

# Knowde: A Visual Search Interface

Maurice Schleußinger<sup>1,\*,+</sup> and Maria Henkel<sup>2</sup>

<sup>1</sup>*Center of Information and Media Technology; ORCID: 0000-0002-2709-1067*

<sup>2</sup>*Department of Information Science; ORCID: 0000-0002-5000-104X*

*\*Heinrich Heine University*

<sup>+</sup>*Corresponding author: [maurice.schleussinger@hhu.de](mailto:maurice.schleussinger@hhu.de)*

March 16, 2018

## Abstract

Information Visualizations are well-established to represent high density information in an intuitive and interactive way. There are no popular general retrieval systems, however, which utilize the power of information visualizations for search result representation. This paper describes Knowde, a search interface with purely visual result representation. It employs a powerful information retrieval system and works in a common web browser in real-time. This working prototype, with three different variations of network graphs will assist us in exploring current issues in visualization research, such as the challenge of system evaluation.

**Keywords:** Visualization, Information Retrieval, Information System, Visual Search Interface, Human-computer interaction

## 1 Introduction

In the digital age, the quantity and complexity of information we have to deal with in everyday life, has reached a whole new level. And while there is an abundance of services that help us with the retrieval and processing of needed information, the drive for more efficiency and a competitive advantage leads us to innovate systems, processes and services endlessly. Especially businesses recognize "the need for more effective tools for extracting knowledge from the data warehouses they are gathering" (Heer, Van Ham, Carpendale, Weaver, & Isenberg, 2008) and are implementing information visualization in their work flows. When it comes to human understanding of information, the visual approach is a very powerful one. "Visual displays provide the highest bandwidth channel from the computer to the human", which helps us to understand "large-scale and small-scale features of the data" observed and supports us in gaining new insight, for example, by recognizing properties of and also problems with the data, that were not anticipated before" (Ware, 2004). As a great quantity

of information can be processed more quickly with the help of visualization, it is not only helpful for “scientists and analysts”, but also for “commercial and personal use” (Heer et al., 2008). And yet, current search engines and interfaces work with text-based lists as means of result representation. And while information visualization is a research focus of growing interest over the recent years, little to none systems that utilize interactive visualizations for information retrieval purposes have established themselves as major player in the information systems and services market. Earlier approaches in combining information search and visualization in one system exist, but yielded mixed results in mostly modest evaluation attempts (Ellis & Dix, 2006) and were sometimes limited by bulky and slow hardware. Nowadays, we can say that we possess the required technology to enable better and more appealing design for retrieval and visual representation of information - not only on the web, but also in three-dimensional environments. To evaluate how visualization can optimize and enhance the information seeking and understanding process, we designed a visual search engine called Knowde. Additionally, we aim to tackle the several challenges and difficulties in evaluation of visual search interfaces. Knowde (**Knowledge Node**), is an information retrieval system that employs elements of information visualization as integral part of the result presentation. It adapts network graphs into a web search interface, using them as the sole means of search result representation and main focus for user interaction.

The purpose of this work is to introduce the Knowde system in its design and function. For which we briefly review related work and continue to present the system itself. Followed by a quick overview over issues and challenges such as evaluation methodology and cognitive load.

## 2 Related work

The idea of combining search and visualization in one system is not new. Earlier approaches, the information space prototype by Rohrer and Swing (1997), SENTINEL and NIRVE (Cugini, Laskowski, & Sebrechts, 2000), attempted 3D visualizations, requiring the user to navigate a three-dimensional space, using mouse and keyboard input and a 2-dimensional computer monitor. User evaluation yielded mixed results. Yi, ah Kang, Stasko, and Jacko (2007) developed Jigsaw, a system which already focused on the interactive exploration of relationships between entities (e.g. humans) in reports. It features a variety of user interfaces combined in one system. However, search query functionality is limited and not focus of the research. The *SIZL* (Searching for Information in a Zoom Landscape) system (Grierson, Corney, & Hatcher, 2015) features multiple search filters which can be combined to limit the result set in a 2.5D interface (an interface with 3D elements in a fixed 2D perspective). They create a specialized system which can accumulate the result sets of multiple searches for later review, but do not aim to provide a visual search interface for broad document search. Heer and Boyd (2005) developed an interactive User Interface called *Vizster* for browsing social networks. Its visualization is realized by a graph which centers



