Defining an Analysis: A Study of Client-Facing Data Scientists

Abigail Mosca, Shannon Robinson, Meredith Clarke, Rebecca Redelmeier, Sebastian Coates, Dylan Cashman, Remco Chang
Tufts University

Abstract

*As the sophistication of data analyses increases many subject matter experts looking to make data-driven decisions turn to data scientists to help with their data analysis needs. These subject matter experts may have little to no experience in data analysis, and may have little to no idea of what exactly they need to support their decision making. It is up to data scientists to determine the exact analysis needs of these clients before they can run an analysis. We call this step of the analysis process **initialization** and define it as: translating clients' broad, high-level questions into analytic queries. Despite the fact that this can be a very time consuming task for data scientists, few visualization tools exist to support it. To provide guidance on how future tools may fill this gap, we conducted 14 semi-structured interviews with client-facing data scientists in an array of fields. In analyzing interviews we find data scientists generally employ three methods for **initialization**: **working backwards**, **probing**, and **recommending**. We discus existing techniques that share synergy with each of these methods and could be leveraged in the design of future visualization tools to support **initialization**.*

CCS Concepts

• ***Human-centered computing*** → ***Human computer interaction (HCI); Visualization;***

1. Introduction

Data science has been rapidly advancing analytic techniques, producing increasingly sophisticated and accurate tools. These tools, coupled with greater availability of data, allow for analyses and insights previously unattainable even in recent years. Driven by the increasing complexity of analyses, many subject matter experts turn to data scientists to help with their analysis needs.

Given this trend, companies are employing more and more data scientists who directly interact with subject matter experts. These subject matter experts may have little to no experience in data analysis, and it is up to the data scientists to determine their exact needs before they can run an analysis. We call this step of the analysis process **initialization** and define it as: translating clients' broad, high-level questions into analytic queries.

However, few visualization tools exist to support **initialization**, and little research has been done to understand the kickoff stage of data scientist and subject matter expert collaborations. Understanding the specific processes data scientists employ and what pain points they encounter during this phase of analysis is essential to developing tools to support it. Most existing studies of data scientists focus on understanding their workflow and needs during the technical part of performing analyses. For example, Kandel et al. [KPHH12] identify data cleaning as a time-consuming process with little support in tools. Similarly, Alspaugh et al. [AZL*19] report the common challenges that data scientists face when performing open-ended data exploration.

More related to our effort are papers that specifically examine the collaboration between data scientists, such as the work by Isenberg et al. [ITC08], or studies by Kastra et al. on the co-located coordination between a data scientist and a subject matter expert during data analysis, referred to as "pair analytics" [KF14]. However, while these studies illuminate the activities and the roles of data scientists while performing the technical aspects of data analyses, little is known about how data scientists who work with subject matter experts identify what analysis their client needs them to run.

To gain a better understanding of the processes data scientists employ and what pain points they encounter when determining what analysis their client needs (i.e. during the **initialization** stage of data analysis), we performed semi-structured interviews with 14 data scientists from a variety of data-intensive fields: market research, biomedical research, policy research, and epidemiological and health research. In each field, we sought out professionals who directly interface with clients and either perform data analytics themselves or manage a team of other data scientists. The findings from these interviews can inform the design of future visualization tools.

In this paper, we provide an in-depth discussion of observed challenges of **initialization** based on our interviews. These challenges center around the level of clarity with which clients are able to express their analysis needs to data scientists. For example, does the client ask the data scientist to *tell me insights in this data* (low clarity), or does the client provide a testable hypothesis (high clarity)?

After analyzing the interviews, we found three common methods data scientists employ to better understand their client's needs: ***working backwards***, ***probing***, and ***recommending***. Each of these

DOI: 10.2312/evs.20191173

methods corresponds to a different level of clarity in the client's need. For example, ***working backwards*** serves a client with a high clarity need who can exactly specifying their desired analysis outcome, such as *I want to show my boss a slide with a statistic supporting targeting 20 - 30 year olds with mobile banking ads*. From this desired outcome, the data scientist can "work backwards" to the appropriate analysis. On the other hand, ***recommending*** serves a client with a low clarity need who may not know what they are looking for. It consists of the data scientist running a number of different analyses in order to see which results are most interesting to the client.

