# From Bubbles to Lists: Designing Clustering for Due Diligence


Winter Wei, Adam Roegiest and Mary Mikhail
Kira Systems
Toronto, Canada
{winter,adam.roegiest,mary.mikhail}@kirasystems.com



## ABSTRACT

In due diligence, lawyers are tasked with reviewing a large set of legal documents to identify documents and portions thereof that may be problematic for a merger or acquisition. In an effort to aid users to review more efficiently, we sought to determine how document-level clustering may help users of a due diligence system during their workflow.

Following an iterative design methodology, we conducted several user studies with different versions of a document-level clustering feature consisting of three distinct phases and 27 users. We found that the interface should adapt to a user's understanding of what "similar documents" means so that trust can be established in the feature. Furthermore, the ability to negotiate with the underlying algorithm is facilitated by the establishment of trust. Finally, while the usage of this feature may be influenced by a user's role, it remains primarily a project management tool.



**ACM Reference Format:**

Winter Wei, Adam Roegiest and Mary Mikhail. 2019. From Bubbles to Lists: Designing Clustering for Due Diligence. In *2019 Conference on Human Information Interaction and Retrieval (CHIIR '19), March 10–14, 2019, Glasgow, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3295750.3298951


## 1 INTRODUCTION

When conducting due diligence, lawyers seek to find and determine the risk in a target company's legal documents (e.g., contracts, employment agreements, leases), often through extraction and analysis of passages from these documents. As there is usually time pressure to review the document collection as quickly as possible, there is no time to explore the collection. Rather lawyers attempt to find the hypothetical "needle(s) in the haystack" that may compromise the deal. Traditionally, these collections are compiled from several sources and so a project manager's job of estimating the amount of work required by the number of documents and assigning documents to lawyers becomes complex. This assignment process is often done "without rhyme or reason" and so may not leverage a particular lawyer's special capabilities. Risks and action items are identified after lawyers review and extract information from documents.

Despite of the lack of visibility into what may be relevant or important in this approach, it remains the most common way of working among most law firms. As pointed out by Russell et al. [15], sense makers tend to choose the lowest-cost method when facing a one-off task with limited time. Accordingly, we sought to provide lawyers with a feature that would allow them to quickly identify these risks and actions items without requiring extensive review of a document collection. One obvious technique to aid in this task would be clustering, such that anomalous/risky documents would ideally be identified as outliers (i.e., ungrouped documents). This resulted in the prototype featured in Figure 1, where the red bubble denotes the outlier group.

This prototype quickly proved to be confusing and a potential hindrance to the users we interviewed. They could find no discernible reason for how and why documents were (not) grouped together when using the feature. This led to a series of prototypes over three distinct phases spanning 8 months and 27 users (e.g., senior partner, junior associates), and marked the transition from a risk identification tool to a project management tool. Overall, we investigated the following research questions:

(1) What is the most appropriate representation for document-level clustering to help users make sense of the collection, while achieving their goals in the due diligence context?
(2) What does it mean for a user to "trust" the clustering algorithm? Should we focus on transparency of information, or a representation that meets their expectations?
(3) What are our users' mental models when it comes to "similar documents"? How do we design an experience to match their mental models?
(4) How would users' expectations of the clustering results and their inclination to negotiate with the clusters be influenced by their particular roles in a due diligence project?

While not all of these questions were answered to complete satisfaction, we found that aligning the interface to the terminology used by lawyers helps to build trust and an inherent understanding of why an algorithm behaves as it does. Once trust is established, users are more likely to want to negotiate with the algorithm to correct mistakes. This willingness appears to be beneficial since it is a lower cost activity than not using the feature.

