

# The Effects of Spatial Visualization versus Ranked Lists on Quality, Time Efficiency, and Interaction

Daniel Roßner  
Claus Atzenbeck  
daniel.rossner@iisys.de  
claus.atzenbeck@iisys.de  
Hof University, Institute of Information Systems  
Hof, Germany

Tom Gross  
tom.gross@uni-bamberg.de  
University of Bamberg, Human-Computer Interaction  
Group  
Bamberg, Germany

## ABSTRACT

Hypertext systems support users in navigating structured data sets and to find relevant information. Various interaction and visualization concepts aim to give users better insight into the data set, by suggesting queries and visualizing elements of interest in a meaningful way. Ranked lists are very common to show some sort of priority, while spatial layouts often help users to trace relations in the data. Only little research has been done in user studies that systematically show and reason about the differences of such spatial layouts and ranked lists. In this paper we report on a systematic comparison of a spatial visualization versus a ranked list layout. For this purpose, we did an between-subject study with 43 participants. One group performed a task with a system providing semantic visualization in 2D, the other group performed the same task with a ranked list. Both interfaces are very similar and only differ in how suggestions are visualized. The results show that users of the spatial layout finished their task in shorter time and have a tendency towards higher satisfaction. At the same time, they had more interactions with the system. Furthermore we discuss some in-depth data of the test sessions, which show that the visualization influences the users' behavior.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in visualization; Visualization systems and tools; Hypertext / hypermedia.**

## KEYWORDS

information retrieval, exploratory search, spatial hypertext, spatial layout, ranked list, user study

## ACM Reference Format:

Daniel Roßner, Claus Atzenbeck, and Tom Gross. 2022. The Effects of Spatial Visualization versus Ranked Lists on Quality, Time Efficiency, and Interaction. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media (HT '22)*, June 28–July 1, 2022, Barcelona, Spain. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3511095.3531286>

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
HT '22, June 28–July 1, 2022, Barcelona, Spain  
© 2022 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-9233-4/22/06.  
<https://doi.org/10.1145/3511095.3531286>

## 1 INTRODUCTION

The perception of associations of objects in spatial contexts depends on various factors, for example, proximity of objects, their alignment, color, and other visual means. The strength of using these rather implicit methods is, that they are very easy to create for human beings, as there are not many formal rules to follow [11]. Hypertext systems usually denote explicit node-link structures. Some of them, however, implement spatial interfaces, that allow users to express their associations visually. They usually follow a “cards on a table metaphor”. [18]. So called spatial parsers enable the machine to interpret the visual cues, making the foremost implicit structure explicitly available. We work on a system, which aims at using this interpretation, to fuel an interactive search interface, by inferring queries with the help of visual cues and by visualizing suggestions within a 2D space.

Information retrieval systems often utilize spatial interfaces to visualize results and allow various interactions, e.g. to refine the shown result. User studies showed superior performance (efficiency, effectiveness and satisfaction), of spatial interfaces compared to conventional one-dimensional list representations [8, 9]. Yet, little research is done in describing how users interact with these interfaces in comparison.

To shed some light on this question, we conducted a between-subject user study with 43 computer science students. Our goal was to test a certain spatial visualization for suggestions of an information retrieval system, and to collect all user interactions to have an in-depth data base to reason about advantages, disadvantages and differences in general. The spatial system works on a 2D space with a map based visualization. For comparison, we implemented a second interface that shows results in a ranked list instead. Both interfaces are similar in their look & feel, functionality and technical framework. This is because, we were particularly interested in the influence of a spatial arrangement on the way user interact with the system.

## 2 RELATED WORK

The user interface fundamental to our study. Driving idea is a spatial hypertext interface, that allows to create, delete, move, and change visuals of *symbols*. In our case, these symbols typically are represented by keywords or short text snippets. The application domains for such generic interfaces are wide spread, especially when thinking about open tasks like story writing [4] or knowledge management [12]. In general, tasks that incorporate brainstorming or evolving structures are suitable, because of the implicit nature

of spatial hypertext. Users are not forced to create any explicit structures or notations, but can rely on altering symbols in the space.

Exploratory search tasks match those described conditions very well: Interactive search systems give feedback on the relevance of resulting suggestions, they may even suggest queries, and offer a valuable visualization. All of this can happen in a spatial hypertext interface.

Our proposed interface is such a spatial hypertext – queries are built by structuring keywords in the 2D space, suggestions are rendered within this user generated structure and adapt to the changes made by the user. The overall system, called *Mother*, will be described in Sect. 2.2.

## 2.1 Visualization

The visualization of retrieved suggestions is related to the extensive work done in the area of information visualization and retrieval in general. For information retrieval, it is not only important to deliver high quality suggestions. As important is to explain them; giving them a reasonable context [3]. Systems like VIBE [14], Bead [5], InfoCrystal [19], VINETA [10], DARE [22], GUIDO [13] and Tofir [21], all with their own specialties, aimed at solving a similar issue, hence influenced our idea of a “meaningful, spatial visualization” [15].

## 2.2 Mother

Mother is a component-based open hypermedia system [1, 2] that can handle various structure types, including node-link and spatial structures. It uses dedicated structure services, implemented on a plug-in basis. Fig. 1 shows an exemplary setup, with several client applications, structure services and data bases in the three layers of the system. Colored in red is the setup used for the information retrieval system:

- (1) The spatial hypertext interface to build queries and receive/show the results.
- (2) A spatial structure service, implementing spatial parsers to analyze the visual structure of the space and to infer queries, which will be answered by
- (3) a knowledge base, containing domain-specific data, usually a weighted, un-directed graph.

Visualizations of suggestion nodes happen within the same space, which is used to (re-)organize the query keywords (controlled by the user). The position and behavior of suggestions is controlled by the machine, claimed to be semantically meaningful [15]. The positioning is a result of a real-time physics simulation, based on forces applied to elements by a spring metaphor. This force depends on *weights*, which are part of the search result. An example of this visualization is shown in Fig. 2, in context of an application to search for movies and actors. The authors argue that the positioning and dynamic behavior support users’ browsing and understanding of the spatial hypertext and, thus, the knowledge base which feeds the system with information [15].

