Redesigning a Document Viewer for Legal Documents

Adam Roegiest
Kira Systems
adam.roegiest@kirasystems.com

Winter Wei
Kira Systems
winter@kirasystems.com

ABSTRACT

In Mergers and Acquisition due diligence, lawyers are tasked with analyzing a collection of contracts and determine the level of risk that comes from a merger or acquisition. This process has historically been manual and resulted in only a small fraction of the collection being examined. This paper reports on the user-focused redesign of our document viewer that is used by clients to review documents and train machine learning algorithms to find pertinent information from these contracts.

We present an overview of the due diligence task and the user stories, generated through analysis of support tickets, user interviews, and usability testing sessions, that we used to redesign our document viewer to accommodate the variety of workflows that our clients employ. Additionally, we detail the important design decisions made and discuss the implications of our redesign beyond our particular use case.

CCS CONCEPTS

• **Information systems** → **Search interfaces**; *Information extraction*; *Structured text search*; • **Human-centered computing** → *Usability testing*;

KEYWORDS

Due diligence, document viewer, legal retrieval, usability study

1 INTRODUCTION

In the due diligence process of Mergers and Acquisitions, lawyers are tasked with extracting and analyzing content from contracts (e.g., start date, value, what happens when one party is bought) to determine any potential risks that would result from a merger or acquisition. Due to the manual effort historically involved, a very small sample of the contracts is analyzed, and senior lawyers then extrapolate to the entire collection. There have been several cases

where a failure in due diligence had substantial fiscal repercussions [1]. One of the most costly was HP's acquisition of Autonomy for $10B USD and the subsequent $8.8B loss reported by HP, which is largely believed to result from poor due diligence on HP's part regarding Autonomy's worth.

Accordingly, there is continued growth [2] of software companies aiming to target the due diligence use case with the goal of improving the review process (i.e., increasing efficiency and effectiveness). A crucial component to any such system is a full-featured document viewer, which allows users to view contracts, highlight relevant text, and summarize text that the system has identified as important to their use case.

In this paper, we describe the design and evolution of our company's document viewer. The original document viewer was based on the past experiences of company employees (former lawyers) and was largely created in an ad hoc manner. This resulted in the original document viewer having serious flaws: including, summary panels of extracted information overlaying the document; inability to zoom; hard to distinguish overlapping highlights; lack of contextual document navigation (i.e., navigation only took folder structure into account).

Through a combination of direct customer feedback, user interviews, and usability testing with clients, we developed a series of personas, each having different goals when using the document viewer. Using these personas, we developed a set of job stories (i.e., particular tasks to be completed) which formed the basis of our design and development process. With all this in mind, we present the evolution of our document viewer and the rationale behind these decisions. The result is, to the best of the our knowledge, one of the first attempts to qualitatively design an interface for document review in legal information retrieval. Indeed, we believe that the iterative design process used herein would benefit the design of UI and UX for other legal retrieval tasks.

In the following section, we place our work in context and distinguish the due diligence task from related legal retrieval tasks. Section 3 describes the old document viewer and the user stories, generated through various interactions with clients, that were the motivation behind the redesign. We then (Section 4) provide a description of the new document viewer and how it was designed and developed to meet these user needs. Section 5 discusses user feedback to the new document viewer, collected during and after the design process. We also provide a new set of user stories based on our goal of continual improvement to the document viewer. We conclude (Section 6) with a discussion of this work in the context of legal retrieval and the takeaways from this work.

[1] See https://www.firmex.com/thedealroom/top-10-due-diligence-disasters/. Accessed October 1, 2017.
[2] See https://www.artificiallawyer.com/2017/09/15/legal-ai-co-s-seal-kira-leverton-show-buoyant-growth. Accessed October 1, 2017.