Finally, we discuss existing tools and techniques that share synergy with ***working backwards***, ***probing***, and ***recommending***. For example, we note similarities between query-by-example [Zlo57, Wat01], commonly used in databases, and ***working backwards***. Likewise, we identify a number of visualization systems that serve as examples of ***recommending*** such as Voyager [WMA*16], and VizDek [KHPA12]. We propose that future work may leverage these synergies to build visualization tools that support **initialization**, thus filling this gap in visualization tools for data scientists.

2. Related Work

Common practices, workflows, and pain points of data scientists have been studied extensively [KPHH12, KBHP14, KZDB16, KS11, FDCD12, CKW09, AZL*19]. These studies contribute valuable design implications for building analytic systems that better serve data scientists' needs. What differentiates our work is a focus on understanding data scientists' practices, workflows, and pain points *in relation to their clients*. Unlike existing work that seeks to build better systems for data scientists as a single entity, our goal is to understand how we might build systems that support data scientists' interactions with non-technical clients.

There have also been studies on supporting collaborative data analysis. Insenberg et al. studied how teams engage in data analysis and problem solving, and explore the idea of collaborative visualization [ITC08, IES*11, IFP*12]. Similarly, Heer et al. provide guidelines for the design and evaluation of collaborative visualization systems [HA07]. Together, these studies provide guidelines to designing collaborative analysis tools. Our work is related to these studies on collaborative analytics. However, instead of focusing on how analysts collaboratively perform an analysis, our focus is on the collaboration between a data scientist and a non-technical client. The difference in expertise between the two results in interactions and collaboration challenges that differ from the previous studies.

Our work is perhaps most similar to *pair analytics* studies. First introduced by Kasstra et al. [KF14], pair analytics is an experiment methodology that aims to better understand the cognitive processes behind collaboration between a visual analytic expert and a subject matter expert in solving an analysis problem. Kwon et al. [cKFY11] implemented this methodology to identify roadblocks (and potential solutions) to novice users of Jigsaw [SGL08]. Our work is similar in spirit to *pair analytics* in that we also seek to understand how data scientists interact with subject matter experts (their clients). However, unlike *pair analytics* which explores collaboration in performing the technical pieces of an analysis, our study focuses on the steps leading up to this point. Specifically, we focus on how data scientists determine their clients' analysis needs.

Parallels can be drawn between user-centered design methodologies [SMM12] and our work. The distinction between this work and ours is that user-centered design focuses on producing a visual artifact, the success of which depends on determining the correct design to answer specific domain questions. In contrast, we seek to understand how data scientists help their clients identify those specific domain questions, an intrinsically more nebulous outcome.

3. Interview Study Design

To better understand the processes data scientists employ during **initialization**, we performed semi-structured interviews with 14 data scientists. Below we describe the specifics of our study. The interview script is available as supplemental material.

3.1. Participants

Fourteen data scientists (8 females) were recruited to participate in the study. These data scientists work in a variety of industries, including market research, biomedical research, policy research, and epidemiological and health research. Some of these data scientists were mathematicians or statisticians by training, while others had backgrounds in marketing or other areas. All of the data scientists have the ability to run analyses ranging from cross tabulations to advanced machine learning algorithms, and all work with clients who have limited experience in analytics.

3.2. Interview Process

Interviews followed a semi-structured protocol where the interviewees were asked a set of predetermined, open-ended questions. There was no time limit for answering a particular question. In line with the semi-structured approach, researchers kept interviews as conversational as possible and were free to ask follow-up and clarifying questions that were not necessarily on the script. Each interview lasted 30 to 60 minutes and occurred in-person (11 interviews) or via a conference call (3 interviews). All interviewees signed a consent form to participate in the study which included agreeing to an audio recording. The full interview script is available as supplemental material.

4. Interview Analysis

Recorded interviews were analyzed qualitatively. First, two researchers used material from a pilot interview to iteratively develop list of themes of interest. Then, for each interview, two of the four researchers on the team individually listened to the interview and marked where themes of interest were discussed. Results were compared, and discrepancies discussed and resolved. Finally, for each theme, another researcher listened to all excerpts marked as pertinent by other researchers and amassed the findings below.

