

ABSTRACT

Title of thesis: UNDERSTANDING SCIENTIFIC LITERATURE NETWORKS: CASE STUDY EVALUATIONS OF INTEGRATING VISUALIZATIONS AND STATISTICS

Robert Gove, Master of Science, 2011

Thesis directed by: Professor Ben Shneiderman
Department of Computer Science

Investigators frequently need to quickly learn new research domains in order to advance their research. This thesis presents five contributions to understanding how software helps researchers explore scientific literature networks. (1) A taxonomy which summarizes capabilities in existing bibliography tools, revealing patterns of capabilities by system type. (2) Six participants in two user studies evaluate Action Science Explorer (ASE), which is designed to create surveys of scientific literature and integrates visualizations and statistics. Users found document-level statistics and attribute rankings to be convenient when beginning literature exploration. (3) User studies also identify users' questions when exploring academic literature, which include examining the evolution of a field, identifying author relationships, and searching for review papers. (4) The evaluations suggest shortcomings of ASE, and this thesis outlines improvements to ASE and lists user requirements for bibliographic exploration. (5) I recommend strategies for evaluating bibliographic exploration tools based on experiences evaluating ASE.

UNDERSTANDING SCIENTIFIC LITERATURE NETWORKS:
CASE STUDY EVALUATIONS OF INTEGRATING
VISUALIZATIONS AND STATISTICS

by

Robert Paul Gove

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2011

Advisory Committee:
Professor Ben Shneiderman, Chair/Advisor
Professor Bonnie Dorr
Professor Douglas W. Oard

© Copyright by
Robert Paul Gove
2011

Acknowledgments

Writing a list of acknowledgements requires both a perfect memory of all people who have been in your life plus an exact algorithm for retrieving those people from memory. I lack both. Below is a partial list based on a faulty memory and an approximation algorithm.

First and foremost, I would like to acknowledge the support, guidance, and thoughtful comments from my advisor, Ben Shneiderman. Studying in the HCIL with him has been a fantastically rewarding opportunity. I also acknowledge the wonderful support and advice from Judith Klavans who helped keep me on track and acted as an unofficial member of my committee. I also thank my committee members Bonnie Dorr and Doug Oard for sharing their very helpful insights which greatly sharpened my thoughts about my work and helped shape the final version of my thesis.

Being welcomed into the HCIL is one of the highlights of my academic career. Perhaps it is overshadowed only by graduating from the HCIL. The friendly atmosphere, amazing research, and kind support have made this a fantastic place to be. Thanks to everyone who made it feel like home. I especially thank Awalin Shopan, Cody Dunne, John Alexis Guerra Gómez, Krist Wongsuphasawat, and Tak Yeon Lee who provided excellent comments, help, and irrelevant conversations while drafting my thesis.

I would also like to thank the others who provided various types of support for moving my research forward: Michael Whidby, David Zajic, Rebecca LaPlante,

Derek Hansen, and Catherine Plaisant.

I thank my many friends and supporters who made graduate school easier. Moving to Maryland wasn't easy, and grad school can be a lonely place, but my friends—old and new—helped keep my sanity in place.

Finally, I would like to thank coffee.

Table of Contents

List of Tables	vi
List of Figures	vii
List of Abbreviations	x
1 Introduction	1
1.1 Problem Definition	1
1.1.1 Exploration of Academic Literature	1
1.1.2 Evaluation of Exploration Systems	4
2 Related Work	7
2.1 Information Seeking Theory	7
2.2 Exploring Academic Literature	9
2.3 Evaluation	28
2.4 Description of Action Science Explorer	33
3 Evaluation and Improvements	40
3.1 First Usability Study	40
3.1.1 Participants	40
3.1.2 Experimental Design	41
3.1.3 Procedure	42
3.1.4 Results	44
3.1.4.1 Predefined tasks	44
3.1.4.2 User-defined tasks	50
3.1.4.3 Additional tasks	52
3.1.5 Discussion	54
3.2 Other Evaluations	58
3.3 Improvements Made to Action Science Explorer	60
3.4 Second Usability Study	63
3.4.1 Participants	64
3.4.2 Experimental Design	64
3.4.3 Procedure	64
3.4.4 Results	65
3.4.4.1 Predefined tasks	65
3.4.4.2 User-defined tasks	70
3.4.5 Discussion	79
3.5 Summary of Evaluations	82
4 Conclusions and Future Work	85
4.1 Information Seeking Behavior in Action Science Explorer	85
4.2 Recommendations	86

