Rapid Understanding of Scientific Paper Collections: Integrating Statistics, Text Analytics, and Visualization

Cody Dunne[1,3*], Ben Shneiderman[1,3], Robert Gove[1,3], Judith Klavans[2] and Bonnie Dorr[2,3]

[1]Human-Computer Interaction Lab, [2]Computational Linguistics and Information Processing Lab, [3]Department of Computer Science
University of Maryland, College Park, MD 20742.
E-mail: {cdunne, ben, rpgove, bonnie}@cs.umd.edu, jklavans@umd.edu
[*]Corresponding author

Abstract

Keeping up with rapidly growing research fields, especially when there are multiple interdisciplinary sources, requires substantial effort for researchers, program managers, or venture capital investors. Current theories and tools are directed at finding a paper or website, not gaining an understanding of the key papers, authors, controversies, and hypotheses. This report presents an effort to integrate statistics, text analytics, and visualization in a multiple coordinated window environment that supports exploration. Our prototype system, Action Science Explorer (ASE), provides an environment for demonstrating principles of coordination and conducting iterative usability tests of them with interested and knowledgeable users. We developed an understanding of the value of reference management, statistics, citation context extraction, natural language summarization for single and multiple documents, filters to interactively select key papers, and network visualization to see citation patterns and identify clusters. The three-phase usability study guided our revisions to ASE and led us to improve the testing methods.

Introduction

Contemporary scholars and scientists devote substantial effort to keep up with advances in their rapidly expanding fields. The growing number of publications combined with increasingly cross-disciplinary sources makes it challenging to follow emerging research

fronts and identify key papers. It is even harder to begin exploring a new field without a starting frame of reference.

Researchers have vastly different levels of expertise and requirements for learning about scientific fields. A graduate student or cross-disciplinary researcher in a new field might find it useful to see the pivotal historical papers, key authors, and popular publication venues. On the other hand, a seasoned academic may be interested only in recent leading work and outlier papers or authors that challenge their preconceptions about the field. Grant program managers and review panel members sometimes have to examine fields they are not familiar with, looking for research trends, emerging theories, and open questions. Moreover, social scientists or scientometric analysts may be interested in how academic communities form over time, comparing citation and publication trends by country, or tracking the adoption of a single innovation.

Tools for rapid exploration of the literature can help ease these difficulties, providing readers with concise overviews tailored to their needs and aiding the generation of accurate surveys. Digital libraries and search engines are useful for finding particular papers or those matching a search string, but do not provide the additional analysis tools required to quickly summarize a field. Users unfamiliar with the field often find it challenging to search out the influential or groundbreaking papers, authors, and journals.

Specialized tools compute statistical measures and rankings to help identify items of interest, and other tools automatically summarize the text of multiple papers to extract key points. However, these tools are decoupled from the literature exploration task and are not easily integrated into the search process. Visualization techniques can be used to provide immediate overviews of publication and citation patterns in a field, but are uncommon in literature exploration tools. When present, they usually do not display much data (e.g., only ego networks) or provide the interaction techniques required to analyze the publication trends and research communities in a field. More ambitious goals for visualizations include helping users reach sufficient understanding to enable decision-making, such as which fields are promising directions for researchers, appropriate for increased/reduced funding by government or industrial program managers, or worthy of investment by a venture capital organization.

This paper presents the results of an effort to integrate statistics, text analytics and visualization in a powerful prototype interface for researchers and analysts. The Action Science Explorer (ASE) [1] is designed to support exploration of a *collection* of *papers* so as to rapidly provide a summary, while identifying key papers, topics, and research groups. ASE uses 1) bibliometric lexical link mining to create a citation network for a field and context for each citation, 2) automatic summarization techniques to extract key points from papers, and 3) potent network analysis and visualization tools to aid in the exploration relationships. ASE, shown in Fig. 1, presents the academic literature for a field using many different modalities: tables of papers, full texts, text summaries, and visualizations of the citation network and the groups it contains. Each view of the underlying data is coordinated such that papers selected in one view are highlighted in the others, providing additional metadata, text summaries, and statistical measure rankings about them. Users can filter by rankings or via search queries, highlighting the matching results in all views. Specifically,

---

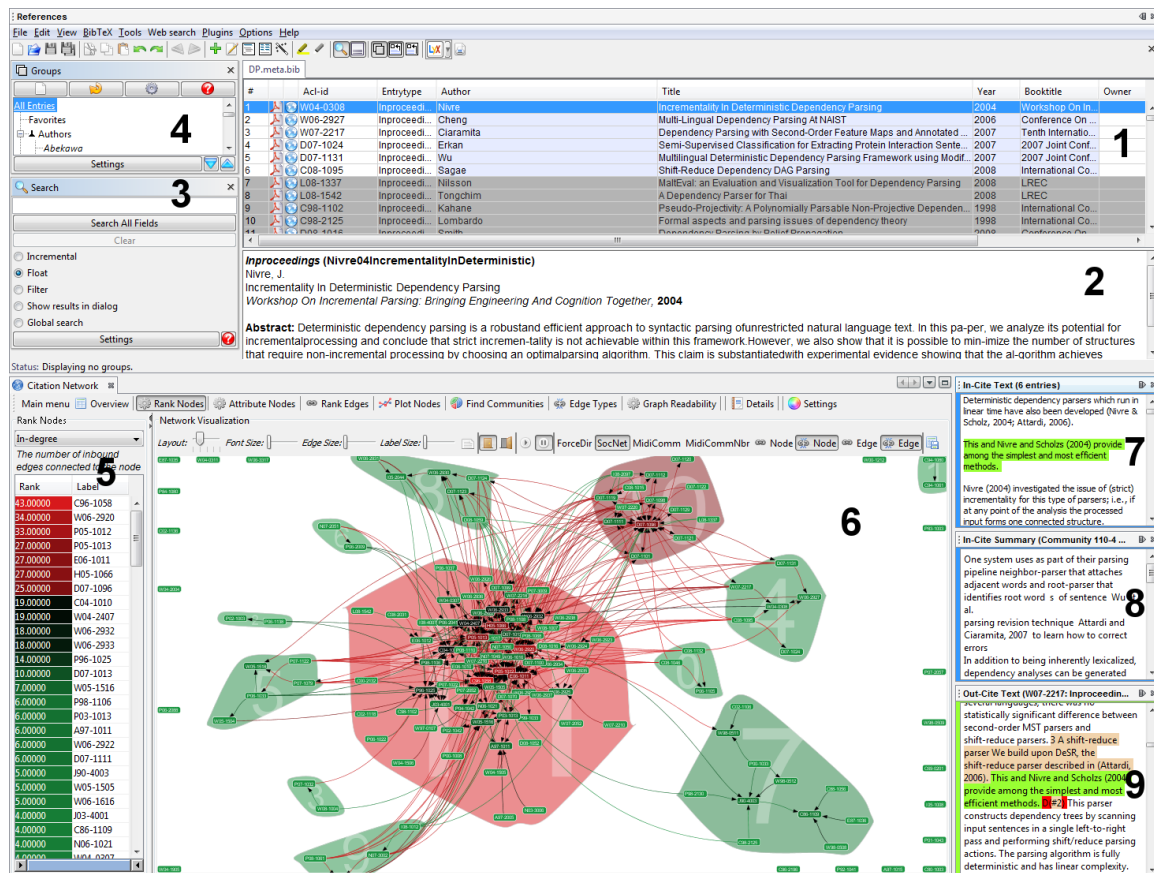[1] For videos and more information visit http://www.cs.umd.edu/hcil/ase

Figure 1. : The main views of ASE are displayed and labeled here: Reference Management (1–4), Citation Network Statistics & Visualization (5–6), Citation Context (7), Multi-Document Summaries (8), and Full Text with hyperlinked citations.

the contributions of this paper are:

- A discussion of the motivation for creating a prototype literature exploration tool, its sophisticated design, and the challenges involved;

- The novel integration of visualization with text analysis of citation context and multi-document summarization;

- Results from early evaluations which demonstrate the effectiveness of our multiple coordinated view design for exploring collections of papers; and

- A set of user requirements and evaluation recommendations for future systems we realized through our evaluations.

Related Work

To accomplish the goals laid out in the introduction, a complete system needs to support a variety of services. Initially users would search a large collection and import the relevant papers to deepen their understanding of the desired scientific field. Most research database systems support searching the collection and return a list of papers, but only a

few provide sufficiently powerful tools to explore the result set. Natural operations would be to sort and filter the result set by time, author name, institutions, key phrases, search term relevance, citation frequency, or other impact measures. These help users to identify the key papers, researchers, themes, research methods, and disciplinary links as defined by publication venue.

As users invest time to gain familiarity with individual papers, they study the list of authors, read the abstracts, scan the content, and review the list of citations to find familiar papers, authors, and journals. Another source of insight about a paper is to see how later papers describe it and to see what other papers are cited concurrently. Studying such citation contexts is a fruitful endeavor, but is difficult in most systems.

After studying 5–50 papers users usually begin to understand the field, key researchers, consistent topics, controversies, and novel hypotheses. They may annotate the papers, but more commonly they put them into groups to organize their discovery process and facilitate future usage. Accelerating the process of gaining familiarity would yield enormous benefits, but a truly helpful system would also improve the completeness, appropriateness, and value of the outcome.

Once users have gained familiarity they may dig deeper to understand the major breakthroughs and remaining problems. Breakthroughs and problems are rarely spelled out explicitly as a field is emerging, although review papers that look back over a decade or two are likely to contain such insights. Reading citation contexts is helpful for gaining insights into the field, but can be time consuming even in a well-designed system and might give only a narrow focus. Visual analytics can help in this task by offloading some of this effort to users' perceptual processes, which excel at finding patterns and relationships in the high data density of visualizations (National Visualization and Analytics Center, 2005). Visualizations group information together in compact representations, reducing search space, and simplify and organize information through abstractions and aggregation.

Standard charts like ranked lists and scatterplots can provide interesting views of a collection, but citation network visualizations expose a different set of relationships entirely. Network visualizations have been only marginally effective in the past, but improved layout, clustering, ranking, statistics, and filtering techniques have the potential for exposing patterns, clusters, relationships, gaps, and anomalies. Users can quickly appreciate the strength of relationships between groups of papers and see bridging papers that bring together established fields. Even more potent for those studying emerging fields is the capacity to explore an evolutionary visualization using a temporal slider. Temporal visualizations can show the appearance of an initial paper, the gradual increase in papers that cite it, and sometimes the explosion of activity for "hot" topics. Other temporal phenomena are the bridging of communities, fracturing of research topics, and sometimes the demise of a hypotheses.

Techniques of natural language processing can speed up the analysis of a large collection by extracting frequently occurring terms/phrases, identifying topics, and identifying key concepts. Multi-document summarization and document clustering have the potential to help users by providing some forms of automated descriptions for interesting subsets of a collection.

Since accomplishing these complex tasks in a single scrolling window is difficult, many systems provide multiple coordinated windows that enable users to see lists or visualizations in one window and make selections for displays in other windows. A more advanced

technique is brushing and linking, which allows selection in one display to highlight related items in another display.

Existing systems provide some of these features in various combinations, though none allow users to leverage all of them in a single analysis. For a comparison of the capabilities of several common systems, see Table 1 of Gove, Dunne, Shneiderman, Klavans, and Dorr (2011). This table shows 13 capabilities we have identified as important for literature exploration tools, and their support by current search engines, reference managers, and summarization and recommendation techniques.