5. Findings

Interviews revealed a wide variety in the types of questions clients ask data scientists. Some clients have high clarity in their needs and go to a data scientist with something as specific as a testable hypothesis, for instance *Are 5th graders in New Hampshire better at math than 5th graders in Massachusetts?*. On the other end of the spectrum, clients may have little to no clarity in their needs and request something as ill-defined as *Tell me about this data*. Regardless, the data scientist must interact with their client to identify an analysis query. We found data scientists typically default to three tactics for **initialization**, discussed in detail below. Interestingly, the success of each tactic typically coincides with a specific level of clarity.

5.1. High Clarity Needs: Working Backwards

In the case of high clarity analysis needs, clients have a solid sense of what they want from the analysis but may not know what to ask in order to achieve their end goal. For example, a market researcher may have a clear goal of seeking evidence to support targeting millennial consumers in advertisements. However, because the client lacks analysis expertise, they may ask the data scientist to *"tell me about millennials"*.

In order to construct a formal analysis query for a high clarity request, data scientists explained that they will start by asking clients to define the end goal of their project and try to find the *why* behind a request. We call this technique ***working backwards***. To continue the example above, the data scientist would want the client to express their ultimate goal – evidence to support targeting millennials in ads. One data scientist explained that this tactic is effective because it provides a well-defined analysis outcome from which they can extrapolate the appropriate approach.

To find that "well-defined end point" data scientists focus on questions that dig into the purpose of the analysis. In other words, they try to determine why a client came to them and what ideally the client wants to walk away with. Data scientists will ask: *"What are you trying to accomplish? What is the gap in the knowledge or literature?"*, *"How do you see the final product?"*, *"Why are we talking? What do you need?"*, *"What don't you have?"*, *"What are your research questions?"* .

If the client answers these questions, the data scientist can rely on their own analysis expertise to work backwards to the appropriate analysis approach. For example, one data scientist explained how they proceed from a high clarity need: *"I'll repeat the [research] question back to them in a way that I know I can work with it, but I also want to make sure that I understand what they're asking."* In other words, this data scientist confirms their own understanding of the client's need by presenting it as a research question, then implements their training in analysis to choose the appropriate technique.

Sometimes, the client has a high clarity need and may request a specific, yet incorrect, analysis. Despite confusion on analysis technique, the clients' high clarity need still allows the data scientist to ***work backwards*** to an analysis approach. For example, one data scientist shared a time when the client specified their need as: *"I want to see if [X] is overused, and I want to have different people to rate [X]"*. The client (being unfamiliar with the functionalities of different statistical tests) requested a kappa test to answer their question. However, upon confirming the client's end goal, the data scientist realized that they did not need a statistical test at all. Instead, a simple qualitative analysis would suffice.

These examples highlight the importance of a well-defined end goal or analysis outcome. Essentially, with a well-defined outcome, the data scientist can identify and perform the necessary analytics, independent of the client.

5.2. Medium Clarity Needs: Probing

Sometimes a client will struggle in answering the data scientist's questions about a specific end goal, but still have a sense of what they would like to know from the analysis. This is characteristic of clients with medium clarity needs. For example, a client might know that they want to write a report that says *"drinking milk is healthy"*. However, the population for which they want to make this claim is not clear, nor is the definition of 'healthy'. In this case, data scientists will employ ***probing*** to understand the clients' problem space and determine an analysis strategy. One data scientist explained that you have to *"poke and prod to find the most important information."* Nine data scientists reported using probing questions to determine analysis approach. Techniques include: Getting to know the data with questions like: *"How many samples do you have? What is your data like (binary, continuous, categories etc)?"*. Understanding the problem space with questions like: *"What story are you trying to tell here?"*. Encouraging a conversation: *"I ask open ended questions that can't be answered with 'yes' or 'no'"*.

One data scientist explained that the purpose of ***probing*** is to get clients talking so that they provide enough background information for the data scientist to *"fill in the blanks"*: *"Once they get [talking] they can talk for days, so that gives me some information and some background... we need open-endedness so that we can fill the blanks of what they're talking to us about"*.

Unlike questions used to identify a high clarity need for a defined end goal, probing is a way for the data scientist to learn about clients' backgrounds, needs, and restrictions. Once the data scientist has a strong handle on these things, they can then fill in the blanks to define an end goal themselves and proceed as they would for a high clarity need.