(a) The document pane displays an image representation of each page in the document and overlays any highlights that have been made on top of the corresponding text.
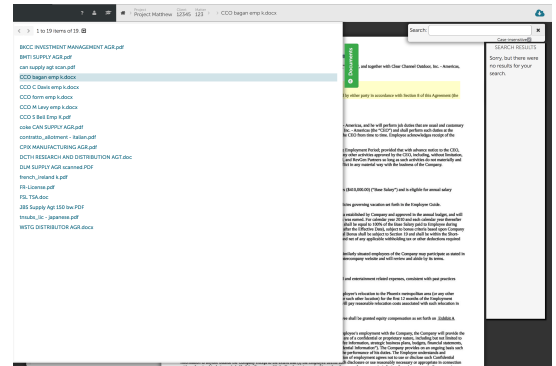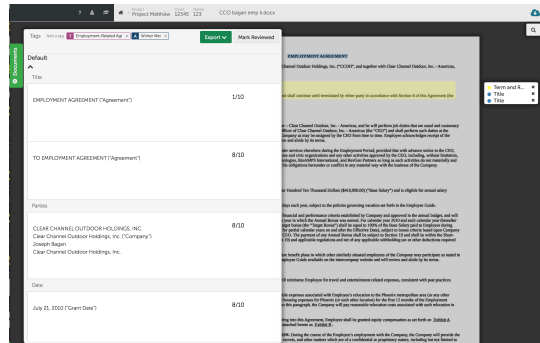

(b) The document navigation tab allows users to move from the current document to any document within the same directory. Navigation through other contexts (e.g., search results) required using browser history.

(c) The annotation summary pane displays highlighted text from the document grouped by topic with "jump to page" functionality.

Figure 1: Examples of the three panes of the original document viewer: the document pane, the summary (of annotations) pane, and the document navigation tab.

2 BACKGROUND

To the best of our knowledge, there has been little scientific research done on the due diligence problem; most published work is primarily for lawyers and other practitioners in legal journals [3, 7, 8]. This is in contrast to the large amount of work that has gone into the related field of electronic discovery from the scientific and legal communities (c.f., [2, 6]). While it is not clear why this is the case, we might speculate that the amount of data required to be processed and the stakes in failing to do it well, have traditionally been much higher in electronic discovery. More concretely, electronic discovery seeks to find all relevant material relating to a particular topic in hopes of finding a "smoking gun"; whereas, due diligence seeks to analyze risk inherit in a particular merger or acquisition through the extraction of relevant information from contracts, bills, etc. Furthermore, electronic discovery is usually a regulated process but due diligence is limited primarily by the effort that lawyers and their clients want to invest in the process.

Electronic discovery has had much more exposure in the scientific community but much of this discusses the trade-offs between manual versus technology-assisted review, the efficacy of

algorithms, or ways to measure performance [2, 4, 10]. Very little research has focused on what makes a good review platform and the human factors involved. While there has been work on sense-making in electronic discovery [1, 5, 9], which deals with understanding the document collection (e.g., through clustering or email threading), there has been little work done in tying this to a review workflow or platform.

Accordingly, the work we present herein is not only applicable to our particular use case but may help inform others in designing systems for related legal high-recall tasks that involve a component of document review. In particular, it would be an interesting study to see how much of the user interface presented here would be applicable to an electronic discovery use case.

3 OLD DOCUMENT VIEWER AND USER STORIES

Our original document viewer (Figure 1) had three constituent components: (a) a document pane that allowed users to read the documents themselves and annotate particular portions of text as being relevant to a particular need; (b) a summary pane that allowed users to view manually and automatically generated annotations;

and, (c) a document listing tab that allowed users to navigate from one document to the next using a directory structure. As seen above, these three components overlay each other in undesirable ways, creating numerous usability issues.

Additionally, it is worth noting that the document pane displayed a rendered image (to promote supporting a wide variety of file formats during ingestion) and overlaid transparent HTML on top of the image's text to facilitate highlighting and extraction. This approach, while effective, suffered from a serious flaw in that documents could not be zoomed or rotated since similar actions could not be performed on the HTML elements. While not a deal-breaker, such inability meant that users occasionally had to struggle when viewing poor quality scans of documents.

The old document viewer was filled with other usability issues including: the inability to toggle the active highlight when several overlapped, the inability to navigate through searched documents without returning to the search, a reliance on the browser's built-in search to look for keywords in a document, and difficulty in getting the summary and document panes to align correctly in the browser or swap between them.

Based upon customer feedback (e.g., support tickets, interviews, and usability sessions) on the original document viewer, we formulated the following user stories for how customers want to use our application in various scenarios:

- When I review the documents assigned to me, I want to go through the list without disruption, so I can keep in the flow.
- When I am reviewing documents, I want to read the document instantly once the document viewer is opened, so I can finish my job faster.
- When I read the documents that contains landscape and portrait pages, I want to zoom in to the document or rotate the pages, so I can read the text easily.
- When I read a document, I want to search by keywords as they are usually a good indication of where the important clauses are, so I can read and review them.
- When I review the summary, I want to only review the ones I care about so I can save time.

