Research Statement

Abram Handler • University of Colorado, Boulder • December, 2021
https://www.abehandler.com/

Large document collections are key resources in many areas of qualitative research. For instance, historians inspect newspaper archives to answer fundamental questions about societies [11], marketers review social media posts to understand consumer behavior [3], and journalists investigate document leaks to report news stories [22]. Yet what is sometimes called *making sense* [19] of large document collections (also called *corpora*) is often challenging. Human time and attention are limited, and in many settings, there often many more documents than a person can possibly read. Therefore, in my research, **I develop new natural language processing (NLP) methods to create new intelligent interfaces for helping people make sense of large text corpora**. There are two complementary and intertwined threads to my work:

- **Developing new NLP methods.** I develop new NLP methods for intelligent interfaces. For instance, in one recent paper I introduced the VertexAddition algorithm [7], which efficiently summarizes single sentences for display in query-based text analytics tools. This thread of my research uses computational research methods and is aimed at NLP audiences. I often evaluate using traditional automatic metrics like latency, precision and recall.
- **Developing new intelligent interfaces**. I also apply new NLP methods to create new user interfaces, designed to help people in their work. For instance, I created the Rookie system which combines traditional visual analytics methods like brushing and linking [12] with text summarization techniques to help journalists learn about new topics (Figure 5). In this thread of my work, I employ research methods from human–computer interaction (HCI), such as studying user needs, developing prototypes, conducting user interviews and measuring human performance [15].

These two threads of my research are united by my interest in using artificial intelligence (AI) to augment rather than replace human cognition. This human-centered perspective on AI has deep roots in computer science [4], and is a focus of much recent research concerned with by developing intelligent systems for people [1]. In my own work, I have found that uniting NLP and HCI is helpful because user interfaces shape, constrain and reveal NLP problems (Section 1), and because focusing on people helps refine NLP tasks (Section 2). In the future, I will continue to pay close attention to user needs, which I believe to be central to conducting high-impact research (Section 3).

1 Developing new NLP methods

As an NLP researcher, I develop new NLP methods for new intelligent user interfaces. From my perspective, NLP and HCI are deeply linked. Because people ultimately interact with language technologies via specific user interfaces, the details of how such interfaces work, and the details of what people expect them to do, often shape, constrain and reveal NLP problems.

For instance, in one paper [10], I focused on NLP for ConceptMap interfaces, which describe relationships between notable terms in a corpus, using a directed graph. ConceptMap interfaces require automatically-generated natural language text that can describe relationships between terms in a corpus. This generated text is displayed along directed arrows on the graph (Figure 1, right). By focusing on how to generate such text, I found that the assumptions underlying baseline NLP methods for building machine-readable knowledge bases were poorly suited to summarizing relationships for people. This mismatch motivated development of a new NLP method to summarize relationships for ConceptMaps [6], based on extracting readable spans from longer sentences (Figure 1). Focusing on ConceptMap interfaces thus showed me limitations of existing NLP methods, and guided my development of an NLP technique.

I had similar experiences when examining the established NLP problem of sentence simplification, in which the goal is to create a summary of a long sentence. I became interested in simplification because I wished to build new intelligent interfaces (e.g. ClioQuery [11]) which could create corpus summaries consisting of shortened sentences, based on a user's query term. Yet when I investigated existing simplification techniques,

