploration: Designing Future Search Interfaces for the Web. *Foundations and Trends in Web Science 2*, 1 (Jan. 2010), 1–97. URL: http://dx.doi.org/10.1561/1800000003, doi:10.1561/1800000003. 1