Users can move, resize or delete their nodes, or accept suggestions (i.e., adding them as query keywords). The spatial structure service permanently analyzes all these changes and infers new queries, which update the resulting suggestions and their position

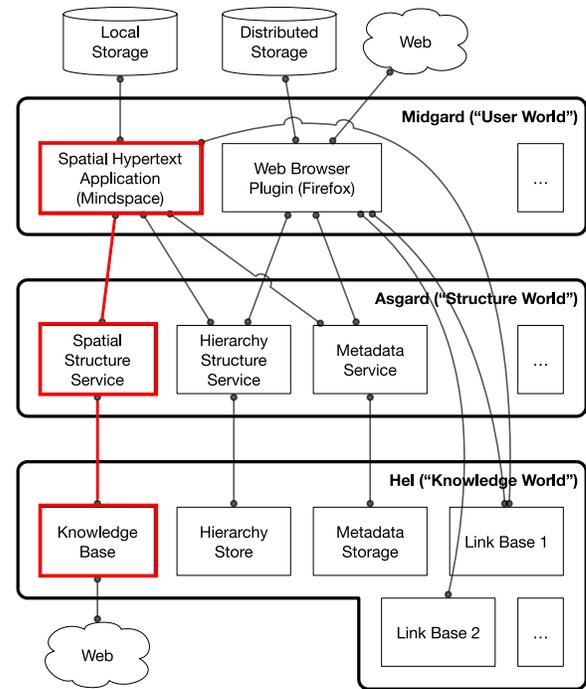


Figure 1: Mother’s architecture (exemplary setup)

in the space. Query generation in general is a complex process: Data can be filtered by various aspects, while filters themselves can be aggregated (AND, OR, etc.) and so on. Various implementations of visual query formulation systems aimed at modeling these complexities into a spatial canvas, e.g. with “movable filters” [6] or a visual hierarchy [17]. Mothers’s approach is not based on the aggregation of filters, but on visual relations, discovered by the spatial structure service. For each set of visual related nodes, the knowledge base is searched for other nodes, which are closely related to the set. Note that a set of visual related notes does not need to form a local cluster, as visual relation can be expressed by many means, other than proximity. This query formulation is not as expressive as the approaches mentioned above, but makes the application simple – technically and for the users. It already showed its applicability in recent prototypes, like the one in Fig. 2 or [7]. Yet, the system is open to be updated with a more sophisticated formulation method.

There are similar information retrieval tools, with focus on exploratory search [20]. Some of them also exploit a 2D space [14], others show results in a ranked list or use both [9]. As for the latter, researchers called their system “visual re-ranking”, because the spatial layout can be used to re-rank (change the order of) the resulting list. An accompanying user study revealed, that comparing to a baseline system without any spatial representation and interaction, perception and retrieval tasks were completed in less time (11% and 70% faster), while the overall effectiveness was not reduced.

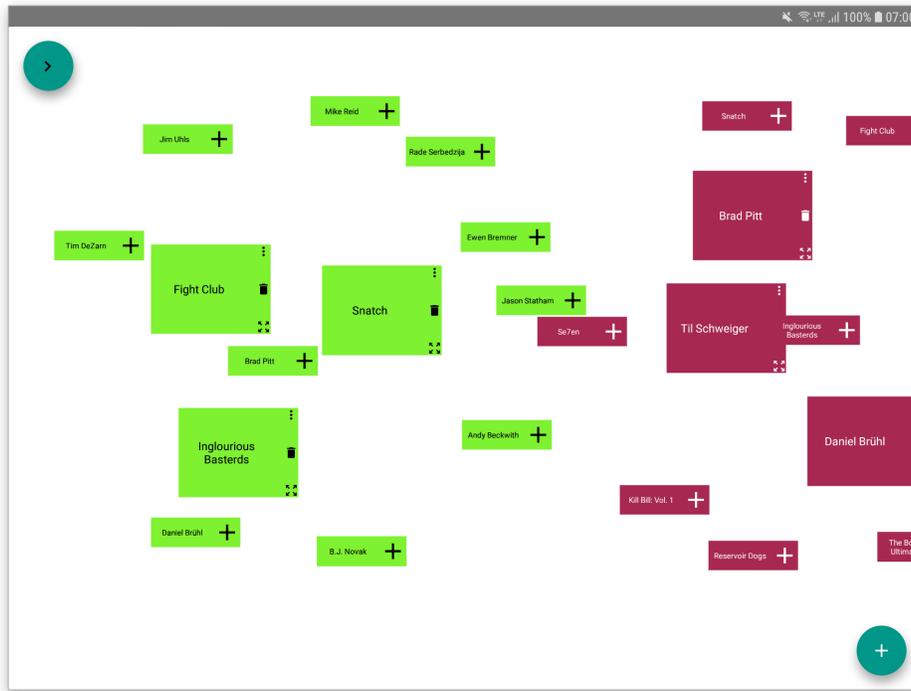


Figure 2: Android application as part of the *Midgard* layer, showing suggestions on query keywords (big rectangles)

### 2.3 Previous Results

Some preliminary results were already published in [16]. They suggest that the spatial visualization shows more efficiency, measured by the time needed to complete a certain research task. Satisfaction, measured by a subjective self assessment, tends to be better, yet this result is not statistically significant.

## 3 SYSTEM

We chose an experiment design with a spatial variant and a list variant. Both were tested remotely, with an identical knowledge base, offering the same suggestions. The general goal was to implement both variants as similar as possible, such that the visualization and interaction mechanisms with the result were the only parts that differ. Hereafter, we describe the system design, the details of both variants and how the knowledge base was created.

In the experiment we used two systems, one based on suggestion visualization as shown in Sect. 2.2, the other (i.e., the reference system) which does not show suggestions in a 2D space, but in a ranked list. Participants got randomly assigned to one of the systems. By adding or removing suggestions, users can alter what is suggested by the system.

Spatial parsers, as those described in Sect. 2.2, have been disabled. Hence, the query is not generated based on the visual structure, but by searching for all suggestions which are related to at least one added keyword. This is because we do not allow users to alter the position of any visual entity in the space, which makes spatial parsing superfluous. Instead, focus is given on how suggestions are

represented, not how users interact with already added keywords. In addition, this makes the comparison between both variants easier.