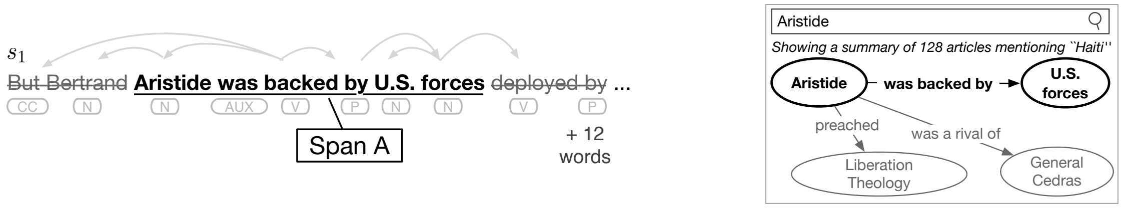

Figure 1: I research new NLP methods for new intelligent interfaces. For instance, I developed the *relationship summarisation* method (left) for use in ConceptMap interfaces (right). Relationship summarization [6] uses linguistic information (shown as curved arrows and rounded rectangles, left) to predict if a sequence of words within a sentence expresses a relationship between a pair of query terms. For instance, Span A (bold, top left) in the sentence s_1 expresses a relationship between the terms "Aristide" and "U.S. Forces". Because Span A expresses such a relationship, it can be extracted from s_1 and shown on the ConceptMap (right).

I found that the only known NLP methods which could meet the needs of these new interfaces required worst-case exponential time, which would likely be too slow in user-facing settings [17]. I thus proposed a new approach to sentence simplification which iteratively built up a shortened sentence, by making predictions about a grammatical representation, called a dependency parse [7]. Because the method ran in linear rather than worst-case exponential time, I observed an 11x speedup over the baseline (Figure 2). In a separate project on sentence simplification [9], I collaborated with an expert in psycholinguistics to examine how people perceived different possible sentence shortenings. I then used a model of such perceptions to guide new query-oriented sentence simplification techniques. This work was partially funded through an "AI for Everyone" grant from the company Figure Eight, Inc.

I have also investigated other NLP problems, motivated by the needs of user interfaces. For instance, I helped develop an NLP method for finding phrases based on part-of-speech tags [8], for display in user interfaces like my Rookie tool [5]. This method has gone on to have significant impact in research and industry. At the time of writing, software for the project has gathered over 150 stars on GitHub, from programmers in Canada, China, Ireland, Singapore and the United States.

- `https://github.com/slanglab/phrasemachine`

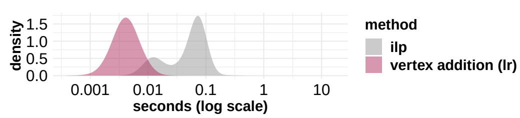

Figure 2: Density plot of observed latencies for my linear VertexAddition method and a baseline method using integer linear programming (ILP).

2 Developing new intelligent interfaces

In addition to developing new NLP techniques, I also apply new NLP methods in new intelligent interfaces, designed to help people in their work. For instance, I developed the ClioQuery system for historians and archivists by following a user-centered and "problem-driven" research process [21]. I first conducted semi-structured needfinding interviews to understand the problems and work of my target expert users, and then worked from my understanding of such needs to develop a series of iterative prototypes. Once I settled on a final system, I conducted evaluation interviews to test ClioQuery "in the wild," by deploying the system and observing its utility for "real users," working on "real problems" using "real data" [21]. The final version of the tool employed text simplification methods I developed for NLP audiences, to create a skimmable summary of each occurrence of a query word across a corpus (Figure 3). In a qualitative evaluation, historians explained how these features helped with actual historical research, and identified limitations for future work.

In another similar project, I created the Rookie system [5] to help journalists learn about new topics. Rookie combines text summarization with traditional visual analytics techniques like brushing and linking [12] to create interactive summaries of unfolding events through time (Figure 5). I began working on