Figure 1: Screenshot of the Knowde system with its three modes (top to bottom)

around the user and displays their friends and their connectedness (via nodes and lines) as well as possible clusters of friends. It also features a list of details for a selected user and allows a simple keyword search. We want to employ a similar system for multimedia documents and emphasize the search functionality. FacetMap (Smith et al., 2006) features a search and filter system which puts emphasis on the navigation of hierarchical categories like date and type. Their visualization focuses on labeled bubbles in a grid system organized into categories. The system utilized data sets with rich metadata, contrary to many real world databases. Their evaluation showed users did not clearly prefer their system over a traditional interface (nor did performance improve significantly).

### 3 System & Data set

Knowdes' core design principle is the Visual Information Seeking Mantra by Shneiderman (2003): A viable visual design provides “overview first, zoom and filter, then details-on-demand.”, a step which is repeated multiple times during the data exploration process. Additionally, we classify our data set with the Task by Data Type Taxonomy by Shneiderman (2003) and Keim's classification of Information Visualization (Keim, 2002), a simplified and modernized variation. The given dataset is a combination of multidimensional data, including temporal and text data. Concluding from this classification, our visual representation should feature a temporal visualization, a representation of document entities as well as their relation and visualization of the multidimensional categorical information. To incorporate this, the system features different variations of node-link diagrams, or network graphs, enhanced by a time line. As mentioned before, some previously developed systems attempted to incorporate visual elements in search interfaces. But few of them actually focus on a powerful search, with state-of-the-art retrieval performance, term completion and multimedia-search (e.g., Office files) in junction with simple, intuitive visualizations on a temporal scale with a clear & modern design. The concrete visual design incorporates graphical codes from the visual grammar by Ware (2004). Entities are closed contours (circular nodes), relationships are lines between entities. Proximity between entities suggests a similarity (grouping). The visual system features three different modes as shown in Figure 1. They consist of a fixed search bar at the top and, depending on the mode, buttons for category selection. The entirety of the screen is dominated by the actual visualization of search results as network graphs. Resulting from the specific structure of the company's data set used in this example, there are two connected node types: reports (blue) and their attachments (green). The search function features a powerful search index which enables access to all textual information provided in the many file formats of the provided data set. Result relevance is expressed in the size of the nodes. Only small chunks (around 20 documents with their attachments) of the full result set are displayed initially, to reduce visual clutter, but more results can be loaded on demand. The user can fluidly zoom and drag in the visualization at any point using the computer's mouse. All resulting reports are sorted on a

time axis based on their creation date.

This interface, *mode 1* is the basis for all other modes (Figure 1, top). It allows us to evaluate the core idea of Knowde, interactive network graphs for result representation as well as more complex variations, in direct comparison. The point of having different modes or variations of the same interface is to gain more out of the evaluation results. Simply said, we want to find out which if any of our ideas perform well with users. In *mode 2*, categorical information is displayed in the form of a third type of nodes (orange). These additional nodes are sized by absolute occurrences in the result set and are connected to all report nodes with the corresponding category (Figure 1, middle). In *mode 3*, the same categorical information is displayed on the y-Axis of the window. The white stripes on the axis represent the total amount of values for the selected category, but only matching values are labeled. The value labels are sorted alphabetically and scale on zooming and panning together with the nodes (Figure 1, bottom). Therefore, at any point the category information for visible nodes can be retrieved and never changes its relative position. While *mode 2* and *3* transport the same additional information (categorical information), the means of displaying it differs. Evaluation will indicate if there is a user favorite and why.