4 NEW DOCUMENT VIEWER

As several of our user stories were impeded by technical constraints in the old document viewer, the new document viewer was built from scratch. This allowed us to take completely different technical approaches to address these issues rather than trying to retrofit existing components. This also included changing how the document viewer fit users' mental model. For example, in opening a document, instead of being redirected to a new page, the document opens on top of the current page and its context, and the user can "exit" the document by closing it (i.e., the 'X' in Figure 2). When this feature was tested, experienced users often took a few seconds to realize they could utilize this functionality, while new users immediately took advantage of the new functionality.

Figure 2 depicts our new document viewer which maintains the document and summary panes from the original document viewer. However, they no longer overlap and are able to have their screen share adjusted or be collapsed. The choice to have an adjustable size for the two panes is relatively straightforward in that it allows users

to customize the document viewer to their particular workflow (e.g., annotating documents versus reviewing annotations). Swapping the order of the panes was done to align more with how users have typically described using the viewer (i.e., document first) and with other web applications where comments/notes appear on the right-hand side (e.g., Google Docs). It is worth noting that the navigation tab was removed and replaced with navigation arrows (top right) which are modally dependent (e.g., using a search context) and fit more closely to the previous/next document workflow.

One of the main benefits of the new document viewer is that the user is able to zoom in on the document and still annotate text [3]. This was accomplished by utilizing the virtual DOM present in modern Web frameworks to tag portions of the image with associated metadata (e.g., bounding boxes for words) which facilitates the annotation under zoom with relatively simple maths.

Features like text annotation under zoom required several data migrations to ensure all necessary metadata was available in the virtual DOM. On the other hand, features such as overlapping highlights, were improved simply through better colour selection, improved CSS usage, and allowing the desired highlight to be brought to the forefront.

We followed an iterative design and development process to help determine what wasn't working for users and refine what was. To this end, we conducted 6 user testing sessions with a combination of in-house annotators, our support team (who have a strong understanding of user pain points), and mid-level associates during different stages of development.

5 RESPONSE TO THE REDESIGN

Generally, the reception to the new document viewer has been highly positive. One of the best examples of this occurred during the phased roll-out: a client firm had the new document viewer turned on a week prematurely, requested it to be turned off, then, three hours later, asked to have it turned back on due to internal user requests for it.

One of the glaring flaws that has persisted in measuring improvement to the document viewer, is a lack of instrumentation to determine how users *actually* use the viewer, primarily due to a lack of resources being allocated to the task.

While we lack quantitative feedback from the document viewer itself, we have some idea of the impact of the redesign on our clients and internal support team. Figure 3 depicts the number of support tickets submitted (pre- and post-redesign) since the company's inception. The release of the new document viewer was roughly August 2017, where we see a precipitous drop in document viewer issue tickets. The general downward trend is an indicator to us that many of the long standing issues have been addressed.

Of course, this isn't to say that all of the feedback we've received has been positive. Indeed, the following are some of the issues that users have had with the new document viewer:

- When I read a document that is too small, I want to *precisely* control the granularity of zoom so I can read the document comfortably according to my level of comfort.

---

[3]Rotation issues were solved by ensuring the rotation of the pages was consistent during the ingestion and OCR phase of document processing.
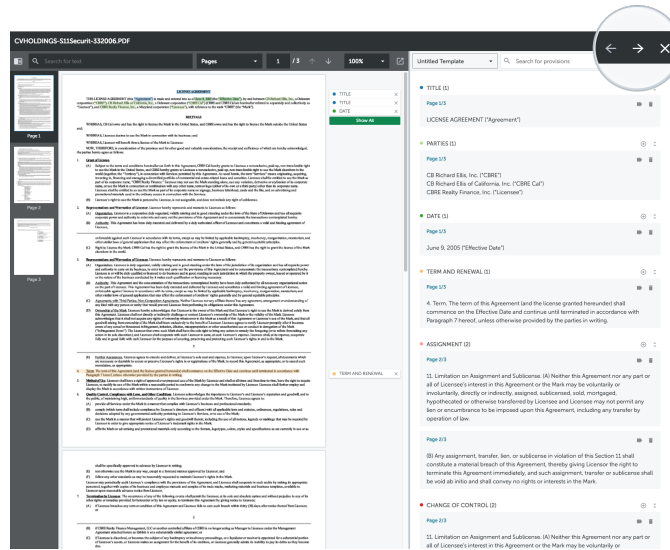
Figure 2: The new document viewer with the document pane on the left, the annotation summary on the right, and the navigation buttons (previous document, next document, and exit document) are magnified on the top-right bar.