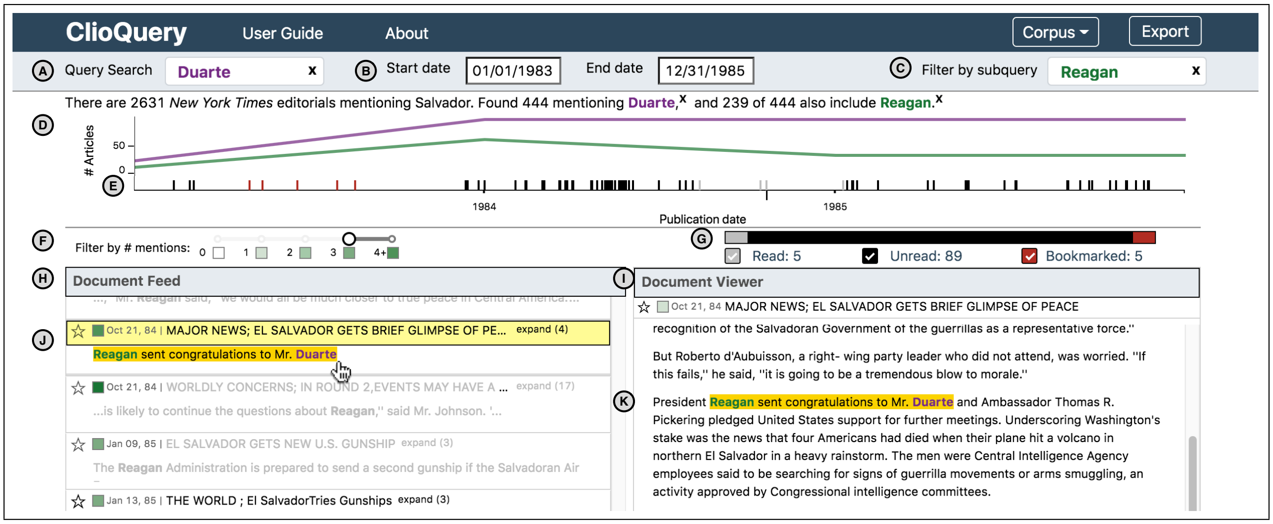

Figure 3: ClioQuery, an interactive text analytics system for helping historians investigate queries in news archives. Features (letters A to K) include: (D) a Time Series View showing the frequency of a user's query through time, (H) a linked Document Feed showing a skimmable query-oriented summary of every mention of the query in the corpus, and (I) a linked Document Viewer showing a selected news story, with text from the query-oriented summary highlighted in yellow. The paper describes each lettered feature in detail [11].

Rookie while working as a computational journalist before starting my PhD, and secured funding for the project through a grant from the Knight Foundation's Prototype Fund.

While working on Rookie and ClioQuery, I collected quantitative measurements of human performance, to complement qualitative feedback. For instance, to evaluate ClioQuery, I asked crowdworkers to perform a historical reading comprehension task, modeled on a real question from a historian. I randomly assigned participants to complete the task using ClioQuery or a baseline information retrieval (IR) system [18], and measured the total number of correct questions for participants in each group. I found that ClioQuery helped people answer significantly more comprehension questions than the IR baseline. In a similar evaluation, I found that the Rookie system helped Rookie users complete a historical sensemaking task more quickly than users assigned to complete the same task using an IR tool (Figure 4).

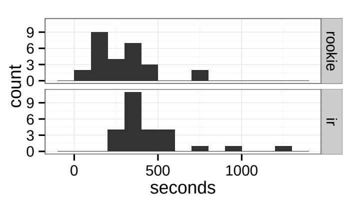

Figure 4: Human task completion times for a historical sensemaking task. Rookie helps people answer historical questions more quickly than an IR baseline.

In my experience, I have found that designing, evaluating and discussing new corpus interfaces provides insights into more traditional method-based NLP research. For example, while working on Rookie, I learned that summary quality was far less important than system speed. Rookie is effective in part because it can update summaries in response to mouse interactions in near real time. (Lags are known to hinder user performance [17]). Similarly, while working on ClioQuery, I found that some historians needed to be able to trust, understand and interpret summary output before they could use summaries in their work [11, Sec. 9]. Yet traditional study of summarization in NLP typically ignores system latency and summary trustworthiness, and instead evaluates with the recall-based ROUGE metric against "gold" references [2]. Such experiences with summarization helped me learn that testing language technologies with actual users is a crucial complement to traditional methods-based study in NLP. Working with people can refine and direct NLP tasks, by revealing and challenging assumptions ingrained in purely computational research.