## 2 RELATED WORK

There has been extensive work into data visualization techniques at-large, Bilal et al. [2] and Kucher and Keren [8] provide good overviews of the most recent developments and best practices. Chuang et al. [4] emphasized the importance of facilitating interpretations and trust when designing for model-driven visualizations. Kang et al. [1] suggest that visualization tools for domain experts should allow considered augmentation of the visualization



*CHIIR '19, March 10–14, 2019, Glasgow, United Kingdom*



ACM ISBN 978-1-4503-6025-8/19/03…$15.00
https://doi.org/10.1145/3295750.3298951

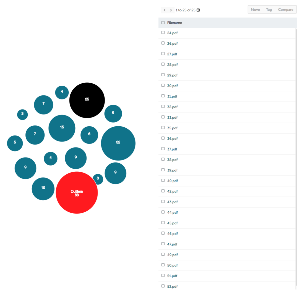

**Figure 1: An example of our initial clustering prototype based primarily on the belief that users would want to examine groups of similar documents and the outlier (red bubble) documents.**

such that users do not draw invalid conclusions. Along those lines, Russell et al. found that certain types of representation and manipulation were preferable to others for some tasks as they would have a lower cost [14]. As project managers often have to make sense of an unknown collection, Sarrafzadeh et al.'s suggestion that hierarchical visualizations can be better at providing an overview in uncertain situations [16] is an apt one for due diligence.

In the legal domain, while there are existing studies that examine the information seeking behaviour of lawyers, they mostly focus on case law [9] and academic lawyers [10] whose main tasks and therefore information seeking behaviour are different than those of transaction and corporate lawyers. As such, the design implications of visualization in the legal domain in existing literature have largely revolved around email collections and how to view the different relationships (e.g., email threads, social networks, etc) for the purposes of electronic discovery [3, 17, 18]. While such results may be useful in some diligence contexts (e.g., involving email), the results are not generally useful as the end goals between due diligence and electronic discovery do not necessarily align.

## 3 PHASE ONE

This phase represented our initial prototype (Figure 1), which was developed largely using our intuition about what users would want (i.e., risk identification). Accordingly, our hypotheses were that:

(1) Dissimilar documents would be prioritized for review, as we assumed they were anomalous documents.
(2) Reviewing documents that are similar would make review faster.
(3) Visualization is desirable in providing a high level overview to the structure of the document collection [6].

### 3.1 Methodology

We tested our initial prototype (Figure 1) with a selected group of interested users that all found the feature confusing. To determine what caused this confusion, we conducted a round of generative research [7] by interviewing 5 senior partners to understand their current workflos and goals. As part of this process, we presented them with several abstract visualizations (Figure 2) for different scenarios and asked what they might prefer. Following this round, we conducted several rounds of paper-prototyping and think-aloud studies to determine which interface may be of the most use. These

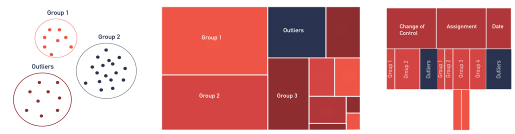


**Figure 2: Examples of the various cluster visualizations that were presented to senior partners during a user study. From left to right: circle clusters, treemaps, icicle diagrams.**

sessions tested 7 different designs with 16 legal professionals, including senior partners, project managers, and associates. Additional feedback was collected from in-house legal professionals to determine out how they might use such a feature.

### 3.2 Results

We find that for due diligence, abstract visualizations, like that in Figure 1, did not yield something lawyers can easily interpret. As one user opined, *"The Coca-Cola bubbles are more trouble than its worth, I like percentage, number and name but no bubbles."* Lawyers, when facing a tight deadline, tend not to find non-structured or non-hierarchical interfaces useful. They effectively want "just the facts" and this desire must be facilitated by the system and tool builders. As Tesler's Law of Conservation of Complexity [12] states that the complexity of a system is constant, the onus is upon the system builders to facilitate that desire at their expense and not the user's.

One aspect of user behaviour that became apparent in this phase is that users are not readily willing to trust an opaque algorithm. Fundamentally, they did not want to have to spend the time understanding how and why a cluster is formed. Users frequently expressed opinions such as *"At least highlight the provision that made it an outlier. There's gotta be a punchy way of explaining why something is an outlier,"* and *"I don't know what the computer is looking at, I'm hesitant to trust it. I don't know how it's grouping."* These attitudes highlight that "trust" and "transparency" are needed to achieve buy-in from users.

We believe the lack of trust in our original prototype stemmed from several factors. The first was that the visualization itself was overly simplistic and did not facilitate an understanding of why documents were grouped as they were. Secondly, our definitions of "outlier" differed from theirs in that we saw them as "documents that are not similar enough;" whereas, they define them to be "documents that should be similar but are not." Due to the above, we believe that we were not able to convey the appropriate information to the users and that this lack of knowledge transfer led to their dissatisfaction. Indeed, it has been shown [5] that humans will avoid algorithms when they have seen them "fail" to produce their desired outcome. Finally, our hypothesis that risky documents would be dissimilar ones did not bear fruit. Risk, as it turns out, can manifest in several different ways (e.g., unfavourable terms shared across multiple documents) rather than a single document not looking like others.