4.2.1	User Requirements for Understanding Scientific Literature Networks	86
4.2.2	Evaluation Strategies	87
4.3	Limitations	89
4.4	Conclusion	90
	Bibliography	92

List of Tables

2.1	Table of capabilities by system, grouped by system type (search engines, reference managers, and summarizers and recommenders). . . .	25
3.1	Table of capabilities in my first usability study used by each participant for the first predefined task (identifying important authors and papers).	47
3.2	Table of capabilities in my first usability study evaluation used by each participant for the second predefined task (collecting evidence). .	49
3.3	Table of capabilities in my first usability study used by each participant for the user-defined tasks.	53
3.4	Table of capabilities in my second usability study used by each participant for the first predefined task (identifying important authors and papers).	69
3.5	Table of capabilities in my second usability study used by each participant for the second predefined task (collecting evidence).	71
3.6	Table of capabilities in my second usability study used by each participant for the user-defined tasks.	79

List of Figures

2.1	Query results page on CiteSeerX for “Support-Vector Networks”. (1) Search results can be ranked by search relevance, number of citations, publication year, or recency (which attempts to provide results that are both relevant and recent). (2) Excerpts from the documents which match the search query are shown in bold italics, and the citation count is shown at the bottom of each record.	11
2.2	CiteSeerX’s detailed view of a paper, “Support-Vector Networks” by Corinna Cortes and Vladimir Vapnik. (1) Clicking the link for the number of citations shows all the documents which cite “Support-Vector Networks” and allows users to see the citation context of each citation. (2) The list of papers which “Support-Vector Networks” cites, with citation counts for each citation.	12
2.3	Query results page on Google Scholar for “Support-Vector Networks”. Search results can be filtered by recent years. Excerpts from the documents which match the search query are shown in bold text, and the citation count is shown at the bottom of each record.	13
2.4	A view in GoPubMed of the author Antonella Tosti which shows (1) a summary of top terms, and (2) a bar chart of publication counts over time.	15
2.5	A view in GoPubMed of the author collaboration network for Antonella Tosti.	16
2.6	A citation map in Web of Knowledge of “Literate Programming” by Donald Knuth. This network can be filtered by publication year (1)—in this case, the network is filtered to show only papers published between 1970 and 1995. The citation map (2) shows the backward citations (i.e., papers cited by “Literate Programming”) and forward citations (i.e., papers that cite “Literate Programming”).	17
2.7	Mendeley can aggregate information from all users’ libraries and show users (1) the top read articles in one or all disciplines, and (2) the most read articles in one or all disciplines.	19
2.8	For users’ own personal libraries, Mendeley can show (1) the growth in users’ libraries over time, (2) the top authors, (3) the top publication outlets, and (4) the most frequently used tags.	20
2.9	XplorMed users can filter their document set by categories and year (a), see word association scores (b), and filter documents by ranked word chains (c).	21
2.10	Boyack and Börner’s visualization in VxInsight of Behavioral and Social Research (BSR) Program grants and publications clustered by Medical Subject Heading (MeSH) terms. Labels show common keywords, and height indicates number of documents. Image is from [8].	22

2.11	CiteSpace II can highlight sudden increases in citations with bursts of color (a) and show co-citation clusters over the timeline of the network (b).	24
2.12	The Action Science Explorer system: (1) the bibliography list, (2) bibliographic entry details, (3) bibliographic search, (4) bibliographic groups, (5) the attribute ranking view, (6) the citation network visualization with Find Communities enabled, (7) the In-Cite Text, (8) the multi-document In-Cite Summary of a cluster, and (9) the full-text article with linked citations.	35
2.13	The different views in Action Science Explorer can be hidden or detached into separate windows. Here, the In-Cite Summary view is hidden, and the Reference Manager (1) and In-Cite Text (2) views have been detached from the main window (3) into their own windows and rearranged to fit a vertical display.	36
3.1	The In-Cite Text view (bottom right) shows several citations which make positive remarks about the paper H05-1066 by McDonald et al.	45
3.2	P2 chose H05-1066 as an important paper after ranking papers by corpus in-degree filtering to show the top most highly cited papers. This revealed that H05-1066 (shown highlighted in blue) is highly cited and cites two of the most highly cited papers. The paper's details are visible on the left, including authors, title, and abstract.	48
3.3	The Attribute Ranking view in ASE was fixed to allow multiple selections that would correctly coordinate selection across all views. Here, the top ten most highly cited papers in the dependency parsing set were selected in the Attribute Ranking (1), which then move to the top of the reference manager and are shown in white and blue (2), allowing us to see that McDonald and Nivre authored eight of the top ten papers.	62
3.4	The pre-computed groups of authors (1) on the left can be selected, which then moves the papers in that group (2) to the top of the list of bibliography entries.	66
3.5	The Plot Nodes view shows a scatter plot of the papers in the citation network with one variable on the X axis (e.g., publication year) and a second variable on the Y axis (e.g., corpus in-degree). Here, the vertical double-ended slider (1) has been adjusted to filter out papers with a low corpus in-degree.	67
3.6	When the Find Communities function is enabled, users can select a cluster of papers (1) which become highlighted in yellow, the In-Cite Summary (2) is shown, and the documents in the community are moved to the top of the reference manager (3).	74
3.7	While exploring the documents in the large cluster (1) and viewing them in the attribute ranking (2) and reference manager (3), P1 was able to discover that a large number of papers in this dataset were published in two conferences from 2006 and 2007 (3).	76