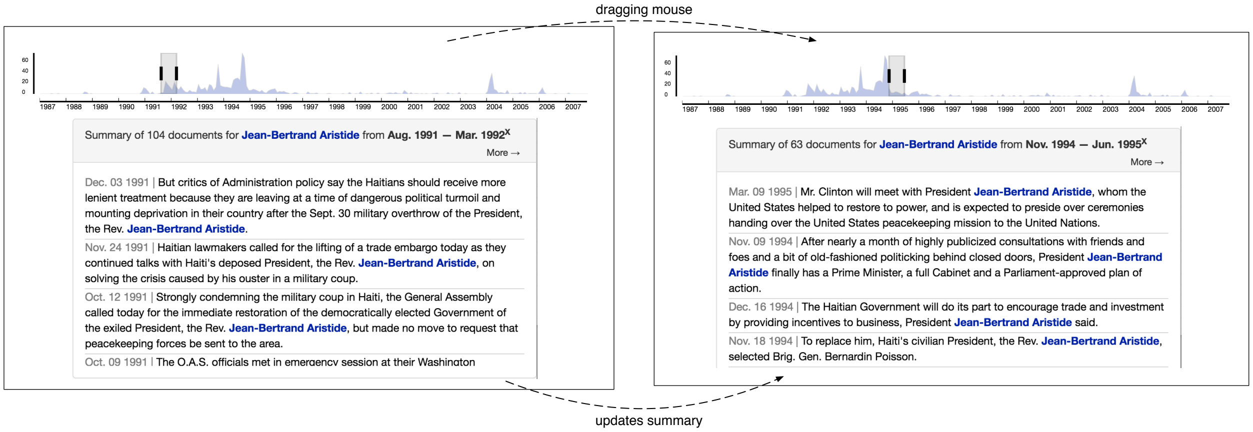

Figure 5: My Rookie system combines traditional visual analytics techniques like brushing and linking [12] with text summarization to help people make sense of news archives. As users drag a cursor across the timeline, a summary updates to show unfolding events.

3 Future work

In the future, I plan to continue investigating new NLP methods and new intelligent interfaces. I am particularly excited to continue to explore three themes from my current work.

New modalities. In the past, I have learned that user interfaces shape and constrain NLP problems (Section 1), and that new kinds of interfaces can enable powerful new language technologies (Section 2). I am thus excited to explore NLP for new interface modalities in future work. For instance, I am currently working on an audio podcast application designed to help people learn new languages. The tool will automatically translate words and phrases from an English language podcast into audio clips in a second language, using machine translation and text-to-speech technologies. Because people acquire new vocabulary in part from contextual cues [13], I plan to automatically intersperse these second-language clips with original English-language audio to create bilingual podcasts. Thus far, I have created a working prototype, and am exploring different possible approaches to sound design, in collaboration with an audio researcher and podcaster. I am particularly excited about this project because the number of potential users (anyone who wants to learn a language) is very large. The effort is inspired by prior work on generating bilingual text [20].

Text visualization and text summarization. In the past, researchers have proposed many systems and methods for helping people make sense of large document collections. Often, these tools adopt a data visualization approach, by considering how to map high-dimensional text to a two-dimensional graphical format [14]. Yet in my work on Rookie and ClioQuery, I began to explore a different paradigm, inspired by other new work in text visualization [23]. Instead of mapping text to graphics, Rookie and ClioQuery show spans of text, extracted from a corpus, which are then displayed in an interactive interface (Figure 5). Future efforts might consider how to combine this extractive approach, based on summarization in NLP, with more traditional data visualization techniques. For instance, we might use NLP methods (e.g. Zhang et al. [24]) to annotate time series plots of word frequency (e.g. Figure 5) with textual summaries of events.