5.3. Low Clarity Needs: Recommending

When ***probing*** and ***working backwards*** fail, the client probably has a low clarity need. For example, a market researcher may say to the data scientist *"tell me about millennials"*. In contrast to a client with a high clarity need, they may not be able to communicate what they specifically want to say about millennials (define an end goal), or provide additional specificity on their interest in millennials when asked probing questions. In cases like this, where the data scientist is unable to have a productive dialogue with their client, most will run several plausible analyses and present them to the client to see which is of interest. This process of running and presenting analyses will continue until the client identifies one as useful. We call this tactic ***recommending***. Eleven of the 14 data scientists interviewed mentioned recommending analyses to clients.

When ***recommending***, data scientists are employing a "best guess and check" method. They were previously unable to establish a discussion of needs with their client, so instead they take a guess as what analyses might be of use to them. They then present the results to see if the client does in fact find any of them useful. To do this, data scientists either utilize their expertise in the clients' domain or spend time talking with the client to gain a deeper understanding of that domain. From there, they make an educated guess at what the client's need may be, provide an analysis result based on that understanding, and then adapt based on client feedback. In the words of one data scientist: *"I'll get them that quick outline of, yes we have something and here's what I think it would be, and then I ask them to confirm."* Another data scientist described using this method in cases where the client attempts to use analysis jargon but may not do so correctly: *"[I] Predict what [the clients] mean to say when they use the 'wrong' words."* The data scientist would then recommend of what they perceive as the "right" analyses to the client and ask for confirmation.

6. Discussion

We found that during **initialization** data scientists are often faced with with murky and ill-defined questions from which they need to define a specific analysis query. Even though this process can be tedious and time consuming, current visualization tools offer limited support for defining an analysis need based on high-level, broad questions. This presents an opportunity to develop tools that better support **initialization**. Below, we discuss several tools and systems that share synergy with ***working backwards***, ***probing***, and ***recommending***, which we believe may be modified to support **initialization**.

Working backwards is similar to Query by Example techniques [Zlo57], or Visual Query Systems [LG17] in that both ask the user to provide an example of their ideal outcome. For instance, Tan et al. developed a system that that takes a table and reverse engineers a query that would produce such a table [TZES17]. Similarly, Jayaram et al. and Bonifati et al. developed systems that accept tuples from users and return a corresponding knowledge graph query or join query, respectively [JKL*14, BCS16]. Query by example is also utilized in the visualization field. For instance, Wongsuphasawat et al. built an interface that allows a user to query event sequences by providing a timeline of events similar to what they are interested in [WPTMS12]. Similarly, QuerySketch accepts as input a user-drawn line graph and returns stock prices for stocks whose price histories are similar to that input [Wat01]. The ability to generate database queries hinges on the relational algebra underlying databases. However, there is no such algebra describing the relationship between broad, high-level questions and analysis queries. Coming up with such an algebra would be the first step towards adapting these techniques to support ***working backwards***.

Probing could be achieved through systems that leverage incremental query construction techniques [ZZM*09]. A recent example of such technology is Zhao et al.'s semantic-enhanced query expansion system for retrieving medical image notes [ZFL*18]. Zhao et al. present a system for querying medical images with an interface that walks the user through building an image query via a series of drop-down menus [ZFL*18]. This technique could easily be adapted to support ***probing***, with a specialized set of drop-down menus for analysis query formation. However, such a system would only succeed with a reasonably bounded number of potential analysis queries. In order to achieve this, the types of data and analyses the system could support would have to be limited.

Similarly, techniques such as fuzzy querying [Tah77, MPS18] could help the user express partial knowledge when constructing a query. However, unlike fuzzy querying which takes as input an imprecise request and returns a set of potential answers, ***probing*** is more iterative in nature and the outcome of ***probing*** is a set of *additional questions* for the client. Despite this difference, fuzzy querying could be adapted to support ***probing*** by modifying the underlying algorithm to return a set of potential analysis queries and the questions they answer to the user as a set of potential answers.