For their initial exploration, users frequently use academic search tools like Google Scholar (Google, 2011) and Microsoft Academic Search (Microsoft Research, 2011). Subscriber-only general databases are used frequently at universities and research labs, such as ISI Web of Knowledge (Thomson Reuters, 2011b) and SciVerse Scopus (Elsevier, 2011). Additionally, many field-specific databases exist such as PubMed (National Center for Biotechnology Information, 2011) for Life and Biological Sciences. Computer and Information Sciences have databases like the web harvesting CiteSeer (Giles, Bollacker, & Lawrence, 1998; Bollacker, Lawrence, & Giles, 1998), arXiv (Cornell University Library, 2011) for preprints, and the publisher-run ACM Digital Library (Association for Computing Machinery, 2011) and IEEE Xplore (Institute of Electrical and Electronics Engineers, 2011).

These search tools and databases generally provide a sortable, filterable list of papers matching a user-specified query, sometimes augmented by faceted browsing capabilities and general overview statistics. Some enable users to save specific papers into groups to review or export later, though this is via a separate interface and annotation is not usually supported. ISI Web of Knowledge is rare in that it includes a visualization, showing an ego network of an individual paper including both incoming and outgoing citations. However, it is a hyperbolic tree visualization that has little dynamic interaction. Furthermore, visualizations are most useful for finding overall trends, clusters, and outliers–not for looking at small ego network subsets.

An emerging category of products called reference managers enhances these paper management capabilities by supporting additional search, grouping, and annotation features, as well as basic collection statistics or overview visualizations. Some examples are JabRef (JabRef Development Team, 2011), Zotero (Center for History and New Media, 2011), EndNote (Thomson Reuters, 2011a), and Mendeley (Mendeley Ltd. 2011).

Many academic databases now use citation extraction to help build the citation network of their paper collections for bibliometric analysis, and some such as CiteSeer (Giles et al., 1998; Bollacker et al., 1998) and Microsoft Academic Search (Microsoft Research, 2011) expose the context of those citations. The benefit of showing citation context is that readers can quickly learn about the critical reception, subsequent and similar work, and key contributions of a paper as seen by researchers later on. Analyses of paper collections from citation text has also been demonstrated to be useful for a wide range of applications. Bradshaw (2003) used citation texts to determine the content of papers and improve the results of a search engine. Even the author's reason for citing a given paper can be automatically determined (Teufel, Siddharthan, & Tidhar, 2006).

Natural language processing techniques for document and multi-document summarization can produce distilled output that is intended to capture the deeper meaning behind

a topically grouped set of papers. Citation texts have been used to create summaries of single papers (Qazvinian & Radev, 2008; Mei & Zhai, 2008). Nanba and Okumura (1999) discuss citation categorization to support a system for writing surveys and Nanba, Abekawa, Okumura, and Saito (2004) automatically categorize citation sentences into three groups using pre-defined phrase-based rules. Other summarization approaches exist for papers (Teufel & Moens, 2002) or news topics (Radev, Otterbacher, Winkel, & Blair-Goldensohn, 2005). For a cogent review of summarization techniques, see Sekine and Nobata (2003).

Academic research tools apply bibliometrics to help users understand collections through network visualizations of paper citations, author collaborations, author or paper co-citations, and user access patterns. Network Workbench (NWB Team, 2006) provides an impressive array of statistics, modeling, scientometric, and visualization algorithms for analyzing bibliometric datasets. Another tool designed for analyzing evolving fields is CiteSpace (Chen, 2004; Chen, 2006; Chen, Ibekwe-SanJuan, & Hou, 2010), which is targeted at identifying clusters and intellectual turning points. Similarly, semantic substrates can be used for citation network visualization (Aris, Shneiderman, Qazvinian, & Radev, 2009), showing scatterplot layouts of nodes to see influence between research fronts. Unfortunately these visualizations are weakly integrated into the rest of the exploration process and are yet to be widely used.

Part of the challenge of integrating visualizations effectively is making them visible concurrently with the search result list. Effective designs would move from the traditional single scrolling windows to *multiple coordinated views* that support brushing and linking to highlight related items (North & Shneiderman, 1997). The power of a spatially stable overview and multiple detail views is especially appropriate for browsing large collections of papers. However, many bibliometrics tools that present several views of the collection would benefit from better integration, easier linking, and common user interfaces across windows. For example, Network Workbench (NWB Team, 2006) is a collection of tools from different providers whose interface design and workflow strategies sometimes require extra work on the part of users. The many useful visualizations of Network Workbench provide little user interaction, no linking between visualizations, and a diverse array of independent interfaces.

Existing theories of information seeking are helpful for reminding us of process models that start from identifying the goal and end with presenting the results to others (Hearst, 2009). One example is Kuhlthau's six stages: initiation, selection, exploration, formulation, collection, and presentation (Kuhlthau, 1991). Marchionini (1997) describes an 8-stage process in his early book, and offers a richer model in his more recent descriptions of exploratory search (Marchionini, 2006). These and other information-seeking processes (Bates, 1990) provide a useful foundation for the complex task of enabling users to understand emerging fields. This complex task also benefits from theories of sense-making and situation awareness, since the goal is to understand multiple aspects of emergent fields such as the key papers, authors, controversies, and hypotheses. A related goal is to understand the relation to other fields which could be sources of insight and fields which have parallel or duplicate results that are not recognized. A further goal is to determine which topics have the greatest potential for advancing a field, thereby guiding researchers, program managers at funding agencies, or venture capital investors who see commercial potential.

Evaluating complex creativity and exploration tools can be challenging. The scope of the features used and the intellectual effort required for exploration render quantitative

laboratory techniques infeasible for capturing many important aspects of the tool usage (Bertini, Perer, Plaisant, & Santucci, 2008; Chen & Czerwinski, 2000). One way that individual tools can be analyzed and compared with others is based on the insights into the data users find with them (Saraiya, North, and Duca, 2005; Saraiya, North, Lam, and Duca, 2006). Alternatively, Shneiderman and Plaisant (2006) make the argument that qualitative evaluation methods are becoming common, accepted, and effective techniques for analyzing visual analytics tools. Examples of these techniques are demonstrated by Seo and Shneiderman (2006) and Perer and Shneiderman (2008).

ASE Design

The goal of Action Science Explorer (ASE) is to help analysts rapidly generate readily-consumable surveys of emerging research topics or fields they are unfamiliar with, targeted to different audiences and levels. The design of an effective literature exploration tool is complex and requires significant thought about which techniques to use to display the collection, how to arrange the screen space to minimize distracting window manipulation and occluding overlaps, and how to use rich forms of brushing and linking to produce relevant highlights in related windows. The philosophy of our design is to integrate statistical, visual, and text representations that are each relevant to the task of scientific literature exploration. All of these modalities are linked together in multiple coordinated views, with brushing and linking such that any selection in one is reflected in the others. We hope the design and ideas we demonstrate with ASE will provide inspiration for designers of many similar commercial and research tools.

This section describes the design and various features of ASE, which is illustrated in Fig. 1, in addition to the challenges we encountered in its creation. For more technical details and discussion about the challenges we faced with data processing and text summarization, see the later Implementation Details section.

Search & Data Import

ASE builds on familiar literature exploration interfaces: the search engines and databases often used when conducting literature reviews. A typical ASE session begins with a keyword, phrase, or topic search of a database to define a target corpus that is retrieved and processed. In our examples, we use the 147 papers returned by a search for "Dependency Parsing" a collection of 17,610 Computational Linguistics papers from the ACL Anthology Network (AAN)[2] (Radev, Muthukrishnan, and Qazvinian, 2009b; Radev, Joseph, Gibson, and Muthukrishnan, 2009a). The AAN includes a network of the citations between papers as well as the full text of each paper, its metadata, abstract, references, and citation sentences.

Reference Management

The search results are loaded into ASE and displayed using the JabRef reference manager (JabRef Development Team, 2011) component, shown in Fig. 1 (1–4). This provides users with a table of papers and their bibliographic data (1), from which the URL or DOI, full text PDF, plain text, and any other files for each paper can be opened. The reference

---

[2]http://clair.si.umich.edu/clair/anthology/

version of the selected paper is shown along with its abstract and any user-written annotations (2), and additional metadata can be shown or entered by double-clicking on an entry. The table can be sorted by column and searched using regular expressions (3), and papers can be organized into hierarchical overlapping groups (4).

As the underlying data structure used by JabRef is the BibTeX bibliography format, ASE can be easily used in conjunction with LaTeX and, with the appropriate plugins, Microsoft Office or OpenOffice.org. Moreover, there are numerous export filters to copy selected entries to websites, other formats, or tools to allow rapid sharing of findings and easy import into survey writing software.

Citation Network Statistics & Visualization

Once analysts have reviewed the data using standard reference management techniques, they can view visualizations of the citation network of the papers in the SocialAction network analysis tool (Perer & Shneiderman, 2006) (Fig. 1 (5–6)). Using these visualizations of the citation network we can easily find unexpected trends, clusters, gaps and outliers. Additionally, users of visualizations can immediately identify invalid data that is easily missed in tabular views. For example, if there are large disconnected or loosely connected components in a citation network visualization it may mean that the imported search query matched several unrelated concepts (or independent research groups). Similarly, ranked list and scatterplot visualizations of node attributes easily show empty numerical data coded as '-1', '999' and the like at the extremes.

The left view (Fig. 1 (5)) shows a ranking of papers by dynamically computed network statistics such as their in-degree, which is the number of citations to that paper within this dataset. This can be switched to the attribute 'InCites', which is the number of citations to that paper within the entire AAN corpus. Additional network statistics include betweenness centrality, clustering coefficient, hubs or authorities, and any numeric attributes of the papers like year or externally computed measures. This ranked list can be filtered using the double-ended slider at its bottom, removing the top- or bottom-ranked papers in the list dynamically from the visualizations.

The papers in the collection can be viewed in standard charts like scatterplots (e.g., Fig. 5) to see trends and outliers, but visualizations of the network topology are more suited to finding research communities and tracking evolution over time. The node-link diagram of the network (Fig. 1 (6)) shows papers as rounded rectangle nodes, colored by their statistic rankings and connected by their citations using spline arrows. The nodes are arranged using a force-directed layout algorithm such that tightly connected nodes are placed in proximity to each other while loosely connected ones move to the extremes. As users filter or group nodes in the visible network the layout algorithm continues to run, updating the layout to reflect any changes. Nodes can also be colored by categorical attributes (not shown here) and users can compare nodes using scatterplots of their statistics. Edges can also be colored using statistical rankings, such as edge betweenness centrality or externally computed measures like citation sentiment analysis. While we focus on citation networks, this approach can be easily extended to visualize other paper collection aspects like co-citation or co-authorship networks.

Papers can be grouped manually or using Newman's fast community-finding heuristic (Newman, 2004), which finds groups of papers that tend to cite each other more often

than external papers. The found communities are shown using colored convex hulls surrounding the group, with the color representing the maximum ranking of any of its entities. For example, in Fig. 1 (6) the central community is shown in bright red to indicate the high number of citations one of its entities (a 1996 paper by Eisner) received. The inter-community spring coefficients for the force-directed layout are reduced by an order of magnitude to separate them visually more than the basic layout would. Community-finding algorithms are most useful when exploring large datasets, though there are at least two meaningful communities shown in our examples, discussed below in Scenario: Dependency Parsing.

