A User Centred Perspective on Structured Data Discovery

Laura Koesten*
University of Southampton; The Open Data Institute
Southampton/London, UK
laura.koesten@theodi.org

ABSTRACT

Structured data is becoming critical in every domain and its availability on the web is increasing rapidly. Despite its abundance and variety of applications, we know very little about how people find data, understand it, and put it to use. This work aims to inform the design of data discovery tools and technologies from a user centred perspective. We aim to better understand what type of information supports people in finding and selecting data relevant for their respective tasks. We conducted a mixed-methods study looking at the workflow of data practitioners when searching for data. From that we identified textual summaries as a key element that supports the decision making process in information seeking activities for data. Based on these results we performed a mixed-methods study to identify attributes people choose when summarising a dataset. We found text summaries are laid out according to common structures, contain four main information types, and cover a set of dataset features. We describe follow-up studies that are planned to validate these findings and to evaluate their applicability in a dataset search scenario.

KEYWORDS

data search, human data interaction, data discovery, data portals

ACM Reference Format:
Laura Koesten. 2018. A User Centred Perspective on Structured Data Discovery. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3184558.3186574

1 PROBLEM

As the use of data-driven technology and with it the economic value of services based on data is growing [7, 22], the availability of data sources that are published on the web is increasing. More than one million datasets were made available by governments worldwide by 2013 [7]. In Europe there were more than 750, 000 datasets published by regional and national authorities, as indexed in September 2017 by the European Data Portal[1]. The site Data Planet[2] lists no less than 4.9 billion statistical datasets, many of which are public. This work focuses on structured data: data that is organised explicitly - such as in spreadsheets, web tables, databases or maps - whatever its format.

The scenario we focus on is when a user tries to find and use data that matches an information need. This can be someone trying to find an answer to a question requiring data, or someone searching for a full dataset to analyse. This work aims to understand how the discovery of new data sources can be improved by supporting users in assessing their fitness for use when selecting data sources from a pool of search results.

Let us take as an example a data journalist writing an article about the runway expansion at London's main airports in the UK. Examples of such stories can be found, for instance, on the Data Blog of the Guardian newspaper[3]. The journalist will look for factual evidence to substantiate her story, in the form of reports, news on similar topics, as well as data about the economic, social, and environmental ramifications of the project, arguing for or against expansion plans. A large share of the relevant data is already available online, published by governmental agencies, researchers, and other journalists. However, finding it is not always straightforward. The journalist could use regular search engines, fact checking services[4,5], and other channels in the same way she does when looking for less structured kinds of information. She might also know of the existence of a particular data catalogue, which offers access to collections of data resources released by one or several organisations. She might even query the API (Application Programming Interface) of a trusted data provider, or crawl the web looking for bespoke data snippets. Once she has identified several matches, her next step would be to explore the most promising among them and assemble a set of datasets most relevant to the narrative of her article. Depending on the tools used to discover the data, this step might involve downloading data files; working with different formats (e.g., CSV, XML, HTML, RDF, relational tables); choosing between several versions of a dataset; alongside sensemaking tasks such as establishing what exactly a dataset covers (its 'attributes' or 'schema'), and how accurate, complete, or up-to-date the data is.

This example illustrates the unique characteristics of the information seeking process for structured data opposed to for textual documents from a user perspective. Therefore the high level research question of this work is formulated as follows: *How can we help people select data that is relevant and useful for their task?*

This paper present two studies. The first focuses on the information seeking process of data professionals searching for structured data on the web, to better understand search strategies, tasks and selection criteria in dataset search. The second explores the characteristics of text summaries for data and their usefulness in a dataset selection scenario.