Finally, ***recommending*** is already widely implemented in visualization systems such as Voyager [WMA*16], VizDek [KHPA12], Small-Multiples-Large-Singles [vdEvW13], and Foresight [DHPP17], to name a few. Recent work has produced further specialized visualization recommendation systems. For example, DataVizard provides visualization recommendations for structured data [ALB18]. The draw back to these systems is that they do not communicate to users what analysis questions the resulting visualization answers, an essential part of ***recommending***. However, recent visualization research on explaining machine learning models such as [CPM*18], and [DCCE18] holds the potential to solve this problem. Augmenting visualization recommendation systems with explanatory capability so that they communicate results, as well as the questions those results answer, would be a way to augment visualization recommendation systems to support ***recommending***.

While none of the tools mentioned above directly support **initialization**, we have provided starting points to modifying them to do so. Pursuing these avenues not only has the potential to lessen the burden of **initialization** on data scientists but also allows us to start conceiving of a world where clients can interact with a system that walks them through initialization independent of a data scientist.

7. Conclusion

This paper investigates a data analysis step we call **initialization** and define as: translating broad, high-level questions into analytic queries. **Initialization** is a preliminary step in data scientists' analysis process that is currently underserved by visualization. In order to identify specific tasks and pain points of data scientists when performing **initialization**, we conducted 14 semi-structured interviews with client-facing data scientists in an array of fields. Our qualitative analysis of interviews shows that data scientists generally employ three methods for **initialization**: ***working backwards***, ***probing***, and ***recommending***. We find the efficacy of these tactics corresponds with how clearly clients are able to articulate their analysis needs to the data scientist. Given these three tactics, we present existing tools and techniques that share synergy with each and that we believe could be leveraged in the design of future visualization tools to support **initialization**.

8. Acknowledgements

This work was supported by DARPA D3M FA8750-17-2-0107, and NSF CAREER IIS-1452977.

References

[ALB18] ANANTHANARAYANAN R., LOHIA P. K., BEDATHUR S.: Datavizard: Recommending visual presentations for structured data. In *Proceedings of the 21st International Workshop on the Web and Databases* (New York, NY, USA, 2018), WebDB'18, ACM, pp. 3:1–3:6. 4

[AZL*19] ALSPAUGH S., ZOKAEI N., LIU A., JIN C., HEARST M. A.: Futzing and moseying: Interviews with professional data analysts on exploration practices. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 22–31. 1, 2

[BCS16] BONIFATI A., CIUCANU R., STAWORKO S.: Learning join queries from user examples. *ACM Trans. Database Syst. 40*, 4 (Jan. 2016), 24:1–24:38. 4

[cKFY11] C. KWON B., FISHER B., YI J. S.: Visual analytic roadblocks for novice investigators. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct 2011), pp. 3–11. 2

[CKW09] CHIN G., KUCHAR O. A., WOLF K. E.: Exploring the analytical processes of intelligence analysts. In *CHI* (2009). 2

[CPM*18] CASHMAN D., PATTERSON G., MOSCA A., WATTS N., ROBINSON S., CHANG R.: Rnnbow: Visualizing learning via backpropagation gradients in rnns. *IEEE Computer Graphics and Applications 38*, 6 (Nov 2018), 39–50. 4

[DCCE18] DAS S., CASHMAN D., CHANG R. C., ENDERT A.: Beames : Interactive multi-model steering , selection , and inspection for regression tasks. 4

[DHPP17] DEMIRALP Ç., HAAS P. J., PARTHASARATHY S., PEDAPATI T.: Foresight: Rapid data exploration through guideposts. *CoRR abs/1709.10513* (2017). 4

[FDCD12] FISHER D., DELINE R., CZERWINSKI M., DRUCKER S.: Interactions with big data analytics. *interactions 19*, 3 (May 2012), 50–59. 2

[HA07] HEER J., AGRAWALA M.: Design considerations for collaborative visual analytics. In *2007 IEEE Symposium on Visual Analytics Science and Technology* (Oct 2007), pp. 171–178. 2

[IES*11] ISENBERG P., ELMQVIST N., SCHOLTZ J., CERNEA D., MA K.-L., HAGEN H.: Collaborative visualization: Definition, challenges, and research agenda. *Information Visualization 10*, 4 (2011), 310–326. 2

[IFP*12] ISENBERG P., FISHER D., PAUL S. A., MORRIS M. R., INKPEN K., CZERWINSKI M.: Co-located collaborative visual analytics around a tabletop display. *IEEE Transactions on Visualization and Computer Graphics 18*, 5 (May 2012), 689–702. 2