As Knowde was developed in cooperation with a large, internationally operating company, a real data set was used. It consisted of 4571 reports, which linked to 7350 attachments. This sums up to a total of 4.3 GB of text files, images, office documents & videos. The prototyped system aims to be easily accessible. To allow all this, the final prototype only requires a modern web-browser running Javascript. The user simply visits a website and starts searching and exploring. The search index itself is implemented with Elasticsearch<sup>1</sup>, it provides a highly configurable search index for a variety of input data, and scales well for even larger data sets. Since the data set also contains a variety of office files, much of it is only accessible if the search index can cover such files as well. Elasticsearch provides a plugin which extracts text information from common formats (PPT, XLS, PDF and more). It can also recognize text in images and make them searchable too. The system contains an inverted index of all existing documents and processes keyword queries by splitting into separate keywords at whitespaces and joining them with the logical OR operator. Since Elasticsearch boosts documents matching multiple keywords, we still get the most relevant results first. For relevance ranking, a variation of the Vector Space Model with TF/IDF calculations (Gormley & Tong, 2010) is applied. Most of the extensive configuration possibilities remain at their default values and result in an acceptable (basic) IR system at this point (Belkin & Croft, 1987; Shneiderman, Byrd, & Croft, 1997; Baeza-Yates & Ribeiro-Neto, 1999). Data transfer for the interface and the visualization data is handled with the Meteor web framework, which provides web pages to the client's browser and keeps data between the server and the browser continuously synchronized. The user interface itself is implemented in styled HTML. For any user interactivity and styling of dynamic

---

<sup>1</sup><https://www.elastic.co/products/elasticsearch>

content (the visualization), the Javascript library D3<sup>2</sup> is used. It provides basic building blocks and paradigms for data visualization of any kind, ranging from a simple bar chart to complex force layouts. In summary, we designed an Information System based on established paradigms (Visual Information Seeking Mantra, Ware’s design grammar). The prototype features three modes with different variations of the design to allow meaningful evaluation. It uses using modern and powerful technologies for all components (Backend, User Interface, IR system). However, there are important issues to address as we step forward in the development and the evaluation of Knowde.

## 4 Conclusion and Current Issues

### 4.1 Addressing the Issue of Scale

Even before the development of the prototype began, the limit of scale for basically any visual representation of data entities became apparent. Using the web browser and D3 JavaScript library, more than 1000 data points render the interface unusable (Bostock, Ogievetsky, & Heer, 2011). Using significantly better performing technologies (e.g., with programming languages like C(++) or Java) would decrease both prototyping efficiency and user accessibility. There are, however, other possible solutions: van Ham and Perer (2009) claim there are use cases or technical limitations which make a grand overview of a data set impractical. They suggest a focus on user-relevant subgraphs and to enable continued browsing of subgraphs (loaded from a server as needed) to simulate an uninterrupted user-experience. Klouche, Ruotsalo, Micallef, Andolina, and Jacucci (2017) system displays nodes (search results) on a 2D plane aligned by their closeness to the entered search terms. The user can re-rank the results by tapping on the area between these search terms. *Mode 2* of Knowde is similar. The user can filter by clicking a category node to only show results of that category. Klouche et al. (2017) evaluation suggests an improvement in retrieval precision for complex tasks. Both van Ham and Perer (2009) and Klouche et al. (2017) offer solutions for a real limitation of Knowde. Scale. While the performance of Knowde is very good for even hundreds of items, the UI would become unresponsive if there was not a cut-off for the number of search results at some point. A future version could feature a real “overview first” (Shneiderman, 2003) showing an aggregated overview (e.g. documents grouped by year) of the entire data set, followed by specific sub-graphs determined by user input queries and implemented in a way as suggested by van Ham and Perer (2009) or Klouche et al. (2017) (“details-on-demand”).

### 4.2 Evaluating Visual Search Systems

To prove and improve the value of visual search interfaces, they have to be thoroughly evaluated and compared to conservative information systems. This

---

<sup>2</sup><https://d3js.org/>

kind of evaluation, however, is a very challenging task. Ellis and Dix (2006) reviewed “65 papers describing new visualization application or techniques” and found that only 12 described any evaluation and only two of them were deemed successful by the authors. They list several challenges and difficulties in visualization evaluation but also offer advice. To summarize: A good evaluation of for visualization systems should be based on real data, with relevant test subjects. It should not be limited to improve a system in small steps, but instead focus heavily on the exploration of truly novel insight regarding the research question(s). It should feature an iterative approach similar to the *user-centered design* introduced by Nielsen (1993). With its “holistic and comprehensive approach” the ISE Model qualifies as a general guideline for evaluating an information service (Schumann & Stock, 2014). It provides a number of evaluation techniques which are grouped in five dimensions, and allow extensive evaluation of key areas of any information system. It was not explicitly designed for visual information services but can still be adapted for such systems. Wares’ guidelines for evaluation techniques of visual systems mentions many of the methods proposed by ISE as well (Ware, 2004). The evaluation of Knowde employs and emphasizes the techniques of ISE which are useful for the evaluation of the visual interface component of such services. The first evaluation was conducted with 24 participants, employees of the company which provided the data set. Interviews consisted of an introduction where the system was explained, followed by a questionnaire regarding the dimensions of ISE. Most users found Knowde to be easy to use, useful and fun. It was ranked to be significantly better than the previously used system at the company. We also found a high overall system acceptance among the test users. Additionally, as part of the critical incident technique and as a thinking-aloud task, test subjects were asked to comment on anything extraordinary or unusual (positive or negative) and mention anything which comes into their minds. This allowed us to collect user interface improvements as well as some qualitative statements about the system. More details about the evaluation results and their implications will be discussed in a separate article.