3.8 A scatter plot of corpus in-degree and HITS Authority reveals an outlier (1) (also highlighted in the citation network (2)) which is shown to be a paper by Collins (3). 77

List of Abbreviations

ACL	Association of Computational Linguistics
ASE	Action Science Explorer
DL	Digital Library
IR	Information Retrieval
IRB	Institutional Review Board
IS	Information Seeking/Seeker

Chapter 1

Introduction

1.1 Problem Definition

A frequent challenge for researchers and scholars is to quickly learn about a new research area. There are many reasons someone may need to do this. Researchers who wish to apply a state-of-the-art technique to their research will need to search for leading papers and recent breakthroughs. A review panel which is reviewing grant proposals from unfamiliar fields may need to identify solved problems, open questions, current trends, and emerging fields. Graduate students who need to become familiar with the research in their chosen areas may search for historical papers, leading authors and research groups, appropriate publication venues, current research methodology, and new directions for research. These tasks can be challenging due to the disparate sources of literature and the relative lack of powerful tools to support exploring and summarizing academic literature.

1.1.1 Exploration of Academic Literature

Users with different profiles may wish to explore academic literature to answer different but overlapping questions. Therefore, systems for exploring academic literature should ideally support a variety of tasks and levels of users, thus introducing a significant challenge for system designers: A wide range of use cases need to

be explored. For example, students may seek to identify a central paper for a class presentation, interdisciplinary researchers will need to learn about research methods from other domains, and new researchers will want to dig deep into a field of study.

As time goes by, digital libraries continue to expand due to new literature being written and old literature continuously being digitized, thus perpetually enlarging the pool of literature to explore. Although improving the accessibility of literature should benefit researchers, it puts pressure on literature exploration systems to support larger data sets, thus making in-depth exploration of digital libraries a more challenging process.

Existing systems, such as Google Scholar [22] and CiteSeerX [42], provide summaries on a per-paper basis. For a paper A , these summaries may include the paper's title, the abstract, list of authors, year of publication, number of citations, a list of papers citing A , and excerpts from other papers showing the text that cites A . These data and statistics can be useful for understanding the paper's domain and impact, but may be less helpful for exploring the domain or for understanding the domain as a whole.

Furthermore, there is surprisingly little overlap in supported functionality across existing systems, possibly indicating that there is no consensus regarding which features are most salient for exploring academic literature. Search engine-based systems naturally center around searching, but users may not have a clear goal in mind, or the goal may be expressed in a query that a literature search engine cannot easily answer, such as identifying the most prominent authors in a given field of research. My first contribution is a survey in Section 2.2 of existing systems

which can be used to explore scientific literature.

The discovery process potentially could be improved by providing summaries of an entire academic field instead of summaries on a per-paper basis. To this end, researchers have worked on summarizing literature textually and providing graphical overviews, e.g., through citation summarizations [36] and graphical models of citations [3]. This could be augmented through additional visualizations, statistics, and metadata. For example, combining a citation visualization with citation summaries may be more powerful than either approach when used alone.

Action Science Explorer (ASE) was developed to address these issues. The Visual Information Seeking Mantra [57] suggests providing an overview, allowing users to zoom and filter, and giving details on demand. This can be a beneficial model for exploring academic literature. In this situation users may wish to view an overview of the papers in the domain, filter out uninteresting papers, and access specific details on a given paper. ASE supports these actions by integrating SocialAction, a network analysis tool with statistics and a dynamic network visualization, and JabRef, a reference manager. SocialAction provides an overview by displaying a collection of papers as a citation network, and allows users to rank and filter papers by statistical properties. Selecting a paper will show the details about the paper in JabRef plus display all sentences from other papers that cite the selected paper. These are some of the many features supported by ASE, which is described in more detail in Section 2.4.