Citation Context

The node-link diagram shows users the number of citations to a paper and topology patterns, but it can also be useful to examine the context in which each of those citations were made in the citing paper. The context of citations often includes detailed and descriptive statements about the cited paper (Garfield, 1994) such as a summary, the paper's critical reception, and citations to follow-up papers (Giles et al., 1998; Bollacker et al., 1998).

From the full text of each paper ASE extracts the sentences containing the citations and their locations in the paper. Then, for any selected papers of interest, the context sentences of all citations to them are displayed in the citation context/in-cite text view (Fig. 1 (7)). If several papers are selected, all their context sentences are shown. Each sentence is a hyperlink that, when clicked, displays the full text of its source paper with the citation highlighed in the full text/out-cite text view (Fig. 1 (9)). Users can then see the broader context of the citation when the citing sentence alone is not sufficient.

Moreover, each citation in the full text is colored and hyperlinked to the target papers, allowing users to rapidly view the cited papers' metadata, full text, statistics, and network location while reading. The hyperlinks also provide immediate access to any cited follow-up papers. Each citation sentence is hyperlinked to the first found citation within it, with subsequent citations hyperlinked to indices at the end of the sentence (e.g., the additional citation represented in Fig. 1 (9) as D(#2)). One item for future work is to hyperlink each citation within sentences separately.

Multi-Document Summarization

Viewing the citation context for a paper or its abstract and keywords can give users an idea of its contribution to the field. However, highly-cited papers have too many citations to read through them all (see the scroll bar in Fig. 1 (7) for an example, with the small "thumb" showing how little of the window is visible). Furthermore, when looking at multiple papers selected manually or through the community-finding algorithm it can be difficult to understand the group's key focuses and contributions.

To aid users in these tasks ASE provides automatically generated multi-document summaries for any selected set of papers, shown in Fig. 1 (8). Summaries of the full text of papers can be useful, but citation contexts and abstracts are richer in survey-worthy information. Mohammad et al. (2009) show that summaries based on citation contexts contain crucial survey-worthy information that is not available or hard to extract from

abstracts and the full texts of papers. Likewise, they demonstrate that abstract summaries contain information not present in citation contexts and full texts. For these examples we will focus on citation context summaries instead of using abstracts or full text, though ASE is modular in design and supports showing multiple summaries simultaneously. Showing additional summaries of the abstracts or full texts would help with network analysis tasks by providing another view of the content. Moreover, these summary types would be critical for understanding datasets without an underlying citation network like news articles or recently published papers with few citations. We are limiting our scope to this particular problem space of citation texts so that we could determine the impact of this form of data for our particular application. We expect future work to explore additional inputs, including the content and structure of the input articles.

Among the four summarization techniques compared by Mohammad et al. (2009), the best at capturing the contributions of papers was Multi-Document Trimmer (MDT) (Zajic, Dorr, Schwartz, Monz, & Lin, 2005; Zajic, Dorr, Lin, & Schwartz, 2007), originally designed to summarize news articles. MDT is an extension of the original Trimmer which summarized single news articles (Dorr, Zajic, & Schwartz, 2003; Zajic, Dorr, & Schwartz, 2004). ASE uses MDT to provide summaries of citation context, but because citation sentences have metadata inline, we made some modifications to better handle this data. First, any inline metadata is identified as grammatically part of the sentence (syntactic) or not (non-syntactic). Non-syntactic citations can be easily removed without changing the sentence meaning. Syntactic citations are replaced with uniquely identifying placeholder text, seen as an out-of-vocabulary noun by the parser. After summarization, the metadata is reinserted for clarity. This modification significantly improves confidence scores from Trimmer's parser and generates higher-quality candidate sentences (Whidby, Zajic, & Dorr, 2011). While we show only multi-document summaries in ASE, single-document citation context summaries using Trimmer or Cluster-Lexrank (Qazvinian & Radev, 2008) for highly-cited papers could be easily added to provide another perspective.

Linking the Views

Each window presents a distinct view of the underlying scientific literature, each with its own advantages and disadvantages. While seeing paper metadata and opening the full text is easiest from the reference management view, determining the relationships between them is best done with the network visualization. Each of the data views becomes more powerful when they are tightly coupled together, such that interactions in one are visually reflected in the others. This technique is called *multiple coordinated views* (North & Shneiderman, 1997).

Each of the views in ASE are linked to all the other windows. When users select papers in the reference manager the selection is also highlighted in the citation network visualization and the statistics ranked list. Likewise, the detail views show the papers' abstracts, reviews, reference forms, citation context, and generated summaries. Selecting nodes in the network visualization or any other view performs similarly, highlighting the nodes in all other views and showing their details.

The only exception to this linking is the full text view, which has two planned use cases. Once users bring up the full text view, they may wish to click on the hyperlinked citations within the text as they read. Clicking a citation selects the cited node in each of

the other views, but to prevent users from losing their place does not update the full text view.

The other use for the full text view is to only update when users select a citation in the citation context view to see the surrounding context in the citing paper's full text. We display which mode is currently being used by updating the border color of the view to show how it is currently interacting with others. Green, as seen in Fig. 1 (9), indicates that the full text view is showing the citation context for a selected citation. Blue, on the other hand, means that a citation within the full text has been selected, highlighting the cited paper in each other view.

In some situations screen space may be limited or users may wish to focus on a subset of the views. ASE provides a docking window manager interface that allows users to hide individual windows, resize or rearrange them, or even drag them to separate monitors. Revealing additional views to users as they gain experience can help speed learning for new ASE users.

Limitations

The design of ASE also has many limitations that may undermine its advantages. The multiple windows require a large screen display to be useful and may increase perceptual and cognitive loads as users make selections which cause changes in multiple windows. Moreover, there is currently no undo feature or exploration history view to show or return to previously viewed states. History awareness is an important aspect of visual analytics systems, and integrating history views into the tool can improve task recall, result in more efficient search strategies, and enable asynchronous collaboration between users (Dunne, Riche, Lee, Metoyer, & Robertson, 2012). Also, the integration of existing components in our prototype means that there are differences between interfaces, especially in the consistency of color highlighting, tool bars, and layouts. We worked to reduce these differences as much as possible, though several remain. Moreover, the rich set of data required for all the views of ASE means that preparing a collection for analysis can be time consuming, thereby limiting our flexibility in conducting evaluations. The below section on Implementation Details elaborates on loading alternate datasets as well as computing multi-document summaries.

Scenario: Dependency Parsing

Imagine Karl, a student new to the field of Dependency Parsing (DP). DP is a small field of Computational Linguistics (CL) dedicated to analyzing sentences based on which of their components are dependent on each other. Karl first runs a search on the ACL Anthology Network (AAN) for papers containing "Dependency Parsing", which returns a subset of 147 papers and the citations between them. After loading the dataset ASE displays the initial windows shown in Fig. 2.

The top view of Fig. 2 shows Karl a reference management interface with a table of all the papers matching the search. In the bottom left he can see a statistical overview of the citation network, including the number of nodes and edges, average in- and out-degree of nodes, and the number of unconnected components. In the rest of the bottom half he can see the topology of the citation network in a node-link diagram, with individual papers colored by the number of citations they have received.
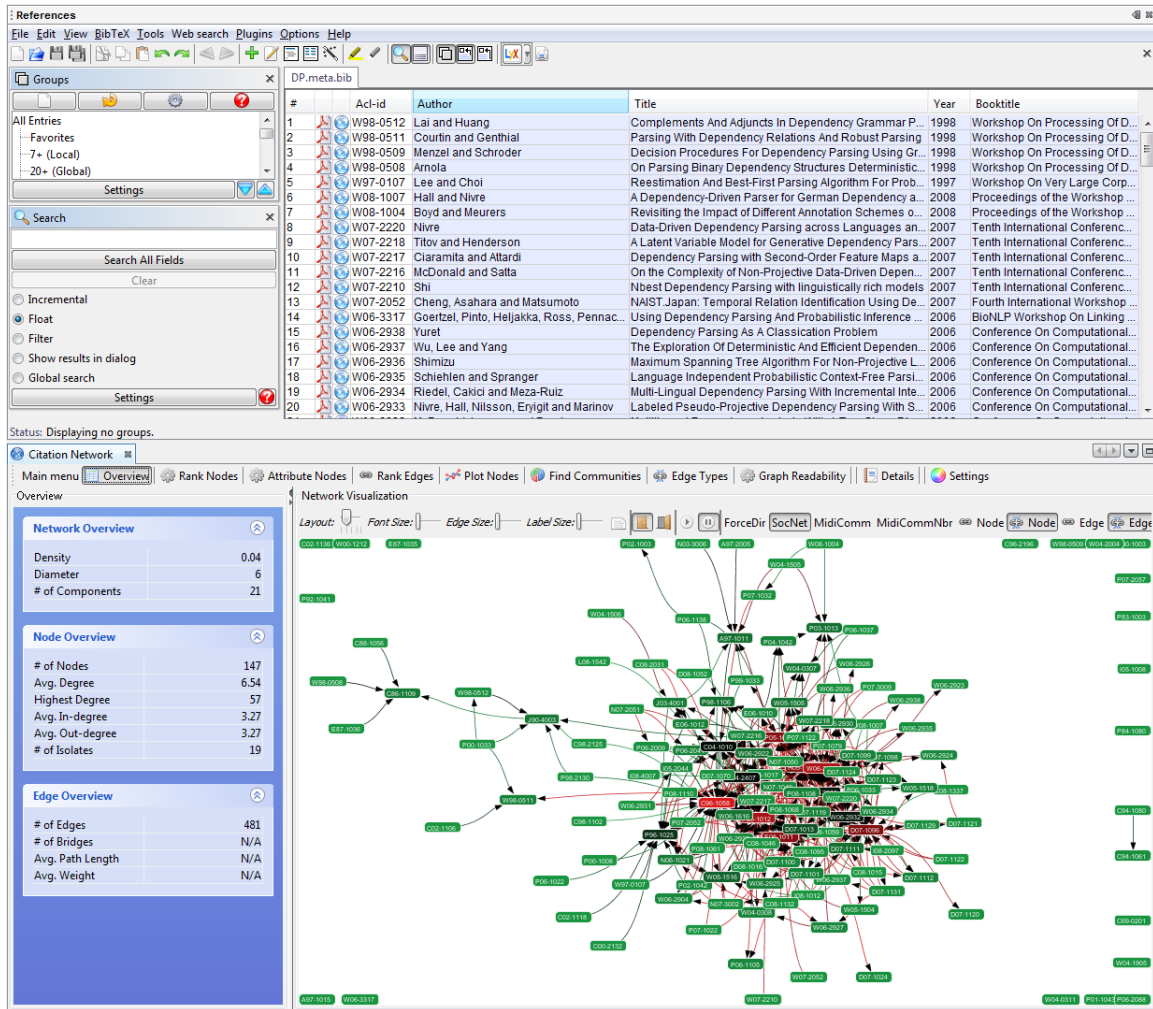
Figure 2. : Starting interface with "Dependency Parsing" (DP) query loaded.