### 3.3 Design Implications

Based on the results of this first phase, our subsequent designs focused on aligning the interface and algorithms with the mental

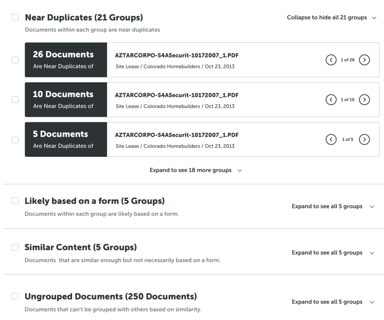


**Figure 3: A screenshot of our clustering user interface for Phase Two. This showed the four different classes of similarity that we determined users were interested in.**

models of the users. Instead of a one-size-fits-all clustering technique, we would seek to use several to identify near duplicates, documents that are "Based on a Form," documents that have otherwise "Similar Content" (e.g., specific types of leases), and documents that do not have a home elsewhere ("Ungrouped"). These categories are drawn from interviewees' own words to provide familiarity and align with their underlying goals more so than technical jargon or elaborate descriptions and justifications. This is similar to the findings of Attfield and Bandford [3] that, in electronic discovery, identifying and specifying the different classes of relevant and irrelevant material was a prerequisite for success.

During this phase, we rapidly designed and prototyped different variations of two major design paradigms to test the coarse grain effect of the designs. We started with representing a cluster of documents as a deck of cards with descriptive information on the card thumbnails, and soon discovered from user testing that it introduced more cognitive load as the card paradigm was not used to represent document collections in other parts of the user interface. This led to a quick decision to change the representation of document collections to lists, which conforms more to users' mental models. We found that users were quicker in identifying what they were looking at and making sense of the document collections.

This phase also made it clear that users need to be able to test actual working prototypes to discover how well their needs are met. Accordingly, in all subsequent phases, we encouraged users to use their own data, when possible, or to use data provided by us that simulates a due diligence collection.

## 4 PHASE TWO

Our main goal for this phase was to validate the effectiveness of the design and algorithmic changes made based upon our discoveries in Phase One. These changes are shown in Figure 3. In our testing for this phase, we sought to investigate the following:

(1) Increased algorithmic accuracy is essential to establish trust and increase likelihood of using the feature.
(2) Different categories imply different levels of similarity (e.g., near duplicates are more similar than "Similar Content" documents).
(3) That additional metadata is needed for understanding why documents are grouped together as filenames do not convey sufficient context.
(4) This feature is most likely to be used by project managers, and functionality should be optimized for their tasks.

### 4.1 Methodology

This prototype feature was provided to three users, who ran 10s to 100s of documents through it. A short survey was also sent to them to provide consistent feedback. In particular, we asked about the number and type of documents, their general impressions, how accurate they found each category, whether they might use this feature day to day, and in what capacity. We allowed them to define their own task goals to get a sense of the real-world applicability.

### 4.2 Results

From post-test interviews, we found an increased level of trust in, and understanding of, the feature. Users were able to understand the meaning behind the categories and subsequently form concrete ideas about what to expect. However, users felt there was some ambiguity between "Near Duplicates" and "Based on a Form" as they could imagine scenarios where there may be overlap that this design did not allow. We also discovered in the interviews that users would like the ability to correct/negotiate with the algorithm by moving documents from one cluster to another.

The metadata we provided in this phase (i.e., document title, parties in the document, and document date) proved to be a boon for users towards understanding why documents were grouped together. However, the paginated card view made observing the relationships between documents in a cluster hard for users, which increased sensemaking cost. Accordingly, the ability of project managers to assign users to review documents was hindered because they could not get a sense of what was in a particular cluster.