### 3.1 Spatial Variant

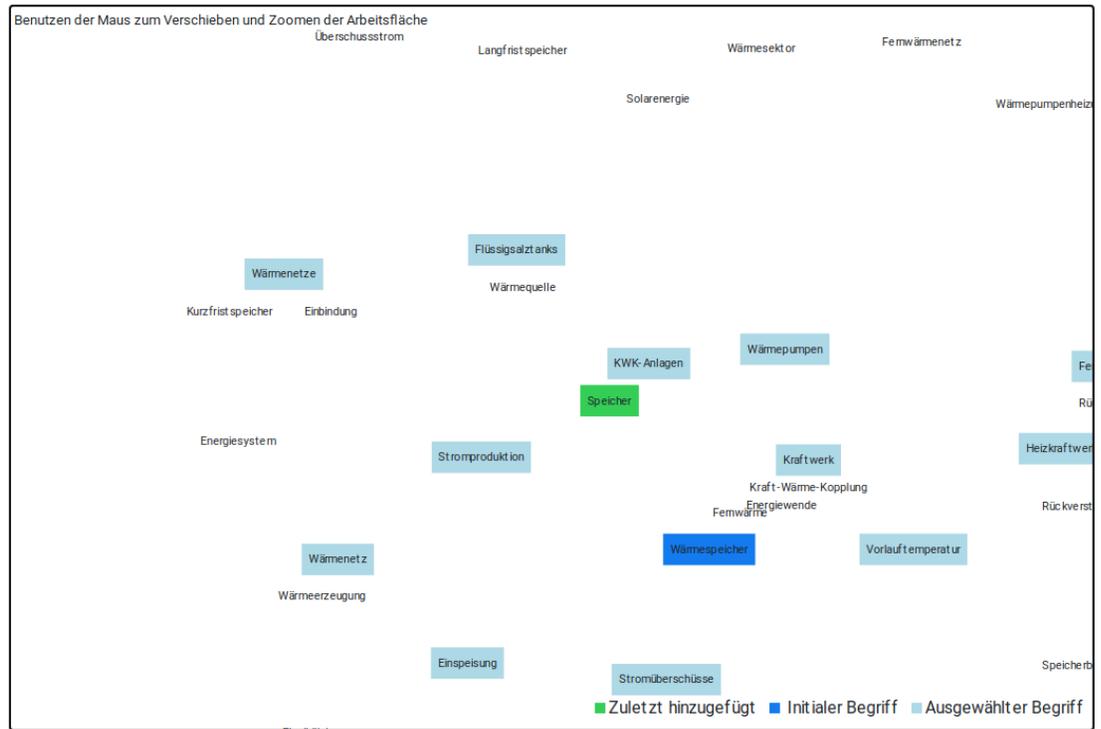
The spatial variant uses a viewport that matches the user's screen size. Some space on the left remains empty, because it is used by the list variant to show the ranked list. Everything – interactions, visualization and browsing – happens insight this frame. An example of an ongoing session in the spatial variant is shown in Fig. 3a. Colorized keywords are symbols added by the user, with

- *green*: initial (given) keyword, which cannot be removed,
- *light blue*: added keywords, that can be removed and
- *dark blue*: as light blue; the last keyword added.

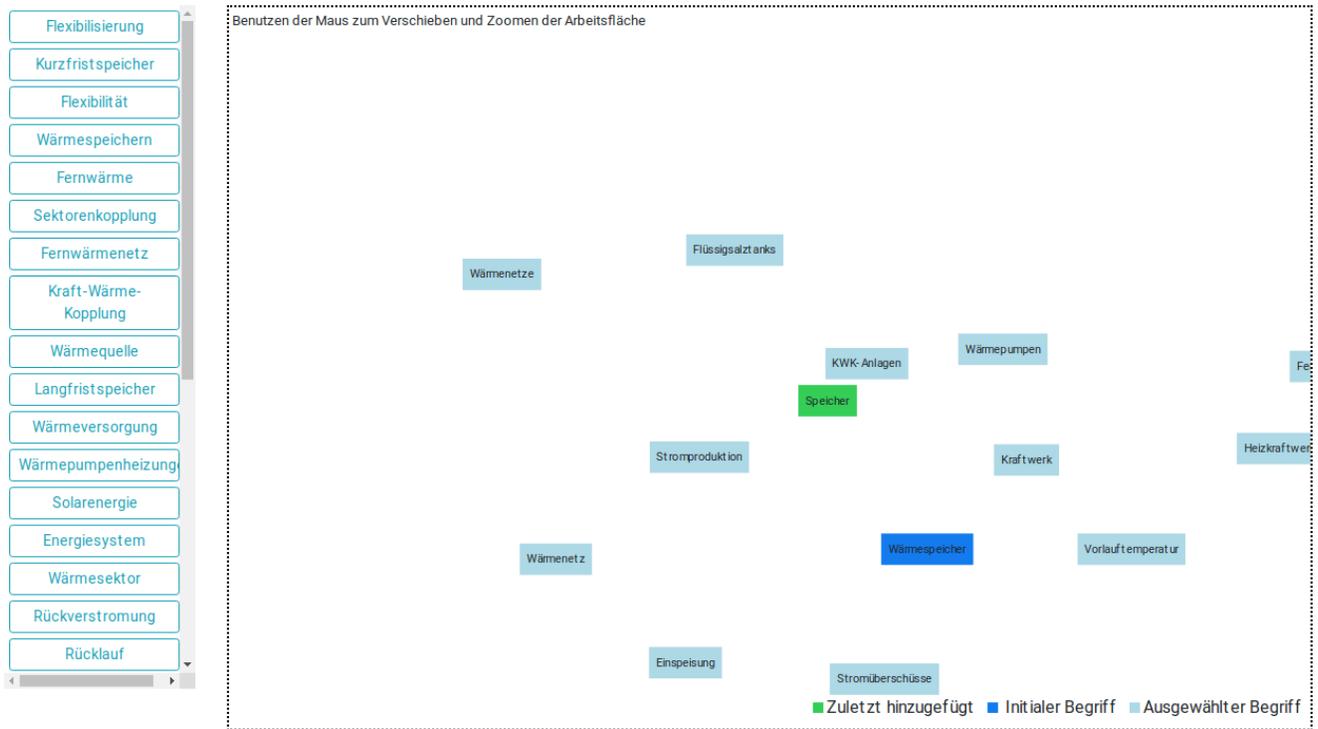
Suggestions do not have any colorization and are positioned with the algorithm described in [15]. In short, the authors describe a physics based approach, utilizing a simulated, annealing spring network to represent parts of a weighted graph (i.e., the knowledge base). Suggestions are generated by querying the  $s$  most, mean relevant nodes for the given set of keywords. The physics simulation reacts 'in real time', adjusting suggestions spatially until the spring damping leads to a stable layout.<sup>1</sup>

As we restricted the interactions to adding and deleting nodes, it is not necessary to actually show this annealing. The annealing is still simulated, but not in real time; only the final positions are rendered and visible to the user. Adding keywords is done by clicking on suggestions, deleting by clicking on the already added keywords. In both cases, the suggestions will update. Added keywords appear

<sup>1</sup>Demo video of the system: <https://youtu.be/GX53yezHDXE>



(a) Spatial Variant, shows suggestions within a 2D space



(b) List Variant, shows suggestions within a ranked list

Figure 3: Comparison of spatial layout (a) and ranked list (b). Suggestions are identical.

at the position where they have been originally suggested to be. The number of suggestions  $s$  depends on the number of selected keywords  $k$ . The formula

$$s = k \cdot 3 + 10 \quad (1)$$

is derived by the authors' experience and reflects that a growing number of spatially represented keywords leads to more space, which can be occupied by suggestions.

Panning and zooming the viewport helps looking into details or gaining an overview. Both actions are triggered by the computer mouse, the latter with the mouse wheel.