### 4.3 Advances in Design vs. System Acceptance

Whether a system is accepted by users is influenced by different aspects. Many models divide system acceptance into the sub-dimensions “ease of use”, “usefulness”, “trust” and “fun” (Stock & Stock, 2013). A visual search interface like Knowde promises to be enjoyable and useful but may not be accepted due to other reasons. Although it has, to us, become a usual and casual activity, searching for information “is a mentally intensive task” (Hearst, 2009). In some cases, even a “spartan presentation” may be “too complex for some people.” (Hearst, 2009). Hence, greater functionality or a higher density of information are no improvements if the resulting system is too hard to use. While there might be the risk of cognitive overload, there is also the simple issue of habit. A visual approach may be unusual for users who are used to a certain design when it comes to information retrieval, for example the text-based lists that are common in web search. Nielsen (1993) explains, that today’s users have a firm mental

model of how a (web) search should look. “Deviating from this expected design almost always causes usability problems.” We still believe that today’s user-habits will change. Many of the design guidelines that are good practice now, might be obsolete for the next generation of knowledge systems. Until then, a hybrid approach may be the solution (Nguyen & Zhang, 2006; Clarkson, Desai, & Foley, 2009; Kraker, Kittel, & Enkhbayar, 2016).

## References

- Baeza-Yates, R. & Ribeiro-Neto, B. (1999). *Modern information retrieval*. Addison-Wesley Longman Publishing Co., Inc. doi:[10.1080/14735789709366603](https://doi.org/10.1080/14735789709366603). arXiv: [9780201398298](https://arxiv.org/abs/9780201398298)
- Belkin, N. J. & Croft, W. B. (1987). Retrieval Techniques. In M. E. Williams (Ed.), *Annual review of information science and technology, vol. 22* (volume 22, pp. 109–145). New York NY: Isevier Science Inc.
- Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2301–2309. doi:<http://doi.ieeecomputersociety.org/10.1109/TVCG.2011.185>. arXiv: [1408.4626](https://arxiv.org/abs/1408.4626)
- Clarkson, E. C., Desai, K., & Foley, J. D. (2009). ResultMaps: Visualization for search interfaces. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1057–1064. doi:[10.1109/TVCG.2009.176](https://doi.org/10.1109/TVCG.2009.176)
- Cugini, J., Laskowski, S., & Sebrechts, M. (2000). Design of 3D Visualization of Search Results: Evolution and Evaluation. *12th Annual International Symposium: Electronic Imaging 2000: Visual Data Exploration and Analysis*, 3960, 198–210. doi:[10.1117/12.378896](https://doi.org/10.1117/12.378896)
- Ellis, G. & Dix, A. (2006). An explorative analysis of user evaluation studies in information visualisation. *Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for information visualization - BELIV '06*, 1. doi:[10.1145/1168149.1168152](https://doi.org/10.1145/1168149.1168152)
- Gormley, C. & Tong, Z. (2010). *Elasticsearch The Definite Guide*. Sebastopol: O’Reilly Media.
- Grierson, H. J., Corney, J. R., & Hatcher, G. D. (2015). Using visual representations for the searching and browsing of large, complex, multimedia data sets. *International Journal of Information Management*, 35(2), 244–252. doi:[10.1016/j.ijinfomgt.2014.12.003](https://doi.org/10.1016/j.ijinfomgt.2014.12.003)
- Hearst, M. (2009). The Design of Search User Interfaces. In *Search user interfaces* (Vol. 1997). doi:[10.1017/CBO9781139644082.002](https://doi.org/10.1017/CBO9781139644082.002)
- Heer, J. & Boyd, D. (2005). Vizster: Visualizing online social networks. In *Proceedings - IEEE Symposium on Information Visualization, InfoVis* (pp. 33–40). doi:[10.1109/INFVIS.2005.1532126](https://doi.org/10.1109/INFVIS.2005.1532126)
- Heer, J., Van Ham, F., Carpendale, S., Weaver, C., & Isenberg, P. (2008). Creation and collaboration: Engaging new audiences for information visualization. In *Lecture notes in computer science (including subseries lecture notes in*