[*] Supervised by Elena Simperl and Jeni Tennison
[1] https://www.europeandataportal.eu/data/en/dataset
[2] https://www.data-planet.com/
[3] https://www.theguardian.com/data
[4] http://factcheck.org/
[5] https://fullfact.org/

WWW '18 Companion, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5640-4/18/04.
https://doi.org/10.1145/3184558.3186574

2 STATE OF THE ART

2.1 Information seeking for structured data

While literature describes user centric models for information seeking in web search, little research explicitly focuses on data as the information source. Information Retrieval, rooted in Computer Science, is concerned with the technologies that support finding and presenting of information [3] and was traditionally focused on whether a system can retrieve relevant documents. Information seeking, as a discipline rooted in Library Sciences, places the people and the finding or searching activity in the centre of attention [3]. Interactive Information Retrieval (IIR) studies users interacting with systems and information, the focus here is whether people can use a system to retrieve information [15]. This work is based in IIR and information seeking, aiming to understand the specific characteristics of how people retrieve structured data as the source of information opposed to other sources, such as textual documents or web pages.

Information seeking that is centred around (structured) data is not often discussed in literature [16]. The user's goal and information need might differ, as the type of information available differs when being able to search at the data level [30]. In a study with social scientists [17] found that users are willing to put higher effort in the searching and selection process of datasets than they do when searching for literature and that the quantity and quality of metadata are far more critical in dataset search than in literature search, where convenience is most important. As one of the few models explicitly focusing on structured data, [25] present the data science process, which is describing activities to collect relevant data, to explore data in order to make sense of it, and finally to build an analysis model to draw conclusions from it. Closer to this area is literature on interaction with databases [6, 13, 28]. The focus there has been mostly on how users compose queries and the degree to which these queries can be translated into SQL or similar [6]. Our scenario is much more open in the range of data sources it targets.

2.2 Data versus documents

There are several aspects that add to the complexity of search tasks for data. Web data is heterogeneous and comes from different sources, it needs to be understood in its context - much more than traditional documents [30]. This presents unique interaction challenges that require thinking about the user, as well as about the underlying system and the design of the interface. In contrast to document search, users need skills to access and download data; interpret different or limited formats the data might be available in; and understand connected licences and metadata. Furthermore, data requires context to create meaning [8], to make sense of it. Sensemaking is defined as the process of constructing meaning from information [3, 23]. Rieh et al. describe the sensemaking process as creating knowledge structures between information that has been acquired through an information seeking task [27]. While this applies to information seeking activities generally, we believe that this process has unique characteristics when the source of information is structured data. In their work on accessing statistical information Marchionini et al. emphasise that people need context as well as means to reveal the story behind numbers [21]. This varies with the level of expertise of the user in terms of technical skills, prior knowledge and data literacy. People looking to find answers with data will sometimes require not just to be pointed to the right database, spreadsheet, or list that might contain the information they need, but to the record that answers their question. This has implications for how data search is implemented, as algorithms that focus on metadata, which are the de facto standard in existing portals, do not have this capability. In addition to technological limitations, there is little research examining data search from a user perspective. This research, as much of the related work in data search and sensemaking with data, is based on the assumption that, in order to offer the best user experience, we cannot simply reuse or re-purpose principles, models, and tools that have been proposed for less structured sources of information.

2.3 Summarisation of datasets

In search scenarios, we are used to being presented with a snippet, the short summarising text component that is returned by search engines. This helps users to make a decision about the relevance of a search result when searching for textual documents [1]. We are able to create snippets that adjust their content dependent on the query, which have proven to be more effective [1]. However we are still far away from being able to provide the same user experience for data search. Johnson defines a summary as a brief statement that condenses information and reflects the central ideas or essence of the discourse [12]. A good summary should be able to represent the core idea, and effectively convey the meaning of the source [32]. Meaningful summaries of datasets can support the discovery process [24] by enabling users to understand the content of the dataset and skip irrelevant search results. We know that textual representations of data can be more effective, comprehensible and helpful in decision making than corresponding graphical representations [20, 29], even when the data is uncertain [11].