Identifying Key Papers & Authors

Karl is interested in identifying and reading the most influential papers in the field, so he clicks the "Rank Nodes" button to replace the overall statistics window with a list of papers ranked by their in-degree (Fig. 3). The in-degree of a paper is the number of citations it has received from other papers within this subset of the AAN. From here Karl selects all papers cited seven or more times (Fig. 4), and that subset is highlighted in the reference manager (top). He then drags these 14 papers to a group he created in the reference manager to keep track of those results (top left).

Karl quickly notices several things by scanning the table of these highly cited papers. First, all but four of the 14 are written by various combinations of the authors Nivre (6), Nilsson (6), and Hall (3) from Växjö University as well as McDonald (6) and Pereira (4) from University of Pennsylvania. Second, they are all written from 2004–2007, except for
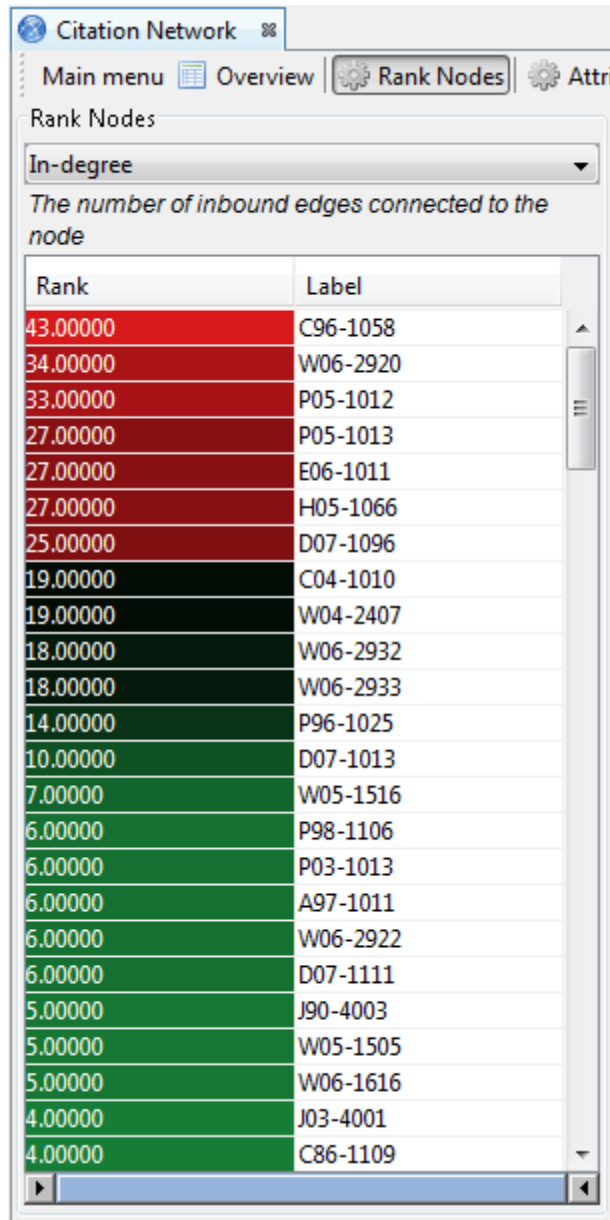
Figure 3. : Ranked list of DP papers by their in-degree (the citation count within this subset).
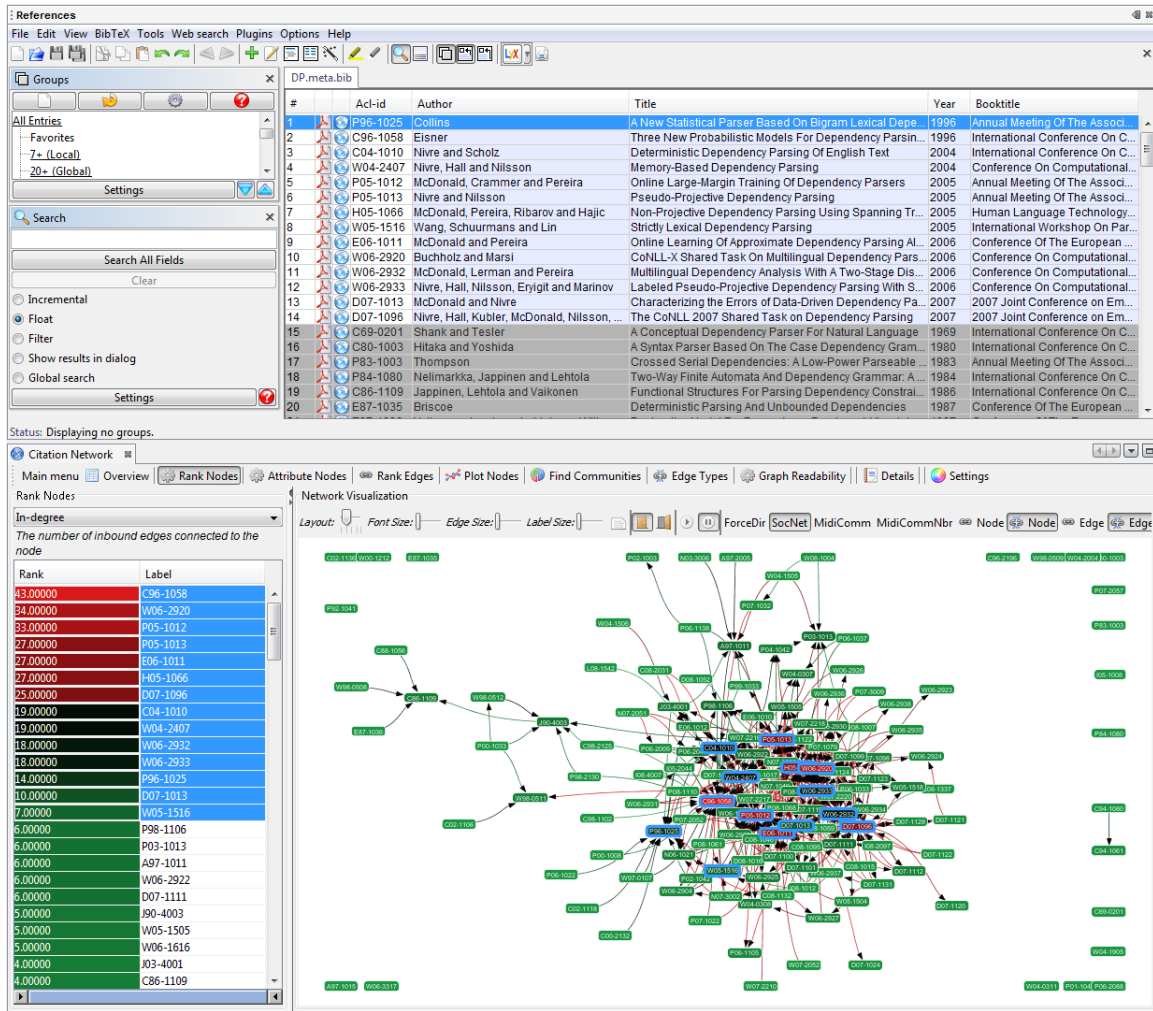
Figure 4. : DP papers with seven or more citations are highlighted in the ranked list (bottom-left), reference manager (top), and node-link diagram (bottom-right).

two written in 1996 separately by Eisner[3] (the most highly cited with 43) and Collins[4] (14). A simple search by author reveals that both Collins and Eisner have additional papers in the dataset, but only in the late 2000s and with few citations.

Karl thinks that he has seen the Collins paper cited before in another field of CL. To compare how many citations it has received among DP papers versus CL papers in general, he creates a scatterplot with the number of citations from DP papers on the horizontal axis and the number of overall citations from CL papers on the vertical axis (Fig. 5). The

---

[3]Eisner, J. M. (1996). *Three new probabilistic models for dependency parsing: an exploration*. In *International conference on computational linguistics*. Retrieved from http://clair.si.umich.edu/clair/anthology/query.cgi?type=Paper&id=C96-1058.

[4]Collins, M. J. (1996). *A new statistical parser based on bigram lexical dependencies*. *Annual meeting of the association for computational linguistics*. Retrieved from http://clair.si.umich.edu/clair/anthology/query.cgi?type=Paper&id=P96-1025.
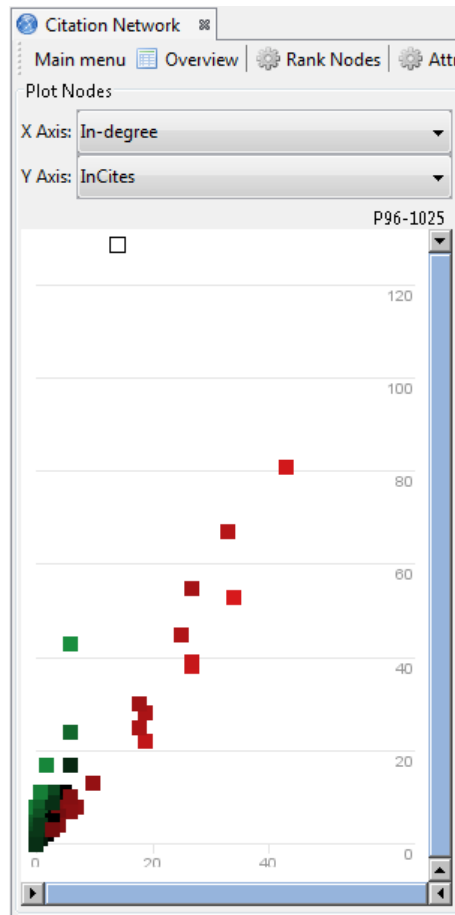
Figure 5. : This scatterplot shows a square for each paper, with the number of citations it receives from papers within the DP subset on the horizontal axis and the number of overall CL citations on the vertical axis. There is a general linear trend, with a green-black-red color scale showing deviations above and below the diagonal. The white box near the top left shows the selected paper by Collins. It is the most highly cited paper in the subset when all CL citations are counted, but when only citations from other DP papers are counted there are several more highly cited papers.
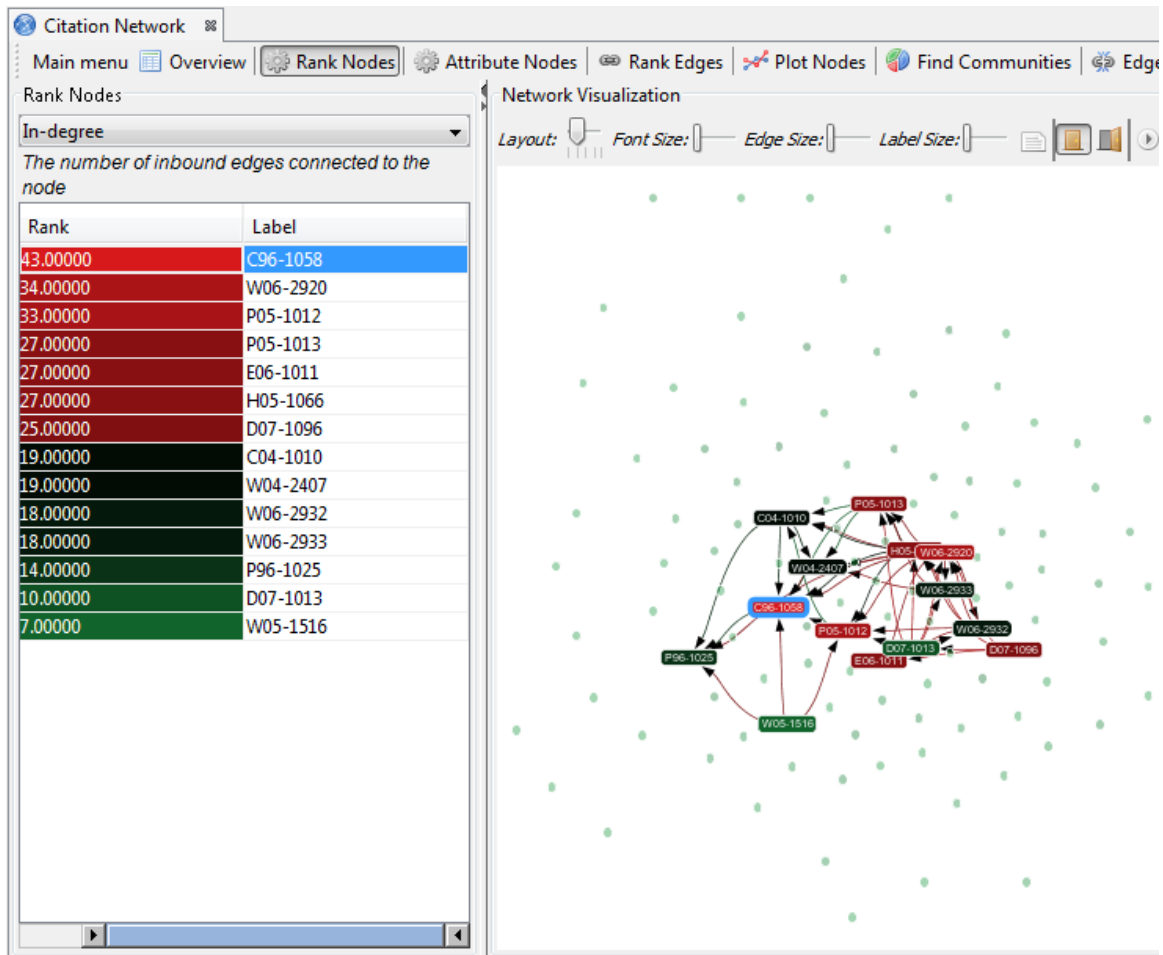
Figure 6. : The ranked list and node-link diagram show only the papers cited more than seven times, filtered using the double-ended slider at the bottom-left.