- When I organize the summary, I want to re-order the results of certain provisions, so they fit the template I was given.
- When I have to manually add summary results, I want to do that in an intuitive and more discoverable way, so that the process is seamless.

6   DISCUSSION AND CONCLUSION

As part of this redesign it became increasingly obvious to us that when designing the UI and UX for complex workflows that the user is a necessary component throughout the design process. Trying to model the user's mental state when performing their daily tasks and the struggles that arise from conflicting workflows is crucial to making a usable and pleasurable design. While there will always be some trade-offs between focusing on micro- versus macro-interactions, the end goal ought to be an effective user experience.
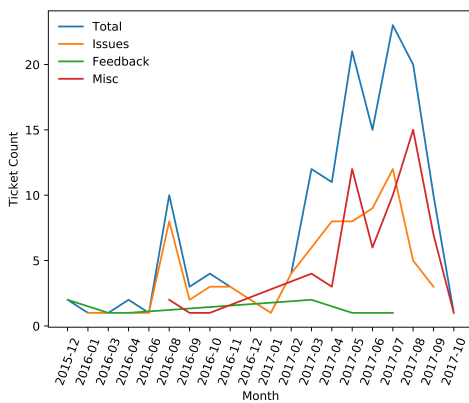

Figure 3: The number of tickets submitting to the support line regarding the document viewer over time as a total number of tickets, identified issues, feedback tickets, and tangential tickets to the document viewer.

While there has been substantial research on user experience in Web search and related tasks, there has been little on complex workflows that encompass legal retrieval and discovery tasks, like electronic discovery and due diligence. Accordingly, there was little to guide us in our initial development and design of the document viewer, but by utilizing user research we have been able to more fully understand and empathize with our users. Ensuring that they get their job done as efficiently and effectively as possible means they're more likely to continue to use our software and speak positively about it. Based upon interactions with our users, we have made great strides towards this but as with any design there is always more to be done.

REFERENCES

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

[2] G. V. Cormack and M. R. Grossman. Evaluation of Machine-Learning Protocols for Technology-Assisted Review in Electronic Discovery. In *SIGIR 2014*, 2014.

[3] B. Klaber. Artificial Intelligence and Transactional Law: Automated M&A Due Diligence, 2013.

[4] D. W. Oard, J. R. Baron, B. Hedin, D. D. Lewis, and S. Tomlinson. Evaluation of Information Retrieval for E-Discovery. *Artif. Intell. Law*, 18(4), 2010.

[5] M. Sathiyanarayanan and C. Turkay. Challenges and Opportunities in using Analytics Combined with Visualisation Techniques for Finding Anomalies in Digital Communications. In *ICAIL DESI VII Workshop*, 2017.

[6] K. Schieneman et al. The Implications of Rule 26 (g) on the use of Technology-Assisted Review. *Fed. Cts. L. Rev. 239*, 2013.

[7] J. A. Sherer, T. M. Hoffman, and E. E. Ortiz. Merger and Acquisition Due Diligence: A Proposed Framework to Incorporate Data Privacy, Information Security, E-Discovery, and Information Governance into Due Diligence Practices. *Rich. JL & Tech.*, 21, 2015.

[8] J. A. Sherer, T. M. Hoffman, K. M. Wallace, E. E. Ortiz, and T. J. Satnick. Merger and Acquisition Due Diligence Part II-The Devil in the Details. *Rich. JL & Tech.*, 22, 2016.

[9] D. van Dijk, D. Graus, Z. Ren, H. Henseler, and M. de Rijke. Who is Involved? Semantic Search for E-Discovery. In *ICAIL DESI VI Workshop*, 2015.

[10] W. Webber. Re-examining the Effectiveness of Manual Review. In *Proc. SIGIR Information Retrieval for E-Discovery Workshop*, 2011.