Textual summaries of datasets could generated automatically but in current practice they are created by people. Community guidelines for data sharing, such as the W3C's Data on the Web Best Practices[6] or SharePSI[7] have a technical focus on the machine readability of data. Dataset descriptions are included in their recommendations, though usually without much specification of what that should contain. This can be seen in metadata standards, such as DCAT, a vocabulary to describe datasets in catalogues[8], or schema.org[9], a set of schemas for structured data markup on web pages. We know that datasets become more useful when metadata descriptions are available and data becomes potentially more understandable [2]. However, we know little about what meaningful summaries of data should look like. In their review of summary generation from text, Gambhir et al. point out the subjectivity of the task and the lack of criteria for what is important in a summary [10] which is reflected in the general lack of guidance in current metadata standards.

[6] https://www.w3.org/TR/dwbp/

[7] https://www.w3.org/2013/share-psi/bp/

[8] https://www.w3.org/TR/vocab-dcat/

[9] http://schema.org/Dataset

3 PROPOSED APPROACH

We break down our high level research question into the following sub-questions for the initial study: (RQ1) How do people currently search for data? (RQ2) What are the characteristics of information seeking tasks for data? (RQ3) How do people evaluate data that they find? (RQ4) What types of tasks do people do with data? (RQ5) How do people explore data that they have found?

We believe that these aspects give us the necessary background knowledge to understand how we can support people in selecting and assessing data in a search scenario. User-system interactions are influenced by factors that are not easily observable or measurable [15]. For this reason, and given the investigative nature of our research, we focused the study on its qualitative element and used in-depth interviews to get the rich data about interaction processes and workflows we were looking for and complemented the results with a search log analysis of a data portal.

Based on the data science process by [25] we discuss the process of working with data from a user perspective in five pillars: *tasks, search, evaluate, explore and use.* This model was used as the basis for the initial study in this work [19]. The assumption is that this process is not linear and involves multiple iterations and backwards movements between pillars for many data centric tasks. The rationale behind this rather general approach is the lack of relevant literature that focuses explicitly on peoples' interaction with structured data on the web.

Based on the findings of this study we designed a second experiment to explore dataset summaries. From a user perspective, summaries of datasets are currently critical in the discovery process; they provide a basis for selecting a dataset from search results. Representing data as text is known to help people make sense of it [20, 29], but we know little about how a good summary should look in a data search context. Therefore this study explores the attributes that people choose to describe a dataset. We defined the following research questions for this study on textual summarisation of datasets:

(RQ1) What types of attributes of data do people choose when summarising a dataset? i) Do these attributes and attribute types vary from one summary to another for the same dataset? ii) Do these attributes and attribute types vary from one dataset to another? (RQ2) Do data practitioners and crowdworkers summarise datasets differently?

4 METHODOLOGY

This PhD project is at the beginning of its third year. After an extensive literature review we conducted two mixed-methods studies, informed by [5] for the purpose of this work. The initial study aimed to shed light on how data practitioners look for data online, with a focus on a qualitative component using in-depth interviews with twenty data professionals from various backgrounds. To supplement the in-depth interviews, we analysed a unique dataset of search logs of a large open government data portal. The sample consisted of a total of more than 100, 000 queries, of which more than 50, 000 were unique queries. This gave us a less obtrusive way to learn about the behaviour of data search users [14], of which our interviewees were a subset of (17 out of 20 participants mentioned they used this portal to search for datasets).

For the second study we conducted a task-based lab experiment in which 30 participants described and summarised datasets in a writing task. Subsequently, we conducted a crowdsourcing study in which we replicated the lab experiment with a larger variety of datasets; and asked crowdworkers to rate the dataset summaries according to perceived quality. This allowed us to get a better understanding of the influence of the underlying dataset and of differences in participants and settings on the resulting summary. We collected 150 long and 150 short summaries from the lab experiment and 250 crowdsourced summaries and analysed these qualitatively and quantitatively.