### 4.3 Design Implications

Based on our findings in this phase, we moved forward with three major changes. The first was to merge the "Near Duplicates" and "Based on a Form" categories into a new "Similar Structure and Content" category. The second was to provide a list view to allow users to view-at-a-glance all documents and document metadata for a cluster, which had the intended goal of reducing sensemaking cost. Finally, we added the ability to move documents between groups, create groups, and delete groups. We refer to these abilities in aggregate as "negotiation," since the movement of documents affects how subsequently added documents may be clustered.

## 5 PHASE THREE

In this final phase, we had what we believed to be a mostly robust solution (Figure 4). To verify this, we investigated the following hypotheses:

(1) The role of the users (e.g., project manager vs. reviewer) would influence their expectation on document grouping and the usefulness of this feature.
(2) The more users need to negotiate, the less likely they will use this feature as the cost of the tasks increases.
(3) Document title is the most helpful meta data for users make sense of the group without having to read each document.

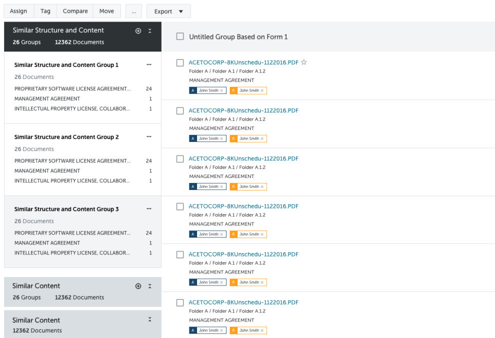


**Figure 4: The final clustering user interface showing the list of documents and document metadata in the same view as the group they belong to.**

## 5.1 Methodology

We asked 8 users (6 project managers, 2 reviewers) from 3 firms to test our new interface. We asked users to conduct three tasks (i.e., examine categories, perform a self-assigned task, negotiate) in a think-aloud fashion and interviewed them afterwards to validated our hypotheses. Unlike the previous phase, we had them use a simulated diligence collection of 103 documents to expedite the testing process and ensure all users had a consistent experience.

## 5.2 Results

One of our primary findings in this phase is that user roles *may* influence how they interact with the similarity categories but is not an absolute. Only one of the reviewers said that they would use "Similar Structure and Content" clusters to identify near duplicates. Similarly, only a slight majority of the project managers said they would use "Similar Content" to assign users to review documents. At best, these trends encourage us to continue to investigate how role influences the usage of this feature once it is deployed.

Contrary to our hypothesis, negotiation appears to be a useful ability for project managers. We posit that this may be due to the fact that clusters can act as "starting points" for assigning reviewers. Traditionally project managers will assign documents to users based upon the unreliable folder structure present in their document collections. Thus, starting with an "almost correct" cluster and negotiating may result in a net gain as they are able to assign documents to users in a more efficient manner.

From this phase and the previous phase, we noticed a trend to want to assign users to documents that are not necessarily similar but are *related* (e.g., amendments, signed versions). While at a high-level, these documents may be similar (e.g., referring to the same contract or deal) they textually look and feel different. Handling these cases is not something the feature currently does. It also highlights that there is no definitive analysis that one can conduct on due diligence document collections.

# 6 DISCUSSION

## 6.1 Matching Representation to Mental Models

Our study has demonstrated that effective designs should reflect users' mental models for due diligence as has been shown for UX design at large [11, 13]. Designers should take the time through generative research to understand their users' jargon, concepts, and goals. Furthermore, we have found that to provide an accurate and useful representation, we must first understand how our mental models and that of our users differ. While we may use the same words, they have different meanings. Accordingly, our different similarity categories evolved from our growing understanding of the different similarity-based tasks our users would like to undertake.

## 6.2 Adapting Interface Based on Tasks

In due diligence, the tantamount goal is to identify whether a particular transaction would put a client at risk. This is a high-stress situation for all involved and users are generally less likely to tolerate sensemaking tasks that require high costs. Efficiency then becomes a primary motivator for lawyers which means that they are more inclined to prefer obviously structural visual representations (e.g., *"This format isn't helpful since circles are misleading. Bars are more simple and straight-forward"*). Once we realized that the the primary task for clustering would be for time and project management, it was much easier to refine the design to meet those needs in a way that was clear to lawyers.