### 3.2 List Variant

The comparison system is very similar to the spatial one, as can be seen in Fig. 3b, and is inspired by the baseline system of [9], where suggestions are shown in a ranked list. Differences to the spatial system are:

- (1) Suggestions are not visible in the 2D space, yet their positions are calculated as in the spatial client. This is done, because we want to have comparable reaction times of the system variants and need them to render selected suggestions at these positions. As an alternative, there is a ranked list on the left of the 2D space. A higher rank means, that a suggestion has the higher mean relevance to all added keywords.
- (2) This user study examines differences in user behavior for visualizing suggestions. Thus, the list variant offers the same 2D space to show *already added keywords*. Users of the spatial client know where a selected suggestions appears (at the position of the suggestions), which is not the case when selecting a list entry. To compensate this disadvantage, the 2D space is panned automatically when a keyword is added, such that the recently added keyword is in the middle of the viewport; without altering the zoom. It is important to note, that users of the list variant can ignore the spatial view at all, if they do not want to remove keywords. In this case, it is just an overview of already added keywords, which the can be kept in their mind.

### 3.3 Knowledge Base

In this test scenario, the knowledge base is a weighted graph with keywords, formed around the topic sustainability, renewable energy and energy efficient buildings. All keywords are extracted nouns of nine German Wikipedia articles:

Windenergie (wind power), Passivhaus (passive house), Fernwärmespeicher (steam accumulator), Nullenergiehaus (zero-energy building), Sonnenenergie (solar energy), Methanisierung (methanation), Power-To-Gas, Wärmedämmung (thermal insulation) and Energieautarkie (self-sufficiency).

The weight ( $W_{i,j}$ ) is calculated by accumulating values, which depend on how often a pair of keywords  $k_i, k_j$  appear ( $a_n(k_i, k_j)$ ) in the same article (+1), paragraph (+2) or sentence (+3). To refine the result, the process analyzes all Wikipedia articles, which are related to at least one of the original ten seed articles. In total 1046 articles were analyzed. To get the final weight, the accumulated scores

are transformed by taking their natural logarithms. The inverted document frequency (IDF) of the keywords is stored as well.

$$W_{i,j} = \ln \sum_{n=0}^N a_n(k_i, k_j)$$

In order to provide matching suggestions, the knowledge base accepts an arbitrary number of keywords, which also must exist in the knowledge base. For all input keywords, an algorithm does the following:

- (1) Iterate all input keywords  $k$ .
- (2) Get all neighbors  $n(k_i)$  of iterated input keyword  $k_i$ .
- (3) Store the products of edge weight (between keyword and neighbor) and IDF of the neighbor.
- (4) The results of step 2 are summed up, when a neighbor is neighbor to more than one keyword.
- (5)  $s$  keywords with the highest relevance are returned as suggestions.

The summed up value is used to rank the list of suggestions, whereas the edge weights, not multiplied by the IDF, are used to feed the physics. In total, there are 2,179 keywords and 130,050 edges in the knowledge base.

## 4 METHOD

### 4.1 Participants

43 students in computer science (11 female, two not specified) with a mean age of 22.6 (ranging from 19 to 30) where asked to participate in a voluntary user study in the beginning of three beginner level lectures. An informed consent was obtained from all participants, prior of gathering any test data from them. During and at the end of the test, all participants got a unique code to delete or request their data from the authors' record. No compensation was given.

### 4.2 Procedure

Due to the current COVID-19 pandemic, there was no laboratory setup. Instead, we implemented a web based test client, which was introduced and explained to groups of participants via Zoom<sup>2</sup>. First, the students got an explanation, that they are invited to participate in a user study in the first 30 minutes of the lecture. After clarifying the optional character of the test, a link was sent in the chat of the Zoom conference client. There, they should click on a button to initialize their test instructions. This initialization is important, because the following page, beside the actual task, explained how the interface works. Thus, participants got the details of their system variant only. Descriptions were as similar as possible and did only vary for differences mentioned in Sect. 3.2. A verbal explanation of the task followed. The test started, when the participant was sure to understand the interface, gave his or her consent and pressed the button 'next'. After completing the task, participants where asked to answer some questions about themselves (gender, age, daily computer usage, domain knowledge) and to provide feedback to their assigned interface. Answers to all those questions were optional.

<sup>2</sup>Video conference tool

### 4.3 Task

All participants solved the same task and were randomly assigned to the test variant. Because we expected a ‘better’ performance for participants who are experts for wind turbines or very used to work on PCs/laptops, we asked them to rate their prior knowledge on that topic and to estimate their daily screen time. Regarding these two variables, we did not find any significant differences between the 21 participants large group, which used the spatial variant and the other 22. Most of the participants rated their prior knowledge as neutral (0.16,  $sd = 0.8$ ), measured on a scale from  $-2$  to  $2$ , with  $0$  equals means “average”. Mean daily screen time was about 9 hours. The letter number seems quite high and may be influenced by learning in online lectures only during this time of the COVID pandemic. As all participants were studying computer science, we can expect that they are used to working on computers with various applications and interfaces anyway. Participants were asked to imagine the following situation:

There are plans to build several new wind turbines (German: “Windräder”) close to your residence. Find as many as possible potential advantages, disadvantages, or other personally relevant topics you want to ask questions about at a soon to be held citizens’ meeting.

The task is completed, when participants have the impression that they had gathered enough information for the citizens’ meeting. Right after, they should write down the topics, which they will ask questions about. As described in Sect. 2, spatial hypertext is especially useful when solving open and exploratory tasks. Therefore, we do not set an explicit goal or time limit, but ask users to end the task whenever they think they are done.

The underlying knowledge base (weighted graph of keywords) was in German, as the participants were German native speakers. In both cases the systems shows an initial keyword, suitable for an open search task given to the users. In turn, they get suggestions based on this keyword and all other suggestions that were added by them.

### 4.4 Measures

During the test, any interaction with the system is recorded and stored, such that a detailed replay of a session could be created. We analyzed the effectiveness, efficiency, and satisfaction of the users. In this study effectiveness is defined as the outcome in the number of chosen topics, which were written down right after task completion. The efficiency is defined as the task duration in seconds, beginning with the first and ending with the last interaction with the system. Satisfaction is the subjective assessment of how helpful the system was, on a scale from  $0$  to  $10$  (“not helpful at all” to “very helpful”). Additionally, we wanted to characterize the respective interaction of the participants with either variant. For this purpose, we measured the number of pans and zooms (called *navigation*) as well as the number of additions and deletions (called *interaction*). A pan starts with a pressed mouse and ends, when the button is released. A zoom starts when the mouse wheel is used and ends with the last wheel usage, if the wheel is not used again within a second.

To gain a better insight, in how the systems are used over time, we record:

- (1) A timestamp for any event.
- (2) List variant: The position in the ranked list of an added suggestion.
- (3) For any deleted keyword, in both variants: Number of other keywords, that were added before this one got deleted.

The latter two are called “position” later on, referring to the position in the list and the position on the stack of added keywords.

## 5 RESULTS

We divide our results in two parts. First, the evaluation of satisfaction, effectiveness and efficiency are shown, by looking at the aggregated measures. This is followed by a detailed evaluation of the user sessions over time, including the discussion of possible reasons of effects discovered in the first part.