5 RESULTS

5.1 Initial study on information seeking for structured data

Key findings from the initial study showed that finding data is challenging, even for data professionals and that searching for data is more often than not exploratory and complex. It was evident that people across different skill sets and professional backgrounds follow common workflows when engaging with structured data. The majority of participants reported trying to obtain recommendations from people working in the respective field who are likely to know about a dataset. The majority of our participants reported often finding it difficult to locate the data they need. For instance, 80% of the participants described finding data online as a complex, iterative, process:

> (By an experiment participant) *I would get some things that looked really promising but weren't and then finally, through some kind of mysterious combination of search terms, I suddenly came across the dataset I'd been looking for the entire time*

Selection criteria for datasets in a search scenario showed unique characteristics. We found that when selecting data on the web people generally do not think they have enough information about the content of a dataset to make an informed decision. We identified dataset relevance, usability and quality as the high level dimensions of selection criteria in dataset search, as listed in Table 1. Key factors that help to select datasets from a pool of search results emerged to be information about provenance and about the methodology of data collection and analysis. The findings also showed that the concept of data quality is inherently task dependent, which is in line with literature (for instance [31]).

Assess	*Information needed about*
Relevance	context, coverage, original purpose, granularity, summary, time frame
Usability	labeling, documentation, license, access, machine readability, language used, format, schema, ability to share
Quality	collection methods, provenance, consistency of formatting / labeling, completeness, what has been excluded

Table 1: Information needs in dataset selection

We further found that the majority of textual summaries of data are perceived to be of low quality and limited usefulness.

The quote below illustrates a common response in this study which was part of the rationale behind the follow-up dataset summarisation study:

> (By an experiment participant) *It's very difficult first when you download new data, to have a quick idea of what the data represents, a quick summary of the data.*

5.2 Dataset summarisation study

As the results of this study are still unpublished we keep this section as an overview for the purpose of describing our approach and the type of results that can be expected. The findings of this study show that textual summaries were laid out according to common structures. We were able to isolate different components that the summaries were made of (four main information types), alongside common structures and detailed features. User-created summaries included judgements of which parts of the dataset are most important and aggregated elements of key importance into semantic groups. The results point to a number of features that could easily be extracted, but for which there is no standard form of reporting in general-purpose metadata schemata.

The study suggests a range of characteristics which people consider important when engaging with unfamiliar datasets. Some of them could be generated automatically, others would still require manual input, for example from the dataset creator or from other (potential) users. We saw for instance that all dataset summaries, as expected, refer explicitly to the content in the dataset. Extracting content features directly from the dataset, and representing them as text is still a subject of research, in particular in the context of extractive dataset summarisation [9] or semantic labelling of numerical data [26].

Our findings can inform the design of these methods by suggesting parts of a dataset that matter for human data engagement. At the same time, our analysis shows that most summaries also cover information that goes beyond content-related aspects. While abstractive approaches to automatically generate summaries exist, we believe that the levels of abstraction and grouping needed for the creation of meaningful textual representations of data are not yet being realised. We believe that to be truly useful, a summary needs to be a combination of extractable features, combined with contextual information, human judgement, and creativity. This applies to selecting the right content to consider, as well as to representing this content in a meaningful way.

Template for dataset summaries Based on our findings, we propose a template for text-centric data summaries, in the form of a set of questions that can be used as guidance in the summary writing process. Each question describes one of the attributes that were common in the participant generated summaries. Studies on text summarisation found that people create better summaries when they are given an outline or a narrative structure that serves as a template, as opposed to having to create text from scratch [4, 18].