- artificial intelligence and lecture notes in bioinformatics*) (Vol. 4950 LNCS, pp. 92–133). doi:[10.1007/978-3-540-70956-5\\_5](https://doi.org/10.1007/978-3-540-70956-5_5)
- Keim, D. A. (2002). Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 7(1), 100–107.
- Klouche, K., Ruotsalo, T., Micallef, L., Andolina, S., & Jacucci, G. (2017). Visual Re-Ranking for Multi-Aspect Information Retrieval. In *Proceedings of the 2017 conference on conference human information interaction and retrieval - chiir '17* (pp. 57–66). doi:[10.1145/3020165.3020174](https://doi.org/10.1145/3020165.3020174)
- Kraker, P., Kittel, C., & Enkhbayar, A. (2016). Open Knowledge Maps: Creating a Visual Interface to the World's Scientific Knowledge Based on Natural Language Processing. *027.7 Journal for Library Culture*, 4(2). doi:[10.12685/027.7-4-2-157](https://doi.org/10.12685/027.7-4-2-157)
- Nguyen, T. N. & Zhang, J. (2006). A novel visualization model for web search results. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 981–988. doi:[10.1109/TVCG.2006.111](https://doi.org/10.1109/TVCG.2006.111)
- Nielsen, J. (1993). *Usability Engineering*. San Francisco: Morgan Kaufmann Publishers Inc.
- Rohrer, R. M. & Swing, E. (1997). Web-Based Information Visualization. *IEEE Computer Graphics & Applications*, 17(4), 52–59. doi:[10.1109/38.595269](https://doi.org/10.1109/38.595269)
- Schumann, L. & Stock, W. G. (2014). The Information Service Evaluation (ISE) model. *Webology*, 11(1). arXiv: [1407.4831](https://arxiv.org/abs/1407.4831). Retrieved from <http://www.webology.org/2014/v11n1/a115.pdf>
- Shneiderman, B. (2003). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The craft of information visualization* (pp. 364–371). doi:[10.1016/B978-155860915-0/50046-9](https://doi.org/10.1016/B978-155860915-0/50046-9). arXiv: [arXiv: 1011.1669v3](https://arxiv.org/abs/1011.1669v3)
- Shneiderman, B., Byrd, D., & Croft, W. B. (1997). *Clarifying search: A user-interface framework for text searches*. Corporation for National Research Initiatives.
- Smith, G., Czerwinski, M., Meyers, B., Robbins, D., Robertson, G., & Tan, D. S. (2006). FacetMap: A scalable search and browse visualization. In *Ieee transactions on visualization and computer graphics* (Vol. 12, 5, pp. 797–804). doi:[10.1109/TVCG.2006.142](https://doi.org/10.1109/TVCG.2006.142)
- Stock, W. G. & Stock, M. (2013). *Handbook of Information Science*. Berlin: De Gruyter Saur. doi:[10.1515/9783110235005.416](https://doi.org/10.1515/9783110235005.416)
- van Ham, F. & Perer, A. (2009). “SEARCH, SHOW CONTEXT, EXPAND ON DEMAND”: SUPPORTING LARGE GRAPH EXPLORATION WITH DEGREE-OF-INTEREST. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 953–960. doi:[10.1109/TVCG.2009.108](https://doi.org/10.1109/TVCG.2009.108)
- Ware, C. (2004). *Information Visualization: Perception for Design* (Second Edi). Elsevier.
- Yi, J. S., ah Kang, Y., Stasko, J. T., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1224–1231. doi:[10.1109/TVCG.2007.70515](https://doi.org/10.1109/TVCG.2007.70515)