## 6.3 Trust and negotiation

Throughout the first phase, users continually expressed the idea that they could not "trust" a black-box algorithm. We later determined that this appeared to be a result of a gap between their mental model and how we were representing the data. Once we were able to align our representation to their model, users were more willing to tolerate the algorithm even if they still did not know how it worked. Consequently, they were willing to tolerate mistakes from the algorithm when given the option to negotiate with it and potentially correct those mistakes. Thus, even when the algorithm does not meet their expectations out-of-the-box, they are able to correct it in a meaningful way with low cost.

# 7 CONCLUSION

This paper has presented an overview of our user-centric approach to designing a document-level clustering tool for due diligence. We have shown that aligning a tool to the intended mental model and tasks of envisioned users is crucial for establishing trust. Furthermore, we have found that when trust is established, users are more willing to tolerate and accommodate flaws in the feature. It becomes more efficient to negotiate with the feature than it does not to use it. While our study is relatively small, we have found that involving users early in the design process and iterating will yield a substantially better product than delivering only an end product that was developed in relative isolation. Further work will be done as additional feedback is gathered after the feature launch. We anticipate that much of our subsequent work will continue to refine the design to further increase trust in and utility of the feature.

# ACKNOWLEDGMENTS

The authors would like to thank all of the users and members of our Legal Knowledge Engineering team who took time out of their busy schedules to help facilitate this work.

# REFERENCES


[1] Youn ah Kang, John Stasko, and Senior Member. 2012. Examining the Use of a Visual Analytics System for Sensemaking Tasks: Case Studies with Domain Experts. (2012).

[2] Bilal Alsallakh, Luana Micallef, Wolfgang Aigner, Helwig Hauser, Silvia Miksch, and Peter Rodgers. 2016. The State-of-the-Art of Set Visualization. *Computer Graphics Forum* 35, 1 (2016).

[3] Simon Attfield and Ann Blandford. 2010. Discovery-led refinement in e-discovery investigations: sensemaking, cognitive ergonomics and system design. *Artif. Intell. Law* 18, 4 (2010).

[4] Jason Chuang, Daniel Ramage, Christopher D. Manning, and Jeffrey Heer. 2012. Interpretation and trust: Designing model-driven visualizations for text analysis. In *SIGCHI '12*.

[5] Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. [n. d.]. Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. ([n. d.]).

[6] Stephen Few and Perceptual Edge. 2007. Data visualization: past, present, and future. *IBM Cognos Innovation Center* (2007).

[7] Erika Hall. 2013. *Just Enough Research*. A Book Apart.

[8] K. Kucher and A. Kerren. 2015. Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*.

[9] Gloria J Leckie, Karen E Pettigrew, and Christian Sylvain. 1996. Modeling the information seeking of professionals: A general model derived from research on engineers, health care professionals, and lawyers. *The Library Quarterly* 66, 2 (1996), 161–193.

[10] Stephann Makri, Ann Blandford, and Anna L Cox. 2008. Investigating the information-seeking behaviour of academic lawyers: From EllisâĂŹs model to design. *Information Processing & Management* 44, 2 (2008), 613–634.

[11] Jakob Nielsen. 2010. Mental Models. (2010). https://www.nngroup.com/articles/mental-models/

[12] Don Norman. 2011. *Living with Complexity*. The MIT Press.

[13] Don Norman. 2013. *The Design of Everyday Things*. Basic Books.

[14] Daniel M. Russell, Malcolm Slaney, Yan Qu, and Mave Houston. 2006. Being literate with large document collections: Observational studies and cost structure tradeoffs. (2006).

[15] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. [n. d.]. Proceedings of InterCHI 1993 ABSTRACT The Cost Structure of Sensemaking. ([n. d.]).

[16] Bahareh Sarrafzadeh, Alexandra Vtyurina, Edward Lank, and Olga Vechtomova. 2016. Knowledge Graphs versus Hierarchies: An Analysis of User Behaviours and Perspectives in Information Seeking. In *CHIIR*.

[17] Mithileysh Sathiyanarayanan and Cagatay Turkay. 2017. Challenges and Opportunities in using Analytics Combined with Visualisation Techniques for Finding Anomalies in Digital Communications. In *ICAIL DESI VII Workshop*.

[18] David van Dijk, David Graus, Zhaochun Ren, Hans Henseler, and Maarten de Rijke. 2015. Who is Involved? Semantic Search for E-Discovery. In *ICAIL DESI VI Workshop*.