6 CONCLUSIONS AND FUTURE WORK

By conducting in-depth interviews with data practitioners, we were able to obtain a better understanding of data-centric tasks, as well as about search, evaluation, and exploration strategies and the data qualities that influence the outcomes of these activities. In contrast to searching for digital objects, such as e.g. physical artifacts in a digital library, datasets contain information which can be used to contextualise them and so support a search process. We currently rely on metadata, which varies in quality and availability. In this work we argue that utilising the original data to enrich metadata can support users in data discovery and, moreover, potentially provide relevant indexable content which could make data search more effective. Our second study presents to the best of our knowledge, a first in-depth characterisation of human generated dataset summaries. This enables us to better define evaluation criteria for textual summaries of datasets and gives insights into selection criteria in dataset search. These results not only support our understanding of how people interact with and communicate about data, but could further inform automatic summary generation and future metadata standards.

We have identified two potential directions of this work, a study on selection criteria in data search and an evaluation of the dataset summaries, described in Section 5.2, in a search scenario. Based on findings from our initial study we hypothesise that selection criteria are specific to dataset search. We believe that in order to make a stronger case for the type of information that should be presented to users in a search scenario, we need a more in-depth study to validate our findings focusing on user defined selection criteria. We are proposing to apply a qualitative approach, using a diary study. We will use thematic analysis to better understand motives and considerations during the selection process of datasets in the context of a particular task. The results of this study can be used to validate our prior findings. More importantly they will contribute to a more in-depth understanding of selection criteria in dataset search.

We further propose to evaluate the effectiveness and usefulness of summaries from our second study in a search scenario as described below. We aim to aggregate the findings of this work in an experiment evaluating five different modes (1-5) of representing the content of summaries in a selection scenario: (a) Textual summaries created with the template, (b) Non-textual representation of summaries (table), (c) Textual summaries created with the template + non-textual representation of the summaries (table), (d) Textual summaries created with the template + visualisation of temporal/geospatial aspects in the data, (e) Sample previews of the data without a summary or other representation. This experiment will focus on the selection of a dataset by a user out of a list of search results and aim to answer the following research questions:

- (RQ1) Are people faster, more confident and do they select more relevant and useful results when they get presented with a summary in a dataset selection scenario?
- (RQ2) Are textual or non-textual summaries most effective and useful?
- (RQ3) Are summaries created based on the template more effective and useful?

Based on the findings of the initial study we hypothesise that textual summaries of datasets will be more effective and perceived as more useful in the selection process of datasets in a search scenario. We further assume that summaries created based on the template will perform better in comparison to summaries created without them. We also assume that textual summaries will be more effective and perceived to be more useful than non-textual summaries. However, the design of this experiment is still in progress and leaves room for discussion of whether it is the right direction and how to select suitable natural tasks, the environment in which such an experiment is conducted, as well as the choice of evaluation metrics.

The two studies outlined in this section would enable us to validate findings of this work. This allows for instance to iterate over the framework for human interaction with structured data which resulted from the initial study. We plan to refine and validate the template for dataset summary creation that resulted from our second study by evaluating the usefulness of the template and it's individual attributes in a dataset selection scenario. The results of this study can inform the development and improvement of manual as well as automatic summary generation, with the ultimate goal of improving people's interaction with data by making the experience more accessible and understandable. Additional work could be carried out on refining a semi-automatic approach to generating summaries, using the template by prompting crowdworkers to extract these elements from datasets.

This research aims to present a novel perspective on the conceptualisation of data centric search tasks, an in-depth analysis of selection criteria in data search and potential solutions to support users in the selection process of a dataset out of a pool of search results by proposing a more standardised approach of creating meaningful textual summaries of datasets for human consumption.

REFERENCES

[1] Lorena Leal Bando, Falk Scholer, and Andrew Turpin. 2010. Constructing Query-biased Summaries: A Comparison of Human and System Generated Snippets. In *Proceedings of the Third Symposium on Information Interaction in Context (IIiX '10)*. ACM, New York, NY, USA, 195–204. https://doi.org/10.1145/1840784.1840813

[2] Bruce E Bargmeyer and Daniel W Gillman. 2000. Metadata standards and metadata registries: An overview. In *International Conference on Establishment Surveys II, Buffalo, New York*.