selected Collins paper is shown with a white square near the top left. It is the most highly cited paper in this subset when all CL citations are considered, but when only citations from other DP papers are counted there are several more highly cited papers.

Karl then wants to see the citation network of only those highly cited papers, so he uses the double-ended slider at the bottom of the ranked list to filter out papers cited less than seven times. The filtered ranked list and citation network visualizations are shown in Fig. 6, and Karl can zoom into it or lay out only the filtered nodes to better see their citation patterns.

Now that Karl has stored a list of interesting papers he starts analyzing them in depth. For each one he selects, the citation context view displays the incoming citations for the paper. After selecting the key Eisner paper and scanning the incoming citations he finds one of particular interest to him: "Eisner (1996) introduced a data-driven dependency parser and compared several probability models on (English) Penn Treebank data" (Fig. 7, bottom). When he clicks on that citation, its surrounding context is displayed in the full
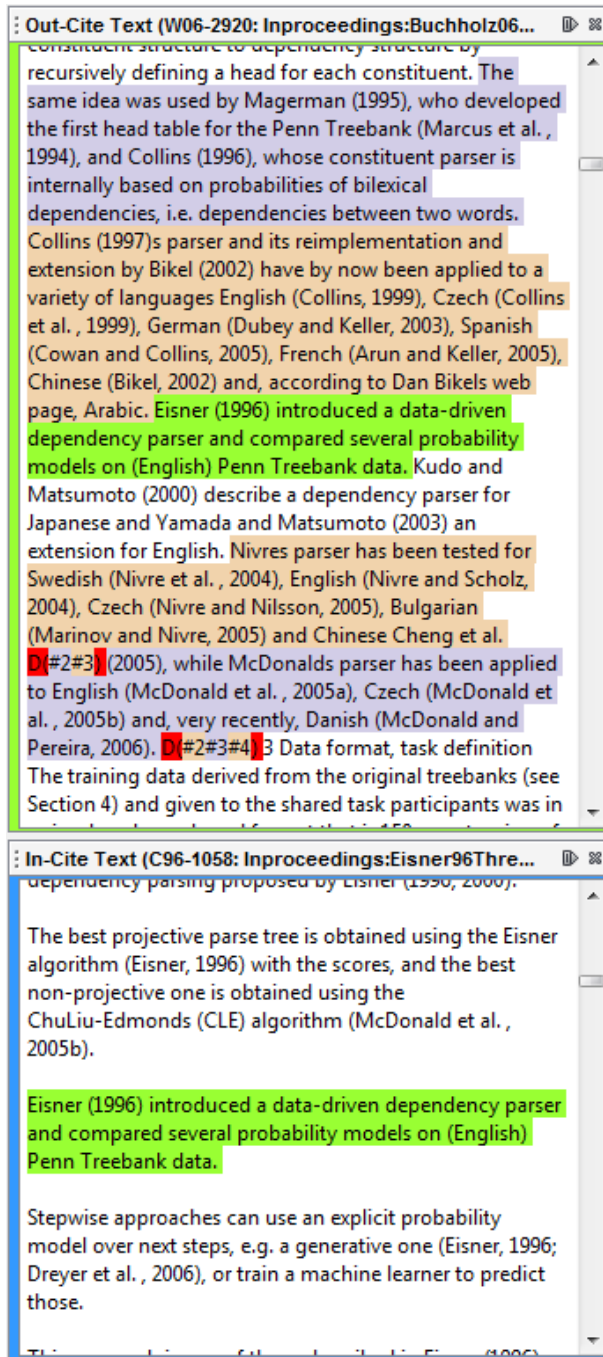
Figure 7. : The citation context for the Eisner paper is shown in the bottom, and the context for the green selected citation is shown above in the full text view.
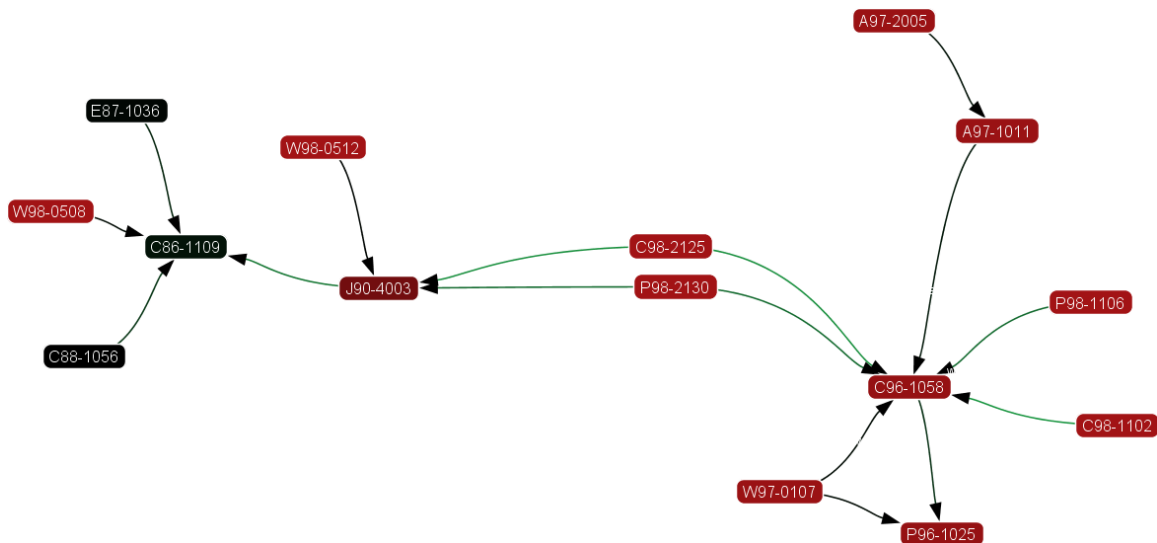
Figure 8. : The connected papers up through 1998 show the original 1986 community on the left and the new 1996 community growing on the right. They are bridged by two duplicate papers in 1998. By filtering to include subsequent years there is an explosion of research focused around the second, right community.

text of the citing paper (top). From here he starts exploring the other hyperlink citations from that paper.

Finally, Karl can view the abstracts for each of those papers and open their full text in his PDF viewer to analyze them in depth. Throughout this process he takes notes in the review field of the reference manager to keep track of his insights.

Tracing the Topic Evolution

Now that Karl has an understanding of the key topics, he wants to trace the evolution of the topic over time. Similar to before, he ranks the papers by the year they were published and uses the double-ended slider to filter out all but the earliest year in the dataset. Then, by slowly dragging the right end of the slider he reveals the papers in the order they were published and the citations between them. He sees the first connected group of papers appearing from 1986–1998, seen in Fig. 8 (left). By CTRL-clicking on each paper, he displays them them in a table in the reference manager and discovers that they center around a research group from the SITRA Foundation in Helsinki, Finland[5].

However, after dragging the slider further Karl sees few papers connected to them in the following years. Starting in 1996, a disconnected group appears beginning with the highly cited Eisner and Collins papers he found in the previous section, which can be seen in the right side of Fig. 8. After filtering up to 1998, two papers (duplicates) by Lombardo and Lesmo[6] appear and cite both the SITRA and Eisner/Collins research communities.

[5]Jappinen, H., Lehtola, A., and Valkonen, K. (1986). *Functional structures for parsing dependency constraints*. *International conference on computational linguistics*. Retrieved from http://clair.si.umich.edu/clair/anthology/query.cgi?type=Paper&id=C86-1109.

[6]Lombardo, V. and Lesmo, L. (1998). *Formal aspects and parsing issues of dependency theory*. *Annual*
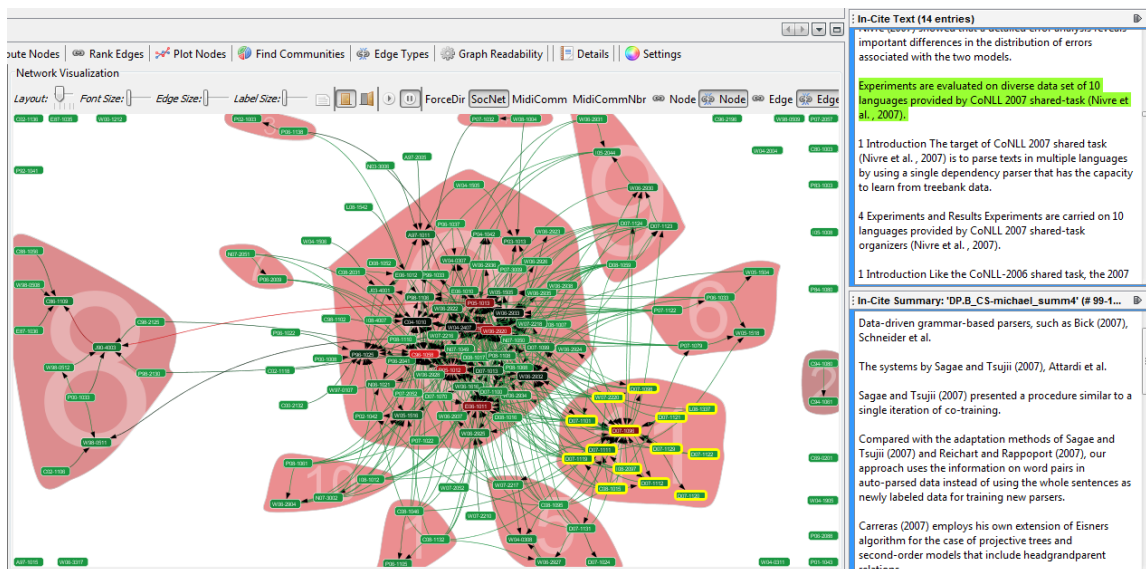
Figure 9. : Algorithmically found communities are shown using convex hulls in the node-link diagram. When selected, all the citation context is shown in the top-right, along with an automatically generated summary of the overall context (bottom-right).