### 5.1 Aggregated Measures

The results for both test scenarios are summarized in Tab. 1 and Fig. 4. The number of chosen topics (written down after the search task) is crucial to this evaluation, as it makes the other measures comparable. Both flavors of the task had very similar outcome with about three chosen topics per user. An assessment of the topics showed a wide range, from very obvious concerns like those about noise pollution to more specific questions on influenced wind flow. This ‘quality’ of the topics could only be estimated by the authors and seems to be evenly distributed among both flavors. Looking at task duration, thus on our definition of efficiency, we observe significant improvement in the spatial setup compared to the list variant. From about 4 minutes (list) to a less more than 2.5 minutes. Fastest user of the spatial variant completed after 16 seconds (list: 66), while the maximum measured completion time was 577 seconds (list: 593). Note that the measured duration starts with the first interaction, which means users can have a first overview, before starting to interact with the interface (e.g. by clicking or zooming).

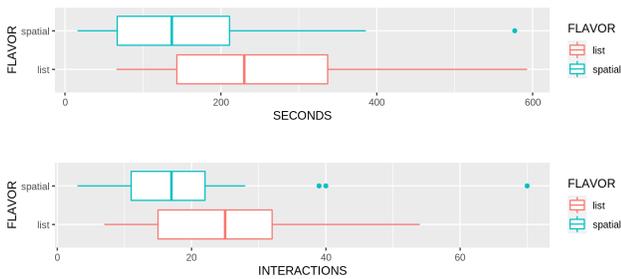
Obviously, users of the spatial interface navigated the space more frequently. Without moving the viewport, it would not be possible to get an overview of the provided suggestions. List users do not necessarily need to navigate the space, as their suggestion are shown in a separate list. Their only benefit would be to see all of their added keywords and to delete them (as keywords need to be in the viewport to be clicked/deleted). Indeed one list user did not pan or zoom at all; for 260 seconds, while adding 17 and deleting one keywords during the session.

In general, list users interacted (in sense of adding/deleting) more with the interface, yet the observed data is not significant (Mann-Whitney:  $p = 0.089$ ). A possible explanation could be that, due to a more helpful visualization, wanted suggestions are easier to identify.

Mean satisfaction (“How helpful was the interface?”) was slightly better for the spatial variant. This is neither statistically significant nor did we expect great differences here: Functionality and suggestions were identical in both cases, while the between-subject design of the study conditioned that users did only knew their assigned interfaces.

**Table 1: Summarized results for list and spatial variants; bold values indicate significance: \*  $p < 0.05$ , \*\*  $p < 0.01$** 

	List (L)		Spatial (S)		L vs. S	
	M	SD	M	SD	Test	
Chosen topics	2.5	1.5	3.3	2.8	$p = 0.5662$	(Mann-Whitney)
Task Duration (seconds)	251.6	147.2	162.5	133.4	<b><math>p = 0.02265^*</math></b>	(Mann-Whitney)
Helpful rating (0–10)	5.7	2	6.2	1.9	$p = 0.2504$	(Mann-Whitney)
Navigation (pan+zoom)	21.8	27.1	50.2	37.6	<b><math>p = 0.000741^{**}</math></b>	(Mann-Whitney)
Interactions (add/remove)	25.3	12.4	20.3	14.8	$p = 0.08811$	(Mann-Whitney)

**Figure 4: Task completion time in seconds and interactions for spatial and list variant.**

The provided feedback in the end of the test, helps to get a better impression, of how the participants experienced the interfaces, the task and the data. All following, cited text was originally written in German and got translated by the authors.

Two participants requested a better test introduction, regarding the explanation of how the interfaces work, e.g. by providing a “video tutorial or tutorial button” or “more images, less text”. Especially list users reported some sort of confusion: “After some time, the number of suggestions increased, which could not be associated correctly”, or “the list got too long and confusing towards the end”. Another list user wrote: “After selecting multiple topics, I lost track of which additional keywords I could add. Maybe it is possible to work with umbrella terms” and someone else: “Building suggestions upon selected keywords works quite good, but it is difficult to come back to a certain topic, e.g. if current focus is on noise pollution and you decide to dig further into animal welfare”.

Spatial users did not report this kind of confusion, but asked for more feedback, when and how the position of suggestions changes: “Maybe providing a connection between selected keywords helps, to trace how you got to the current state”, “the user could be guided to not yet shown suggestions, especially if they are outside of the current viewport” and “because selecting new keywords refreshes the interface and suggestions, you lose keywords you wanted to select next”. Yet “topics are evident by their proximity to each other”. Another interesting comment from the spatial participants concerned the use of colors. They appreciated the current color coding of the keywords (“[...] the color coding was good, too”, “I liked the color design”) and demanded the usage of more color in general: “however, you should bring in a little more color because otherwise it becomes monotonous”.

The relevance of the suggestions were assessed very differently, independently of the assigned interface. Ranging from “some of the suggestions were not very helpful” to “would use [the application] privately”.

## 5.2 Session Evaluation

**5.2.1 Progress over Time.** Interactive search is a dynamic process, which evolves and changes over time. In Sect. 5.1 we looked at the aggregated data, that is, how long did participants work on the given task in total. This analysis reveals that there is a higher efficiency (by duration) for the spatial interface, but left open the question, if this was a general advantage—or just an effect during a certain part of the task. To answer this question, we measured the time between interactions, depending on how many keywords have been already selected. If a user had 5 objects added (they begin with one seed) and needed 20 seconds to add the 6<sup>th</sup> keyword or remove one (such that she continued with 4), we added a data point with  $N(5) = \text{DURATION}(20)$ . Fig. 5 shows the resulting box plots (without outliers, to improve the readability) and loess curves. Some notes on interpreting the figure: For a certain  $N$  value, there can be more data points than participants, because deletions allow users to go back. Furthermore, the higher  $N$  the less participants influenced the result, because many users finished without having 44 (max  $N$ ) keywords added.

The first issue is mitigated by the fact, that deletion was sparsely used by all participants among both variants of the test. The numbers are shown Fig. 6, only around 13.4% of all interactions are deletions. With this method, the second issue cannot be neglected. It is just important to honor the given interval of confidence (0.95) in the figure. Fig. 7 shows that about 50% of the list users finished with 16 keywords, whereas users of the spatial variant finished with 12.