[3] Ann Blandford and Simon Attfield. 2010. Interacting with information. *Synthesis Lectures on Human-Centered Informatics* 3, 1 (2010), 1–99.

[4] Ria Mae Borromeo, Maha Alsaysneh, Sihem Amer-Yahia, and Vincent Leroy. 2017. Crowdsourcing Strategies for Text Creation Tasks. In *Proceedings of the 20th International Conference on Extending Database Technology, EDBT 2017, Venice, Italy, March 21-24, 2017*. 450–453. https://doi.org/10.5441/002/edbt.2017.42

[5] Alan Bryman. 2006. Integrating quantitative and qualitative research: how is it done? *Qualitative research* 6, 1 (2006), 97–113.

[6] Tiziana Catarci. 2000. What happened when database researchers met usability. *Information Systems* 25, 3 (2000), 177–212.

[7] Gabriella Cattaneo, Mike Glennon, Rosanna Lifonti, Giorgio Micheletti, Alys Woodward, Marianne Kolding, Angela Vacca, Carla La Croce, and David Osimo. 2015. IDC, European Data Market SMART 2013/0063, D6 - First Interim Report. (16 October 2015). https://idc-emea.app.box.com/s/k7xv0u3gl6xfvq1rl667xqmw69pzk790.

[8] Brenda Dervin. 1997. Given a context by any other name: Methodological tools for taming the unruly beast. *Information seeking in context* 13 (1997), 38.

[9] Rafael Ferreira, Luciano de Souza Cabral, Rafael Dueire Lins, Gabriel Pereira e Silva, Fred Freitas, George D.C. Cavalcanti, Rinaldo Lima, Steven J. Simske, and Luciano Favaro. 2013. Assessing sentence scoring techniques for extractive text summarization. *Expert Systems with Applications* 40, 14 (2013), 5755 – 5764. https://doi.org/10.1016/j.eswa.2013.04.023

[10] Mahak Gambhir and Vishal Gupta. 2017. Recent automatic text summarization techniques: a survey. *Artif. Intell. Rev.* 47, 1 (2017), 1–66. https://doi.org/10.1007/s10462-016-9475-9

[11] Dimitra Gkatzia, Oliver Lemon, and Verena Rieser. 2016. Natural Language Generation enhances human decision-making with uncertain information. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers*. http://aclweb.org/anthology/P/P16/P16-2043.pdf

[12] Suzanne Hidi and Valerie Anderson. 1986. Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of educational research* 56, 4 (1986), 473–493. https://doi.org/10.3102/00346543056004473

[13] Jozef Hvoreckỳ, Martin Drlík, and Michal Munk. 2010. Enhancing database querying skills by choosing a more appropriate interface. In *IEEE EDUCON 2010 Conference*. IEEE, 1897–1905.

[14] Bernard J Jansen. 2006. Search log analysis: What it is, what's been done, how to do it. *Library & information science research* 28, 3 (2006), 407–432.

[15] Diane Kelly. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends in Information Retrieval* 3, 1-2 (2009), 1–224. https://doi.org/10.1561/1500000012

[16] Dagmar Kern and Brigitte Mathiak. 2015. Are There Any Differences in Data Set Retrieval Compared to Well-Known Literature Retrieval?. In *Research and Advanced Technology for Digital Libraries - 19th International Conference on Theory and Practice of Digital Libraries, TPDL 2015, Poznań, Poland, September 14-18, 2015. Proceedings*. 197–208. https://doi.org/10.1007/978-3-319-24592-8_15

[17] Dagmar Kern and Brigitte Mathiak. 2015. Are There Any Differences in Data Set Retrieval Compared to Well-Known Literature Retrieval?. In *Research and Advanced Technology for Digital Libraries - 19th International Conference on Theory and Practice of Digital Libraries, TPDL 2015, Poznań, Poland, September 14-18, 2015. Proceedings*. 197–208. https://doi.org/10.1007/978-3-319-24592-8_15