Continuing on, Karl finds that the vast majority of later work in DP is built around the later Eisner/Collins community with few citations to the SITRA group.

During 2006–2008 Karl sees an explosion in research on DP, with approximately 30 papers each year. Sorting the papers in the reference manager by year and scanning their venue, he finds that the bulk of the papers come from the 2006 and 2007 Conference on Computational Natural Language Learning (CoNLL) which both addressed DP.

Exploring Research Communities

As part of the topic evolution analysis Karl found two separate research communities using the force-directed layout and filtering. To more effectively find other communities of interest he decides to use the community-finding algorithm built into ASE. The groups of related papers are surrounded by colored convex hulls (Fig. 9), and he quickly spots the two groups he identified at the left and center.

However, the center core group was split by the community-finding algorithm into several smaller groups that were not obvious before. By clicking on the largest of these (bottom-right & highlighted in yellow), Karl sees the table of papers in it in the reference manager, all the citation context for the cluster (right), and an automatically generated summary of the citation context (bottom-right). He then scans the citations to these papers and sees frequent references to the CoNLL 2007 shared task that he saw before.

Zooming in on the community in the citation network to examine the citation edges, Karl notices that there are many unusual bi-directional citations between a central paper

meeting of the association for computational linguistics and international conference on computational linguistics. Retrieved from http://clair.si.umich.edu/clair/anthology/query.cgi?type=Paper&id=P98-2130.

(Nivre et al.[7]) and other papers in the cluster. By viewing the abstract of the Nivre et al. paper Karl finds the reason for the bi-directional citations: these papers were written collaboratively. The Nivre et al. paper provides an overview of the shared task for the year and the datasets used. It also analyzes the differing approaches and results of the submitted systems. Karl reads through the citation context summary for a quick overview of the approaches of these papers. Later, he can dig deeper by reading the entire citation context or by viewing the full text of the Nivre et al. paper.

Implementation Details

ASE is built using Java and the NetBeans Platform (Oracle, 2011) for window and settings management. The reference management view uses a version of the JabRef reference manager (JabRef Development Team, 2011) that was modified to interface with our brushing and linking framework. The citation network visualization and analysis components come from the SocialAction network analysis tool (Perer & Shneiderman, 2006), which was similarly altered to enable integration into our framework and automated loading of datasets. The remaining views in the interface for the citation context, automatically generated summaries, and full text are built using standard Java Swing widgets.

Data Import

The easiest way to get additional data into ASE is to load search results or other subsets of the 17,610 papers of the ACL Anthology Network (AAN) (Radev et al., 2009b, 2009a). The AAN includes a network of the citations between papers as well as the the paper metadata, abstract, plain text, and citation sentences. This data was all generated by the AAN team from the original PDF articles and metadata available in the ACL Anthology (Association for Computational Linguistics, 2011). The full text of each paper was obtained via OCR extraction of the PDFs and manual cleanup, from which the reference list was extracted. The authors and references required substantial cleaning, disambiguation, and correction which were done manually by the AAN team, assisted by an n-best matching algorithm with $n = 5$. The citation sentence extraction was done automatically by using string-based heuristics that match the citation pattern, author names, and publication year within the sentences to the reference list.

Initial loading from the AAN into ASE is done by processing the records to create the standard data files used by JabRef (BibTeX) and SocialAction (HCIL Network Visualization Input Data Format[8]). Each of the paper entries is modified to include unigram and bigram keywords generated from the plain text, a link to the AAN website for that paper, and a full text PDF automatically downloaded from the ACL Anthology. Additionally, the summarization process described below in more detail is used to create multi-document summaries for each possible topologic community. The summarization step is by far the most computationally expensive of the data loading tasks.

---

[7]Nivre, J., Hall, J., Kubler, S., McDonald, R., Nilsson, J., Riedel, S., and Yuret, D. (2007). *The CoNLL 2007 shared task on dependency parsing*. *2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*. Retrieved from http://clair.si.umich.edu/clair/anthology/query.cgi?type=Paper&id=D07-1096.

[8]http://www.cs.umd.edu/hcil/nvss/netFormat.shtml

	Median	Mean	StdDev	Min	Max
Size	15	25.3	24.6	1	126
Sentences (DP)	132	186.6	353.5	1	480
Sentences (AAN)	232	383.9	168.2	1	1021

Table 1:: This table shows statistics for the 126 unique communities found in the DP collection about their size, number of summarized sentences from DP, and number of sentences from the entire AAN that were not summarized. We show the median, mean, standard deviation (StdDev), min, and max for each.

To load arbitrary datasets into ASE, several pieces of metadata must be available or generated for each view. The reference management view requires some level of per-document information like what is available in standard academic databases, which can be expanded by including abstracts, DOIs or URLs, keywords, and PDFs. To show more than histograms or other simple visualizations of paper metadata, a citation network needs to be extracted from an academic database or generated for the collection as was done for the AAN. The former is easier, as many databases have created citation networks for at least a portion of their papers (usually the newest ones). All the citation network statistics and visualizations of ASE can be used with either source of network.

However, the hyperlinked full text, citation context, and citation context summary views require more information than is available from most academic databases. For these, individual citations within the paper full text need to be identified and the containing sentences extracted. Citation context for individual papers is available on Microsoft Academic Search (Microsoft Research, 2011), and used to be available on CiteSeer (Giles et al., 1998; Bollacker et al., 1998), but these services do not provide the citation locations within the full text. The citation context from these sites and generated summaries of it can be displayed in ASE, but without the citation locations the hyperlinked full text view can not be used. CiteSeer exposed only the citation sentences, but the underlying algorithms described in the CiteSeer papers could be used to record the citation locations as well.

Multi-Document Summarization

For multi-document summarization we use a modified version of Multi-Document Trimmer (MDT) (see the Design section on *Multi-Document Summarization*). Our current implementation of MDT processes all the citation context of each document in a selected group, which requires substantial computing time to build some of the summaries. For example, running MDT on 10 papers with 146 citation sentences took 555 seconds, while 16 papers with 338 citation sentences took 2580 seconds. The summaries were computed individually on a 30-node cluster, containing 10 2x4-core Intel Xenon processors with 32GB RAM each and 20 2x1-core Xenons with 8GB RAM each.

The MDT computation time is well beyond the interactive response times needed for ASE, so we decided to compute summaries for several pre-defined groups of papers that users would be interested in exploring. The network visualization view uses Newman's fast community-finding heuristic (Newman, 2004) to find topologically interesting groups of papers at several cutoff thresholds. We pre-computed community summaries for all

communities at each of the cutoff thresholds and wrote them to disk, displaying them when users select individual communities in the node-link diagram. Of the 884 communities found at all cutoff thresholds in the DP collection, only 126 unique communities need summarization. Statistics for those 126 unique communities are shown in Table. 1, including the size, number of citation sentences from within DP (that we summarized), and the number of citations from the entire AAN (that we did not summarize).

The summarization process can be sped up by pre-computing sentence candidates for the selected texts of each paper (citation context, abstract, or full text). The pre-computation uses the syntactic trimming and shortening initial steps of MDT. Then, for each community, the candidate sentences are retrieved for each paper, scored for relevance to the selected set, and chosen based on their features using the remaining steps of MDT. With this optimization the summarization time for the two communities mentioned before is roughly halved, from 555 seconds to 274 and from 2580 seconds to 1531. Additional algorithmic optimizations are possible, as many communities have substantial overlap or incremental additions for lower thresholds. Whether these optimized approaches would be suitable for real-time summarization is an interesting next step.

Newman's community-finding heuristic tends to find larger communities than some other approaches. One way to reduce the computation required is to use community-finding algorithms that find smaller, more tightly connected communities that have fewer citation sentences to summarize. These algorithms can be based on the citation network topology, paper text, or metadata.

Evaluation

In order to evaluate how effectively users could use ASE for exploring collections of papers, we conducted a planned, iterative user study procedure with refinements along the way to both the system and our testing methods. The evaluation consisted mainly of three qualitative usability studies over 17 months. An early formative study with five participants helped identify usability issues, guided the development of ASE, and determined the tasks users were interested in performing with the tool. This helped us plan two subsequent and more structured usability studies.

For all three evaluations we used the same Dependency Parsing dataset described in the Design section *Search & Data Import* and Scenario: Dependency Parsing. It is important to have user study participants analyze data of interest to them, and preferably their own data, to keep them motivated and to give the tool significance (Plaisant, 2004). Thus, we recruited researchers interested in and knowledgeable about Computational Linguistics as our participants for each study.

Here we will focus on a high-level overview of the studies and their results without delving into their details. Highly detailed descriptions of the studies and the results of each participant are described for the second study in Gove et al. (2011) and for both the second and third studies in Gove (2011).

Second Study

Our second study was designed to evaluate the usability and effectiveness of ASE after refining both the tool and testing methods during the formative evaluation.

Participants. There were four participants in the second study: two current Computer Science PhD students and two recent graduates. Of these, two had prior experience with Dependency Parsing.

Procedure. The ASE evaluations were conducted using a 30-inch LCD monitor with a resolution of 1920x1080, running off an Intel Core i3 2.26 Ghz laptop with 4 GB of RAM. Sessions were limited to 120 minutes, starting with a 30 minute training session. For the training phase we showed the participants video clips demonstrating each of the features of ASE. Between videos, we asked them to practice the tasks shown and ask questions if they did not understand the tool or its features.

We provided participants with two predefined tasks determined via our formative studies, taking around 60 minutes to complete. We asked participants to: (1) identify and make note of important authors and papers, and (2) find an important paper and collect evidence to determine why it is important. These open-ended tasks allowed participants to use whatever features of the tool they thought would be useful, while providing a basic benchmark for their performance.

For the remaining 30 minutes, we asked them to identify additional tasks of interest to them using the dataset. From these we selected one or more as individual goals for the remainder of the session and asked the participant to try to perform them using ASE.

Throughout the study we asked participants to use a think-aloud approach, making note of their thoughts and actions. We made note of which capabilities demonstrated in the training videos were used by each participant, for both the predefined and individual tasks. At the conclusion of the session participants were asked to comment on their experiences using the system.