Given the collected data, we can conclude that the spatial visualization improves the mean time needed from interaction to interaction throughout the whole task. Interestingly, both loess curves seem to develop in parallel, showing a rising duration from the start to about 7 keywords, which flattens right after. We believe, that those curves show a training effect (flattening) together with the effect of increasing complexity (initial rising) when adding more keywords and getting more suggestions. Even though, the duration stabilizes without showing any significant tendency in either direction with increasing  $N$  keywords. It is not easy to answer which part of the system usage is responsible for these observations. We expected that most of the time, users skim through the suggestions, making a decision what their next step will be. Similar to Fig. 5,

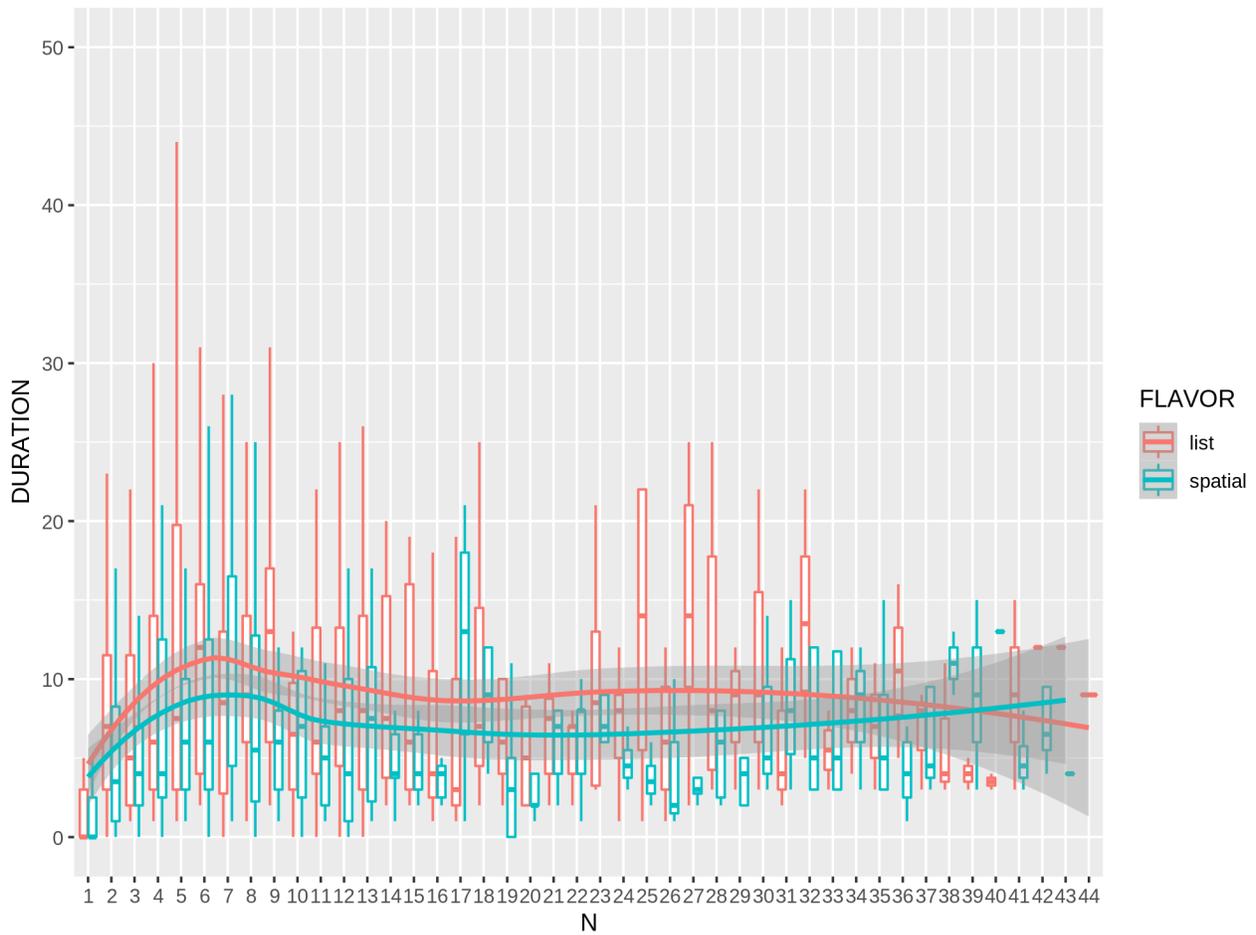


Figure 5: Duration between two interactions (adding/deleting), by number of already added keywords (N).

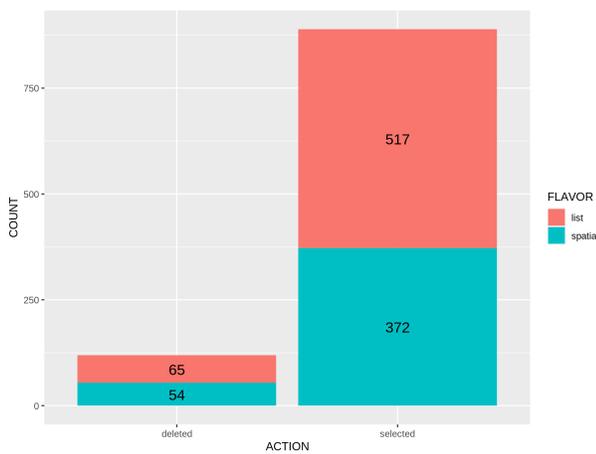


Figure 6: Number of deleted and added interactions by test variant.

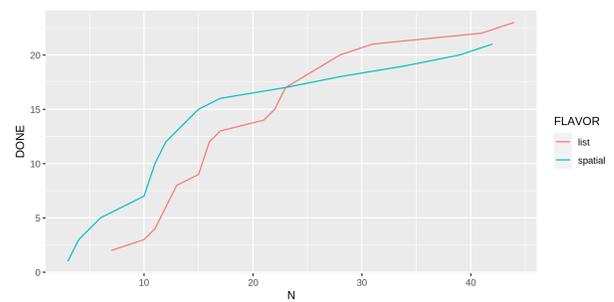
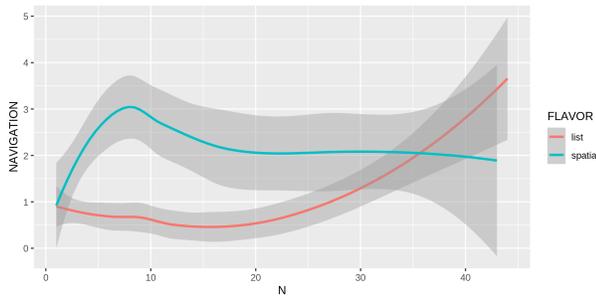


Figure 7: Number of participants, which completed the task (done) by the number of keywords they had added by completion (N).

we analyzed the number of navigation interactions (panning and zooming the space), depending on  $N$ , with the same flaws as mentioned above. The result is depicted in Fig. 8: The blue line (spatial) follows the logic of the duration chart, rising at the start, peaking at



**Figure 8: Number of navigation events (zoom/pan) between two interactions, by number of already added keywords (N).**

**Table 2: Position of deleted keywords.**

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
List	0	0	2	4.50	6	25
Spatial	0	0	2	10.33	14	41

$\approx 7$  and flattening to a stable value. This indeed suggests, that the participants used their time for browsing the results, but they do not invest more time or effort, when more suggestions are shown. For list users, the chart shows something different, but only in the first view. The process of “browsing the results”, does happen in the ranked list, not in the space. In many cases, navigation was not used at all. Probably they used their time to scroll through and read the list, as spatial users did within the space. Unfortunately, we did not record these scroll events. If this assumption is true, the conclusion is the same: Only a certain amount of suggestions, in the list or space, is honored by the user. In our case this means, that it does not make any difference when showing more than about 40 suggestions, in sense of time efficiency ( $N \approx 10$  cf. Eq. 1). Furthermore this seems not to depend on the visualization; a future study should (1) encompass more versions of Eq. 1 and (2) record scroll events within the ranked list to give these thoughts more evidence.