[18] Joy O. Kim and Andrés Monroy-Hernández. 2016. Storia: Summarizing Social Media Content based on Narrative Theory using Crowdsourcing. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW 2016, San Francisco, CA, USA, February 27 - March 2, 2016*. 1016–1025. https://doi.org/10.1145/2818048.2820072

[19] Laura M. Koesten, Emilia Kacprzak, Jenifer F. A. Tennison, and Elena Simperl. 2017. The Trials and Tribulations of Working with Structured Data: -a Study on Information Seeking Behaviour. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1277–1289. https://doi.org/10.1145/3025453.3025838

[20] Anna S. Law, Yvonne Freer, Jim Hunter, Robert H. Logie, Neil Mcintosh, and John Quinn. 2005. A Comparison of Graphical and Textual Presentations of Time Series Data to Support Medical Decision Making in the Neonatal Intensive Care Unit. *Journal of Clinical Monitoring and Computing* 19, 3 (01 Jun 2005), 183–194. https://doi.org/10.1007/s10877-005-0879-3

[21] Gary Marchionini, Stephanie W. Haas, Junliang Zhang, and Jonathan L. Elsas. 2005. Accessing Government Statistical Information. *IEEE Computer* 38, 12 (2005), 52–61. https://doi.org/10.1109/MC.2005.393

[22] Michael Chui Peter Groves Diana Farrell Steve Van Kuiken Elizabeth Almasi Doshi McKinsey Global Institute, James Manyika. 2013. Open data: Unlocking innovation and performance with liquid information. (2013).

[23] C Naumer, K Fisher, and Brenda Dervin. 2008. Sense-Making: a methodological perspective. In *Sensemaking Workshop, CHI'08*.

[24] T. T. Nguyen, Q. V. H. Nguyen, M. Weidlich, and K. Aberer. 2015. Result selection and summarization for Web Table search. In *2015 IEEE 31st International Conference on Data Engineering*. 231–242. https://doi.org/10.1109/ICDE.2015.7113287

[25] Hanspeter Pfister and Joe Blitzstein. 2015. cs109/2015, Lectures 01-Introduction. https://github.com/cs109/2015/tree/master/Lectures. (2015).

[26] Minh Pham, Suresh Alse, Craig A. Knoblock, and Pedro Szekely. 2016. *Semantic Labeling: A Domain-Independent Approach*. Springer International Publishing, Cham, 446–462. https://doi.org/10.1007/978-3-319-46523-4_27

[27] Soo Young Rieh, Kevyn Collins-Thompson, Preben Hansen, and Hye-Jung Lee. 2016. Towards searching as a learning process: A review of current perspectives and future directions. *J. Information Science* 42, 1 (2016), 19–34. https://doi.org/10.1177/0165551515615841

[28] Stefano Spaccapietra and Ramesh Jain. 2013. *Visual Database Systems 3: Visual Information Management*. Springer.

[29] M van der Meulen, R H. Logie, Y Freer, C Sykes, N McIntosh, and J Hunter. 2010. When a graph is poorer than 100 words: A comparison of computerised natural language generation, human generated descriptions and graphical displays in neonatal intensive care. *Applied Cognitive Psychology* 24, 1 (2010), 77–89. https://doi.org/10.1002/acp.1545

[30] Max L. Wilson, Bill Kules, m. c. schraefel, and Ben Shneiderman. 2010. From Keyword Search to Exploration: Designing Future Search Interfaces for the Web. *Foundations and Trends in Web Science* 2, 1 (2010), 1–97. https://doi.org/10.1561/1800000003

[31] Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. 2016. Quality assessment for Linked Data: A Survey. *Semantic Web* 7, 1 (2016), 63–93. https://doi.org/10.3233/SW-150175

[32] Hai Zhuge. 2015. Dimensionality on Summarization. *CoRR* abs/1507.00209 (2015). http://arxiv.org/abs/1507.00209