Results. The second study demonstrated that users were able to use the basic features of the reference manager and network visualization views after the 30 minute video demonstration and practice session. Some users even began using the more advanced features of ASE almost immediately after the tutorial. The overall view available in the node-link diagram was used frequently by participants to orient themselves, as well as to find interesting clusters, trends, and motifs in the topology. This illustrated the value of using multiple coordinated views to provide an overview of the dataset. Disappointingly, most participants were using the same set of features at the beginning of the session as at the end, without branching out to the other features.

By far the most used feature was ranking and filtering by paper metadata or computed network statistics. As the predefined tasks focused on finding "important" papers and authors, perhaps the participants found the provided rankings by quantitative measures to be easy jumping-off points. Similarly, filtering by a metric provides a quick drill-down to the "important" papers (according to that metric).

Several participants made use of abstracts and full texts to explore paper content. One used the provided abstracts to determine which papers presented efficient algorithms. Two others used ASE similar to a digital library: exploring the whole collection, identifying papers of interest, and opening individual PDFs to analyze paper content. One user looked at paper titles and abstracts to help her decide which PDFs to open, which she then scanned to help her make final choices about which to read later. Another user more familiar with

Dependency Parsing opened the PDFs to examine how authors cite other papers, focusing on one in particular as an interesting advance of another.

The participants showed great interest in the citation context view, scanning it for interesting papers, authors, and insights. However, they had problems analyzing more recent or other poorly-cited papers due to the little or no context available. Moreover, the interaction between citation context and the other views was challenging for one user who wanted to open the PDF of each citing paper without changing the visualization focus to them.

The participants were interested in exploring the multi-document summary feature, and two participants used it to successfully understand paper content and guide their exploration. However, the participants were generally dissatisfied with the output quality of the summarization algorithm. MDT is designed to summarize news articles, and we found that citation sentences have several differences that need to be accounted for. For example, inline metadata and the disjoint nature of the sentences reduces the utility of MDT.

The interactions between the full text view and the other views were difficult for participants to understand, as each click on a citation changed the paper selected in all the other views while not changing the full text displayed. Perhaps a better indication of its relationships to the rest would be helpful, but this demonstrates once again that having systematic, homogeneous interactions and consistent highlighting across all views helps users understand the relationships.

From the results of this study, we identified and implemented several improvements for the interactions between the views in ASE. Moreover, we adjusted the MDT summarization algorithm so as to better handle citation context instead of news articles.

Third Study

Six months after our second user study, we conducted the third user study. Our goal was to study the impact of fixes and to evaluate usage patterns of more experienced users.

Participants. The participants of the third study were four current Computer Science PhD students, two of which had participated in the second study as well. All four indicated some knowledge of the concept of Dependency Parsing, if not the associated literature.

Procedure. Our procedure for the third study was identical to the second, with the sole additions of screen and audio capture during the evaluation session for later analysis.

Results. The new participants confirmed our previous observations about the ease of use and value of the coordinated reference manager and network visualization views. Overall the participants used the same general approaches, including extensive use of the ranking and filtering features. However, the two repeat participants that were in the second study used more features their second time around and were able to find deeper insights in the dataset. This demonstrated the value of using extended duration evaluation techniques such as Multi-dimensional In-depth Long-term Case studies (MILCs) (Shneiderman & Plaisant, 2006), which focus on actual use of the system by domain experts solving their own problems. MILCs are well suited to evaluating creativity and exploration tools such as ASE that may be too complicated to understand in a single analysis session, though we were unable to recruit expert users and import their own datasets for a MILC study.

The improvements we applied to the MDT summarization algorithm and the interactions between views helped users with their analyses. The new citation context summaries were used frequently during this study, and the participants were more satisfied with the linguistic structure of the summaries. They found that there were often coherent summaries of the themes in smaller communities, but were unable to find clear themes for larger ones. This is to be expected given the small size of this dataset and the large, diverse central community. Additionally, participants wanted more types of community summaries like topic modeling or using abstracts and full texts instead of only citation contexts.

The automatic community finding algorithm was used by participants for several tasks, however it was limited by the small size of the dataset and by the types of communities it produced. Participants wanted additional clustering techniques for particular tasks and process models, and that were not limited to only clustering based on topology. Moreover, they wanted to select arbitrary sets of papers to summarize instead of being limited to the sets found by the clustering algorithm. This capability is limited by the speed the multi-document summarization algorithm. Unfortunately MDT is not fast enough for this currently, though the Implementation Details section on *Multi-Document Summarization* discusses one potential improvement.

Discussion

These three preliminary user studies provide a basis for interpreting the effectiveness of ASE as a literature exploration tool. These in-depth exploratory studies are becoming more common and are appropriate for understanding the complex intellectual tasks required for insight and discovery in visual analytics tools. While more rigorous and extensive evaluations would be beneficial, these preliminary evaluations helped guide refinements to ASE and provide evidence for its usefulness for specific tasks.

From our three user studies we found that users can understand how to use ASE after 30 minutes of instruction, though they did not use many of the features in their first session. In addition, our repeat participants demonstrated that with more sessions with the tool they can use more features and find deeper insights than they could initially. From the evaluations we discovered several usability issues with ASE, most of which we were able to correct and test again in the last user study. The improvements we made seemed to be effective, especially the coherence of the summaries generated by our modified version of MDT.

The user-defined tasks in the studies helped us to identify several common questions users ask when exploring paper collections. Foremost they wanted to identify the foundations, breakthroughs, state-of-the-art, and evolution of a field. Next, they were looking to find collaborators and relationships between disparate communities. They were also searching for easily understandable overview papers like surveys to help guide their exploration.

We also developed a set of user requirements for exploring scientific literature networks to help guide the design process. First, users want control over the collection they are exploring. They want to choose a custom subset via a query and iteratively refine and drill down into it, putting them in control of the analysis. Next, users appreciate an overview of the subset either as a visualization or text statistics. Overviews help users orient themselves in the subset and allow them to quickly browse via details-on-demand or other multiple coordinated view approaches. Our users made extensive use of the ranking and filtering

features, demonstrating that easy to understand metrics for identifying interesting papers can provide a jump off point for more detailed analyses. Moreover, users should be able to create groups of papers and annotate them with their findings. Grouping and annotating helps users organize their discovery process, and lets them save their analyses so as to come back to them over a period of days or weeks.

Likewise, we identified several recommendations for future researchers conducting similar evaluations. We strongly recommend extended user studies for evaluating complex creativity and exploration tools like ASE. Our 90-minute sessions were helpful and returning participants provided even more useful feedback about ASE's design. One way to improve the tutorial retention is to follow the suggestions of Plaisant and Shneiderman (2005), which suggests having short clips about the features available throughout the sessions for participants to refresh their memory. Similarly, embedded training, animations, or slowly revealing features may help guide users in using the full capabilities of the system. Finally, the importance of motivating participants can not be stressed enough. Identify your target participants early and allow easy import from one or more general data sources of interest to them so they can analyze their own data.

These recommendations would have helped us, as our evaluation is limited by many of these issues. It is difficult to import new datasets into ASE due to the processing requirements discussed in the Implementation Details section. The collection we used contains only 147 papers, though in our evaluations participants were still able to find interesting insights. We had to select participants interested in the research area rather than letting them use their own datasets, which limited the pool of available researchers and their motivation. In the end, we only had six PhD student participants and our efforts to recruit users from other target user groups for a longer MILC study were not successful. However, we still found many useful insights and usability fixes.

Conclusion & Future Work

Understanding scientific domains and topics is a challenging task that is not well supported by current search systems. Fact, document, or exploratory search might require only minutes or hours to attain success, but understanding emerging research fields can take days or weeks. By integrating statistics, text analytics, and visualization we have some hope of providing users with the tools they need to generate readily-consumable surveys of scientific domains and topics. Our prototype implementation Action Science Explorer (ASE) combines reference management, statistics, citation context, automatic summarization, ranking & filtering, and network visualization in several coordinated views. We hope the design and ideas behind ASE provide inspiration for designers of similar commercial and research tools that could benefit from our approach. We do not plan to distribute or support ASE, but our source code is available on request.[1]

The three-phase usability study guided our revisions to ASE and led us to improve the testing methods. These evaluations demonstrated the utility of showing several coordinated views of a paper collection. Moreover, they identified several exploration tasks users are interested in and the benefit of specific functionalities when performing them. The evaluations also found many limitations of ASE including the large screen space required and inconsistent user interfaces between views.

The applicability of ASE to literature exploration, and its future evaluation, depends on the ease to which new datasets can be imported. Many of the views ASE provides can be populated from a wide variety of academic databases, though the citation context and summary views require more extensive datasets and processing. Some academic databases are starting to provide citation context that would be usable by ASE, and we are currently exploring refinements to our multi-document summarization approach to reduce processing time and to better handle the disjoint nature of citation text. Our work was, in large part, an effort to determine the degree to which we could rely specifically on citation text for the integration of our natural language processing and visualization approaches, but we expect future work to explore more content-based efforts toward exploring the full space of papers and authors relevant to a particular scientific topic.

With support for larger, more diverse datasets several interesting studies become feasible. Letting users analyze their own datasets with ASE enables extended user studies that are better suited for evaluating creativity and exploration tools. These users could be from different backgrounds with varying experience and expertise, so as to ascertain the suitability of ASE for these roles. Moreover, the integrated visualization and text analytics approach of ASE could be compared against traditional techniques for numeric analysis of communities and citation patterns. Furthermore, ASE could be used to analyze citation network datasets and to report interesting discoveries about citation structures and patterns.

References

Aris, A., Shneiderman, B., Qazvinian, V., & Radev, D. (2009). *Visual overviews for discovering key papers and influences across research fronts*. *JASIST: Journal of the American Society for Information Science and Technology, 60*(11), 2219–2228. doi:10.1002/asi.21160

Association for Computational Linguistics. (2011). *ACL Anthology*. Retrieved from http://aclweb.org/anthology-new

Association for Computing Machinery. (2011). *ACM Digital Library*. Retrieved from http://portal.acm.org

Bates, M. J. (1990). *Where should the person stop and the information search interface start?* *Information Processing & Management, 26*(5), 575–591. doi:10.1016/0306-4573(90)90103-9

Bertini, E., Perer, A., Plaisant, C., & Santucci, G. (2008). *BELIV'08: BEyond time and errors: novel evaLuation methods for Information Visualization*. In *CHI EA '08: proc. CHI '08 extended abstracts on human factors in computing systems* (pp. 3913–3916). doi:10.1145/1358628.1358955

Bollacker, K. D., Lawrence, S., & Giles, C. L. (1998). *CiteSeer: an autonomous Web agent for automatic retrieval and identification of interesting publications*. In *AGENTS '98: proc. second international conference on autonomous agents* (pp. 116–123). doi:10.1145/280765.280786

Bradshaw, S. (2003). *Reference directed indexing: redeeming relevance for subject search in citation indexes*. In T. Koch & I. Slvberg (Eds.), *ECDL '03: proc. 7th european conference on research and advanced technology for digital libraries* (Vol. 2769, pp. 499–510). doi:10.1007/978-3-540-45175-4_45

Center for History and New Media, G. (2011). *Zotero [Software]*. Retrieved from `http://www.zotero.org`

Chen, C. (2004). *Searching for intellectual turning points: Progressive knowledge domain visualization*. PNAS: Proc. National Academy of Sciences of the United States of America, *101*(90001), 5303–5310. doi:`10.1073/pnas.0307513100`

Chen, C. (2006). *CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature*. JASIST: Journal of the American Society for Information Science and Technology, *57*(3), 359–377. doi:`10.1002/asi.20317`

Chen, C., & Czerwinski, M. P. (2000, November). *Empirical evaluation of information visualizations: an introduction*. International Journal of Human-Computer Studies, *53*(5), 631–635. doi:`10.1006/ijhc.2000.0421`

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010, July). *The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis*. JASIST: Journal of the American Society for Information Science and Technology, *61*, 1386–1409. doi:`10.1002/asi.v61:7`

Cornell University Library. (2011). *ArXiv*. Retrieved from `http://arxiv.org`

Dorr, B., Zajic, D., & Schwartz, R. (2003). *Hedge Trimmer: a parse-and-trim approach to headline generation*. In *HLT/NAACL-DUC '03: proc. HLT/NAACL 2003 text summarization workshop and document understanding conference* (pp. 1–8). doi:`10.3115/1119467.1119468`

Dunne, C., Riche, N. H., Lee, B., Metoyer, R. A., & Robertson, G. G. (2012). *GraphTrail: analyzing large multivariate and heterogeneous networks while supporting exploration history*. In *CHI '12: proc. 2012 international conference on human factors in computing systems*.