**5.2.2 Keyword Positions.** Position, as defined in Sect. 4.4, refers to the list position of suggestions and to the position in the stack of already added keywords, when they are removed from the space. Results show, that list users prefer those suggestions, visible without scrolling. All participants’ screens matched our requirements to show 18 suggestions in the list before a scroll bar is added. Fig. 9 depicts these numbers.

When users add a keyword by clicking on the suggestion (list or spatial) it is added to space; list users’ view is automatically moved to focus these new keywords. Removing happens by clicking on those already added keywords. Their position is determined by the number of keywords added afterwards, which are still in the space (hence: not yet deleted). Table 2 shows the quantiles and average values of these position values. In 50% of the cases, users did delete recently added keywords. Spatial users tend to delete ‘older’ keywords than list users.

## 6 DISCUSSION

Our results match with those of [9], whose test setup inspired this study. On the other hand, the relative, mean time gain (58%) is smaller. Yet, it is still difficult to name the reasons of this effect: Spatial views exploit the available space in 2D, thus one can see more suggestions at once. Zooming and panning may amplify this effect and both features are significantly used more. But also the layout of the suggestions may help to identify the helpful ones, the tendency of less additions and deletions give this claim some evidence. Further studies with more variations are needed to clarify these questions. Satisfaction and effectiveness in the spatial variant is at least as good as in the list variant, even tending towards a better subjective assessment and more results. Yet, the usage of chosen topics as measure for effectiveness is not a perfect solution. The open nature of the task allows this variable to be influenced, e.g. by the personal taste of the participants. Future work should feature a task, which addresses this issue, to give the presented claims more evidence.

The written feedback revealed some shortcomings of the study. Both interfaces can lead to some sort of confusion, which might be reduced by a more advanced introduction. This opens the question, how both interfaces perform in the longer run, with well-trained participants. For this study, we decided to work with keywords only, without e.g. linking to further explanations. This decision was based on the goal, to keep the participants within the interface and to eliminate uncontrolled influence. Further studies should encompass this, to prove that the system is applicable in a real-world scenario. For reasons, mentioned in Sect. 3, we disabled some of the features we usually implement in our spatial hypertext interfaces: visual parsing and dynamic visualization of the suggestions. The comments showed, that the latter feature is really important, as it would have made it easier to trace position changes of the suggestions.

Our session evaluation shed some light on the details, instead of looking at the final numbers only. That spatial users were faster in any increment of the task shows that the increased efficiency is not just influenced by a certain aspect e.g. in the beginning of the task. Users seem to find information they need, faster within a meaningful 2D layout. We believe that this advantage decreases for less open and exploratory search tasks, because the strength of a ranked list is to show the very best suggestions in a prominent position. Another interesting variation for list users, would be the visualization of already added keywords. In our setup, it was the same as for spatial users. A negative aspect of this view is, that users need to navigate the space to get an overview of all added keywords – something list users did less. They might profit from a different view, where their keywords are shown in a simple list or table. Furthermore, the ranked list can be improved by various re-ranking methods. All this should be approached by future studies.

The measure of navigation between interactions suggests that participants invested most of their time in browsing the suggestions, until they trigger the next step eventually. This browsing was easy to measure in the spatial visualization, as reading some keywords, zooming out, moving the view, etc. happen in a very natural order. The list visualization is not aware of such browsing interactions, only scrolling the list is something comparable. All future instances

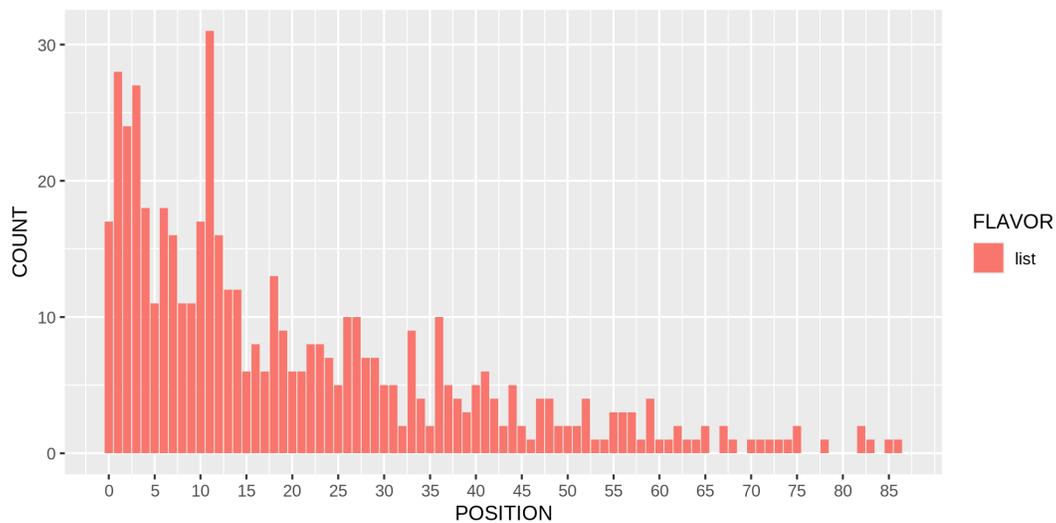


Figure 9: Position of added keywords in the ranked list

of this test should take care about this. It may be worth to find a better measure for ‘navigation’, e.g. by using an eye-tracker.

The positions of added (list only) and deleted keywords were distributed as we expected: High-ranked keywords, in general those visible without scrolling, are added more often than those at the end of the list. Spatial users chose very similar keywords, hence the ranking we provided seems to be reasonable. Usually users delete keywords they recently added. This makes sense, because new keywords influence the resulting suggestions, sometimes in an unwanted direction. This is true for both variants, yet spatial users delete keywords, which are in the space for a longer time, more often. As mentioned above, list users do not have a good overview of already added keywords, if they do not navigate the spatial view. Thus, we expect list users to delete older keywords more often, if they would be available in a simple overview.

## 7 CONCLUSION AND FUTURE WORK

We conducted a between-subject study with 43 computer science students, to compare a ranked list layout and a spatial visualization for suggestions of a information retrieval system. The participants solved a task, which was based on an exploratory search. The most significant, measured differences discovered, where those for efficiency (task duration) and navigations in the space. A detailed examination of the session data over time revealed some interesting findings, like effects of learning and that the number of suggestions actually read by the users seem to have a limit, which is independent of the chosen visualization. However, the discussion revealed ideas for further, future studies, because some questions could not be answered completely. Other flavors and modifications of ranked lists or spatial layouts can be tested within the implemented system.