Interdisciplinary collaboration. In the past, I have collaborated with historians, journalists, psychologists, podcasters, HCI researchers and experts in business analytics. I strongly believe that such work across disciplines reveals new problems and possibilities, and is crucial to developing computational systems and methods that people want and use. During graduate school, my advisor also helped me learn to see computer science (CS) as a helper discipline, much like statistics. While I recognize that CS offers powerful tools and ways of thinking, for me, computational methods are most exciting when put to use helping other researchers and practitioners realize the vast potential of digital technology in their own work. I also gravitated towards CS, in part, because I was interested in so many things that I had trouble picking a single area of focus. CS seemed like a good fit for me because everything is now somehow related to computers. Ultimately I like being a CS researcher for much the same reason Tukey reportedly liked statistics. "*The best thing about being a statistician*," Tukey once said [16], "*is that you get to play in everyone's backyard.*"

References

[1] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz. Guidelines for human-AI interaction. In *CHI*, 2019.

[2] D. Das and A. F. Martins. A survey on automatic text summarization. Technical report, Carnegie Mellon University, 2007.

[3] N. Glance, M. Hurst, K. Nigam, M. Siegler, R. Stockton, and T. Tomokiyo. Deriving marketing intelligence from online discussion. In *KDD*, 2005.

[4] S. Green, J. Heer, and C. D. Manning. Natural language translation at the intersection of AI and HCI: Old questions being answered with both ai and hci. *Queue*, 2015.

[5] A. Handler and B. O'Connor. Rookie: A unique approach for exploring news archives. In *Workshop on Data Science + Journalism at KDD*, 2017.

[6] A. Handler and B. O'Connor. Relational summarization for corpus analysis. In *NAACL*, 2018.

[7] A. Handler and B. O'Connor. Query-focused sentence compression in linear time. In *EMNLP*, 2019.

[8] A. Handler, M. Denny, H. Wallach, and B. O'Connor. Bag of what? Simple noun phrase extraction for text analysis. In *Proceedings of the First Workshop on NLP and Computational Social Science*, 2016.

[9] A. Handler, B. Dillon, and B. T. O'Connor. Human acceptability judgements for extractive sentence compression. *ArXiv*, 2019. URL `https://arxiv.org/pdf/1902.00489`.

[10] A. Handler, P. Ganeshkumar, B. O'Connor, and M. AlTantawy. Summarizing relationships for interactive concept map browsers. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, 2019.

[11] A. Handler, N. Mahyar, and B. O'Connor. ClioQuery: Interactive query-oriented text analytics for comprehensive investigation of historical news archives. In *Under submission*, 2021.

[12] J. Heer and B. Shneiderman. Interactive dynamics for visual analysis. *Commun. ACM*, 2012.

[13] S. Krashen. We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4):440–464, 1989.

[14] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *PacificVis*, 2015.

[15] J. Lazar, J. H. Feng, and H. Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[16] D. Leonhardt. John Tukey, 85, statistician; coined the word 'software'. *The New York Times*, 2000. URL `https://www.nytimes.com/2000/07/28/us/john-tukey-85-statistician-coined-the-word-software.html`.

[17] Z. Liu and J. Heer. The effects of interactive latency on exploratory visual analysis. *IEEE TVCG*, 2014.

[18] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, USA, 2008.

[19] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, volume 5, 2005.

[20] A. Renduchintala, R. Knowles, P. Koehn, and J. Eisner. Creating interactive macaronic interfaces for language learning. In *ACL System Demonstrations*, 2016.

[21] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE TVCG*, 18(12), 2012.

[22] J. Stray. What do Journalists do with Documents? In *Computation+Journalism Symposium*, 2016.

[23] N. Sultanum, A. Bezerianos, and F. Chevalier. Text Visualization and Close Reading for Journalism with Storifier. *IEEE TVCG*, 2021.

[24] C. Zhang, S. Soderland, and D. S. Weld. Exploiting parallel news streams for unsupervised event extraction. *TACL*, 3, 2015.