Elsevier. (2011). *SciVerse Scopus*. Retrieved from `http://scopus.com/`

Garfield, E. (1994, January). *The concept of citation indexing: a unique and innovative tool for navigating the research literature*. Current Contents. Retrieved from `http://thomsonreuters.com/products_services/science/free/essays/concept_of_citation_indexing/`

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). *CiteSeer: an automatic citation indexing system*. In *DL '98: proc. 3rd ACM conference on digital libraries* (pp. 89–98). doi:`10.1145/276675.276685`

Google. (2011). *Google Scholar*. Retrieved from `http://scholar.google.com`

Gove, R. (2011). *Understanding scientific literature networks: case study evaluations of integrating vizualizations and statistics*. (Master's thesis, University of Maryland, Department of Computer Science). Retrieved from `http://hdl.handle.net/1903/11764`

Gove, R., Dunne, C., Shneiderman, B., Klavans, J., & Dorr, B. (2011). *Evaluating visual and statistical exploration of scientific literature networks*. In *VL/HCC '11: proc. 2011 IEEE symposium on visual languages and human-centric computing* (pp. 217–224). doi:`10.1109/VLHCC.2011.6070403`

Hearst, M. A. (2009). *Search user interfaces* (1st). Cambridge University Press. Retrieved from `http://searchuserinterfaces.com/book`

Institute of Electrical and Electronics Engineers. (2011). *IEEE Xplore*. Retrieved from `http://ieeexplore.ieee.org`

JabRef Development Team. (2011). *JabRef [Software]*. Retrieved from http://jabref.sourceforge.net

Kuhlthau, C. C. (1991). *Inside the search process: Information seeking from the user's perspective*. JASIS: Journal of the American Society for Information Science, *42*(5), 361–371. doi:10.1002/(SICI)1097-4571(199106)42:5%3C361::AID-ASI6%3E3.0.CO;2-%23

Marchionini, G. (1997). *Information seeking in electronic environments*. Cambridge University Press. Retrieved from http://www.ils.unc.edu/~march/isee_book/web_page.html

Marchionini, G. (2006, April). *Exploratory search: from finding to understanding*. CACM: Communications of the ACM, *49*, 41–46. doi:10.1145/1121949.1121979

Mei, Q., & Zhai, C. (2008, June). *Generating impact-based summaries for scientific literature*. In *ACL/HLT '08: proc. 46th annual meeting of the association for computational linguistics: human language technologies* (pp. 816–824). Retrieved from http://aclweb.org/anthology-new/P/P08/P08-1093

Mendeley Ltd. (2011). *Mendeley [Software]*. Retrieved from http://www.mendeley.com

Microsoft Research. (2011). *Microsoft Academic Search*. Retrieved from http://academic.research.microsoft.com

Mohammad, S., Dorr, B., Egan, M., Hassan, A., Muthukrishan, P., Qazvinian, V., . . . Zajic, D. (2009). *Using citations to generate surveys of scientific paradigms*. In *HLT/NAACL '09: proc. human language technologies: the 2009 annual conference of the north american chapter of the association for computational linguistics* (pp. 584–592). doi:10.3115/1620754.1620839

Nanba, H., & Okumura, M. (1999). *Towards multi-paper summarization using reference information*. In *IJCAI '99: proc. 16th international joint conference on artificial intelligence* (pp. 926–931). Retrieved from http://portal.acm.org/citation.cfm?id=1624312.1624351

Nanba, H., Abekawa, T., Okumura, M., & Saito, S. (2004). *Bilingual PRESRI: integration of multiple research paper databases*. In C. Fluhr, G. Grefenstette & W. B. Croft (Eds.), *RIAO '04: proc. 7th international conference on computer-assisted information retrieval (recherche d'information assistee par ordinateur)* (pp. 195–211). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.102.9560

National Center for Biotechnology Information. (2011). *PubMed*. Retrieved from http://ncbi.nlm.nih.gov/pubmed

National Visualization and Analytics Center. (2005). *Illuminating the path: The research and development agenda for visual analytics* (J. J. Thomas & K. A. Cook, Eds.). IEEE Computer Society. Retrieved from http://nvac.pnl.gov/agenda.stm

Newman, M. E. J. (2004, June). *Fast algorithm for detecting community structure in networks*. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, *69*(6), 066133. doi:10.1103/PhysRevE.69.066133

North, C., & Shneiderman, B. (1997). *A taxonomy of multiple window coordinations* (Human-Computer Interaction Lab Tech Report No. HCIL-97-18). Retrieved from http://www.cs.umd.edu/local-cgi-bin/hcil/rr.pl?number=97-18

NWB Team. (2006). *Network Workbench [Software]*. Retrieved from http://nwb.slis.indiana.edu

Oracle. (2011). *NetBeans Platform [Software]*. Retrieved from `http://netbeans.org/features/platform`

Perer, A., & Shneiderman, B. (2006, October). *Balancing systematic and flexible exploration of social networks*. TVCG: IEEE Transactions on Visualization and Computer Graphics, *12*(5), 693–700. doi:`10.1109/TVCG.2006.122`

Perer, A., & Shneiderman, B. (2008). *Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis*. In *CHI '08: proc. 26th annual SIGCHI conference on human factors in computing systems* (pp. 265–274). doi:`10.1145/1357054.1357101`

Plaisant, C. (2004). *The challenge of information visualization evaluation*. In *AVI '04: proc. 2004 working conference on advanced visual interfaces* (pp. 109–116). doi:`10.1145/989863.989880`

Plaisant, C., & Shneiderman, B. (2005). *Show me! Guidelines for producing recorded demonstrations*. VLHCC '05: Proc. 2005 IEEE Symposium on Visual Languages and Human-Centric Computing, *00*, 171–178. doi:`10.1109/VLHCC.2005.57`

Qazvinian, V., & Radev, D. R. (2008). *Scientific paper summarization using citation summary networks*. In *COLING '08: proc. 22nd international conference on computational linguistics* (pp. 689–696). doi:`10.3115/1599081.1599168`

Radev, D., Otterbacher, J., Winkel, A., & Blair-Goldensohn, S. (2005, October). *NewsInEssence: summarizing online news topics*. Communications of the ACM, *48*, 95–98. doi:`10.1145/1089107.1089111`

Radev, D. R., Joseph, M. T., Gibson, B., & Muthukrishnan, P. (2009a). *A bibliometric and network analysis of the field of computational linguistics*. JASIST: Journal of the American Society for Information Science and Technology. To appear. Retrieved from `http://clair.si.umich.edu/~radev/papers/biblio.pdf`

Radev, D. R., Muthukrishnan, P., & Qazvinian, V. (2009b). *The ACL Anthology Network corpus*. In *NLPIR4DL '09: proc. ACL-IJCNLP 2009 workshop on text and citation analysis for scholarly digital libraries* (pp. 54–61). doi:`10.3115/1699750.1699759`

Saraiya, P., North, C., & Duca, K. A. (2005). *An insight-based methodology for evaluating bioinformatics visualizations*. TVCG: IEEE Transactions on Visualization and Computer Graphics, *11*(4), 443–456. doi:`10.1109/TVCG.2005.53`

Saraiya, P., North, C., Lam, V., & Duca, K. A. (2006, November). *An insight-based longitudinal study of visual analytics*. TVCG: IEEE Transactions on Visualization and Computer Graphics, *12*(6), 1511–1522. doi:`10.1109/TVCG.2006.85`

Sekine, S., & Nobata, C. (2003). *A survey for multi-document summarization*. In D. Radev & S. Teufel (Eds.), *HLT/NAACL-DUC '03: proc. HLT/NAACL 2003 text summarization workshop and document understanding conference* (pp. 65–72). doi:`10.3115/1119467.1119476`

Seo, J., & Shneiderman, B. (2006). *Knowledge discovery in high-dimensional data: case studies and a user survey for the rank-by-feature framework*. TVCG: IEEE Transactions on Visualization and Computer Graphics, *12*(3), 311–322. doi:`10.1109/TVCG.2006.50`

Shneiderman, B., & Plaisant, C. (2006). *Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies*. In *BELIV '06: proc. 2006 avi workshop on BEyond time and errors: novel evaLuation methods for Information Visualization* (pp. 1–7). doi:`10.1145/1168149.1168158`

Teufel, S., & Moens, M. (2002). *Summarizing scientific articles: experiments with relevance and rhetorical status*. *CL: Computational Linguistics*, *28*(4), 409–445. doi:10.1162/089120102762671936

Teufel, S., Siddharthan, A., & Tidhar, D. (2006). *Automatic classification of citation function*. In *EMNLP '06: proc. 2006 conference on empirical methods in natural language processing* (pp. 103–110). doi:10.3115/1610075.1610091

Thomson Reuters. (2011a). *EndNote [Software]*. Retrieved from http://www.endnote.com

Thomson Reuters. (2011b). *ISI Web of Knowledge*. Retrieved from http://isiwebofknowledge.com

Whidby, M., Zajic, D., & Dorr, B. J. (2011). *Citation handling for improved summarization of scientific documents* (Language and Media Processing Laboratory Tech Report No. LAMP-TR-157). Retrieved from http://hdl.handle.net/1903/11822

Zajic, D., Dorr, B., & Schwartz, R. (2004). *BBN/UMD at DUC-2004: Topiary*. In *HLT/NAACL-DUC '04: proc. HLT/NAACL 2004 workshop on document understanding* (pp. 112–119). Retrieved from http://www.umiacs.umd.edu/~bonnie/publications.html

Zajic, D., Dorr, B. J., Lin, J., & Schwartz, R. (2007, November). *Multi-candidate reduction: Sentence compression as a tool for document summarization tasks*. *Information Processing and Management*, *43*(6), 1549–1570. doi:10.1016/j.ipm.2007.01.016

Zajic, D. M., Dorr, B. J., Schwartz, R., Monz, C., & Lin, J. (2005, October). *A sentence-trimming approach to multi-document summarization*. In *HLT/EMNLP '05: proc. HLT/EMNLP 2005 workshop on text summarization* (pp. 151–158). Retrieved from http://www.casl.umd.edu/node/719