Furthermore, this study utilized a reduced feature set of the original spatial layout system. Especially the possibility to move nodes, while suggestions update their position in the space may

influence the results, as there are more opportunities to interact with the system.

## REFERENCES

- [1] Claus Atzenbeck, Daniel Roßner, and Manolis Tzagarakis. 2018. Mother - An integrated approach to hypertext domains. In *HT 2018 - Proceedings of the 29th ACM Conference on Hypertext and Social Media*. ACM Press, New York, New York, USA, 145–149. <https://doi.org/10.1145/3209542.3209570>
- [2] Claus Atzenbeck, Thomas Schedel, Manolis Tzagarakis, Daniel Roßner, and Lucas Mages. 2017. Revisiting hypertext infrastructure. In *HT 2017 - Proceedings of the 28th ACM Conference on Hypertext and Social Media (HT '17)*. ACM, New York, NY, USA, 35–44. <https://doi.org/10.1145/3078714.3078718>
- [3] Christoph Beckmann and Tom Gross. 2010. Towards a group recommender process model for ad-hoc groups and on-demand recommendations. In *Proceedings of the 16th ACM international conference on Supporting group work - GROUP '10*. ACM Press, New York, New York, USA, 329. <https://doi.org/10.1145/1880071.1880134>
- [4] Mark Bernstein. 2011. Can we talk about spatial hypertext. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia - HT '11*. ACM Press, New York, New York, USA, 103. <https://doi.org/10.1145/1995966.1995983>
- [5] Matthew Chalmers and Paul Chitson. 1992. Bead: Explorations in Information Visualization. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '92)*. ACM, New York, NY, USA, 330–337. <https://doi.org/10.1145/133160.133215>
- [6] Ken Fishkin and Maureen C. Stone. 1995. Enhanced dynamic queries via movable filters. *Conference on Human Factors in Computing Systems - Proceedings 1* (1995), 415–420. <https://doi.org/10.1145/223904.223960>
- [7] E. Herder, D. Roßner, and C. Atzenbeck. 2021. Reflecting on Social Media Behavior by Structuring and Exploring Posts and Comments. *i-com 19*, 3 (2021). <https://doi.org/10.1515/icom-2020-0019>
- [8] Martijn Kagle, Michiel van Wezel, and Patrick J.F. Groenen. 2011. Map Based Visualization of Product Catalogs. *Recommender Systems Handbook* (2011), 547–576. [https://doi.org/10.1007/978-0-387-85820-3\\_17](https://doi.org/10.1007/978-0-387-85820-3_17)
- [9] Khalil Klouche, Tuukka Ruotsalo, Luana Micallef, Salvatore Andolina, and Giulio Jacucci. 2017. Visual re-ranking for multi-aspect information retrieval. In *CHIIR 2017 - Proceedings of the 2017 Conference Human Information Interaction and Retrieval (CHIIR '17)*. ACM, New York, NY, USA, 57–66. <https://doi.org/10.1145/3020165.3020174>
- [10] Uwe Krohn. 1996. VINETA: navigation through virtual information spaces. In *Proceedings of the workshop on Advanced visual interfaces - AVI '96*. ACM Press, New York, New York, USA, 49. <https://doi.org/10.1145/948449.948458>
- [11] C.C. Marshall and F.M. Shipman III. 1993. Searching for the missing link: discovering implicit structure in spatial hypertext. *Proceedings of the fifth ACM conference on Hypertext* November (1993), 217–230. <http://portal.acm.org/citation.cfm?id=168826>
- [12] Catherine C Marshall, Frank G Halasz, Russell A Rogers, and William C Janssen Jr. 1991. Aquanet: A Hypertext Tool to Hold Your Knowledge in Place. In *Proceedings*

- of the *Third Annual ACM Conference on Hypertext (HYPERTEXT '91)*. ACM, New York, NY, USA, 261–275. <https://doi.org/10.1145/122974.123000>
- [13] A Nuchprayoon and R R Korfhage. 1994. GUIDO, a visual tool for retrieving documents. In *Proceedings of 1994 IEEE Symposium on Visual Languages*. 64–71. <https://doi.org/10.1109/VL.1994.363639>
- [14] Kai Olsen, James G Williams, Kenneth M Sochats, and Stephen Hirtle. 1991. Ideation through visualization: The VIBE System. *Multimedia Review* 3 (1991).
- [15] Daniel Roßner, Claus Atzenbeck, and Tom Gross. 2019. Visualization of the relevance: Using physics simulations for encoding context. In *HT 2019 - Proceedings of the 30th ACM Conference on Hypertext and Social Media*. ACM Press, New York, New York, USA, 67–76. <https://doi.org/10.1145/3342220.3343659>
- [16] Daniel Roßner, Claus Atzenbeck, and Tom Gross. 2021. Spatial Layout Versus List Layout: A Comparative Study. In *Human-Computer-Interaction – INTERACT 2021*. Springer, Cham, 495–498. [https://doi.org/10.1007/978-3-030-85607-6\\_65](https://doi.org/10.1007/978-3-030-85607-6_65)
- [17] Tony Russell-Rose, Jon Chamberlain, and Farhad Shokraneh. 2019. A visual approach to query formulation for systematic search. *CHIIR 2019 - Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (3 2019), 379–383. <https://doi.org/10.1145/3295750.3298919>
- [18] Thomas Schedel. 2017. Spatio-Temporal parsing in spatial hypermedia. *ACM SIG-WEB Newsletter Winter* (2 2017), 1–5. <https://doi.org/10.1145/3027141.3027142>
- [19] Anselm Spoerri. 1993. Infocrystal: A Visual Tool for Information Retrieval. In *Proceedings of the 4th Conference on Visualization '93 (VIS '93)*. IEEE Computer Society, Washington, DC, USA, 150–157. <http://dl.acm.org/citation.cfm?id=949845.949876>
- [20] Ryen W White and Resa A Roth. 2009. Exploratory Search: Beyond the Query-Response Paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 1, 1 (2009), 1–98. <https://doi.org/10.2200/S00174ED1V01Y200901ICR003>
- [21] Jin Zhang. 2001. TOFIR: A tool of facilitating information retrieval - Introduce a visual retrieval model. *Information Processing and Management* 37, 4 (2001), 639–657. [https://doi.org/10.1016/S0306-4573\(00\)00042-X](https://doi.org/10.1016/S0306-4573(00)00042-X)
- [22] Jin Zhang and Robert R Korfhage. 1999. DARE: Distance and angle retrieval environment: A tale of the two measures. *Journal of the American Society for Information Science* 50, 9 (1999), 779–787. [https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:9<779::AID-ASL6>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-4571(1999)50:9<779::AID-ASL6>3.0.CO;2-R)