[ITC08] ISENBERG P., TANG A., CARPENDALE S.: An exploratory study of visual information analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, ACM, pp. 1217–1226. 1, 2

[JKL*14] JAYARAM N., KHAN A., LI C., YAN X., ELMASRI R.: Towards a query-by-example system for knowledge graphs. In *Proceedings of Workshop on GRAph Data Management Experiences and Systems* (New York, NY, USA, 2014), GRADES'14, ACM, pp. 11:1–11:6. 4

[KBHP14] KANDOGAN E., BALAKRISHNAN A., HABER E. M., PIERCE J. S.: From data to insight: Work practices of analysts in the enterprise. *IEEE Computer Graphics and Applications 34*, 5 (Sept 2014), 42–50. 2

[KF14] KAASTRA L. T., FISHER B.: Field experiment methodology for pair analytics. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization* (New York, NY, USA, 2014), BELIV '14, ACM, pp. 152–159. 1, 2

[KHPA12] KEY A., HOWE B., PERRY D., ARAGON C.: Vizdeck: self-organizing dashboards for visual analytics. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (2012), ACM, pp. 681–684. 2, 4

[KPHH12] KANDEL S., PAEPCKE A., HELLERSTEIN J., HEER J.: Enterprise data analysis and visualization: An interview study. In *IEEE Visual Analytics Science & Technology (VAST)* (2012). 1, 2

[KS11] KANG Y., STASKO J.: Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct 2011), pp. 21–30. 2

[KZDB16] KIM M., ZIMMERMANN T., DELINE R., BEGEL A.: The emerging role of data scientists on software development teams. In *Proceedings of the 38th International Conference on Software Engineering* (New York, NY, USA, 2016), ICSE '16, ACM, pp. 96–107. 2

[LG17] LLORET-GAZO J.: A survey on visual query systems in the web era (extended version). *CoRR abs/1708.00192* (2017). 4

[MPS18] MOREAU A., PIVERT O., SMITS G.: Fuzzy Query By Example. In *SAC 2018 - The 33rd ACM/SIGAPP Symposium On Applied Computing* (Pau, France, Apr. 2018). 4

[SGL08] STASKO J., GÖRG C., LIU Z.: Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization 7*, 2 (Apr. 2008), 118–132. 2

[SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (Dec 2012), 2431–2440. 2

[Tah77] TAHANI V.: A conceptual framework for fuzzy query processing–a step toward very intelligent database systems. *Information Processing & Management 13*, 5 (1977), 289 – 303. 4

[TZES17] TAN W. C., ZHANG M., ELMELEEGY H., SRIVASTAVA D.: Reverse engineering aggregation queries. *Proc. VLDB Endow. 10*, 11 (Aug. 2017), 1394–1405. 4

[vdEvW13] VAN DEN ELZEN S., VAN WIJK J. J.: Small multiples, large singles: A new approach for visual data exploration. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 191–200. 4

[Wat01] WATTENBERG M.: Sketching a graph to query a time-series database. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2001), CHI EA '01, ACM, pp. 381–382. 2, 4

[WMA*16] WONGSUPHASAWAT K., MORITZ D., ANAND A., MACKINLAY J., HOWE B., HEER J.: Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE Transactions on Visualization and Computer Graphics 22*, 11 (Jan. 2016), 649–658. 2, 4

[WPTMS12] WONGSUPHASAWAT K., PLAISANT C., TAIEB-MAIMON M., SHNEIDERMAN B.: Querying event sequences by exact match or similarity search: Design and empirical evaluation. 55–68. 4

[ZFL*18] ZHAO Y., FESHARAKI N. J., LI X., PATRICK T. B., LUO J.: Semantic-enhanced query expansion system for retrieving medical image notes. *Journal of Medical Systems 42*, 105 (April 2018). 4

[Zlo57] ZLOOF M. M.: Query by example. In *Proceedings of the May 19-22, 1957, nationl computer conference and exposition* (1957), ACM, pp. 431–438. 2, 4

[ZZM*09] ZENZ G., ZHOU X., MINACK E., SIBERSKI W., NEJDL W.: From keywords to semantic queries - incremental query construction on the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web 7*, 3 (2009). 4