These features may be highly beneficial for the exploration and discovery process. With this in mind, we wish to determine whether and how ASE supports

academic literature exploration. Specifically:

- What questions do users need/want to ask when exploring academic literature?
- Does ASE have sufficient features for users to make deep conclusions about a collection of papers?
- How does ASE help users understand the literature in a new domain?
- What possible modifications would improve the system?
- What is the trajectory of use of new users as they learn ASE?
- For mature users, what are the process models of how they employ ASE?

My second contribution addresses these questions with an extended usability study of ASE using qualitative research methods. The extended usability study revealed several effective capabilities in ASE and some potential improvements to ASE, which led me to develop an improved version of ASE. This user study also led to my third contribution, which is a list of questions that users ask and tasks that users wish to perform when they explore scientific literature. Additionally, my fourth contribution is a list of user requirements for literature exploration systems.

1.1.2 Evaluation of Exploration Systems

A problem related to exploration of academic literature is how to evaluate systems that support exploration and knowledge discovery. Some discoveries while exploring the literature might be considered simple and may not require much time

or effort if those actions are appropriately supported by the system, such as determining the number of citations a paper has received. Yet other discoveries may require longer periods of system use, such as identifying emerging trends in research. There is no guarantee that discovery will happen quickly while exploring data [49], and since I wish ultimately to evaluate whether the system supports discovery the evaluation process should account for usage goals that may differ from other types of systems such as productivity tools.

Emerging trends in evaluation methods aim to support effective evaluation of exploration systems. Although controlled experiments measuring time and errors have been a common metric for user interfaces, they may be less relevant for interfaces supporting discovery [49]. To resolve this, researchers have begun emphasizing open-ended protocols and qualitative analysis [39]. These evaluation strategies might be realized through case studies which can be useful for analyzing long-term patterns of use of the system [49], although when used alone with small numbers of participants the results of case studies may not generalize well.

The goal of validity may be more effectively achieved through the use of Multi-dimensional In-depth Long-term Case studies (MILCs), which are especially well adapted to evaluating systems that support creativity, exploration, and discovery [59]. MILCs emphasize long-term use of the system by domain experts solving their own problems where usage data is collected from a variety of sources, such as participant observation, surveys, interviews, and usage logs.

In this thesis I explored the use of extended usability studies for evaluating evolving prototypes of Action Science Explorer. These evaluations yielded insights

into the evaluation methodology. My fifth contribution in this thesis is a set of recommendations for evaluators of literature exploration systems.

Chapter 2

Related Work

In this chapter I review related work in information seeking, literature exploration systems, and evaluation methodology.

2.1 Information Seeking Theory

Hearst [24] provides a review of Information Seeking (IS) theories. These theories frequently assume that seekers operate in a process which involves recognizing an information need, specifying the query, evaluating the results, and possibly reformulating the query. However, when users are unsure what they are looking for, or how to describe it, they may be more inclined to participate in exploration or browsing behavior instead of searching.

Taylor was an early writer on the process of asking questions and information need. He describes four levels of information need [61] [62]:

1. The actual, but unexpressed, need for information.
2. The conscious within-brain description of the need.
3. The formal statement of the question.
4. The question as presented to the information system.

Taylor points out that information seekers frequently cannot define what they want, but they are able to describe why they need it [62].

Kuhlthau conducted five studies to investigate common user experiences during information seeking [33]. Her studies showed that the information seekers go through a six-stage Information Search Process: initiation, selection, exploration, formulation, collection, and presentation. Pirolli and Card [46] discuss information foraging. Analysts tend to select documents which have a potentially high profit and narrow the set down before finally reading the refined set of documents. In this way analysts follow a “scent” or cue about how to direct the search process. Bates [4] presents several information search stratagems and gives several examples, such as the Area Scan where users identify a general area of interest and then browse in that area.

One problem that information seekers frequently face is that they do not know how to express or define what they don't know, or their information need [30] [29] [5] [6] [62] [5]. Hutchins discusses an approach to this problem where the “aboutness” of documents is described in terms of the author's presuppositions about readers [30] [29]. In this approach, information systems contain descriptions of documents in terms of what the document's author presupposes the reader knows. The motivation comes from the fact that even though information seekers may not know how to express their information need, seekers can still define what they do know. Thus, the information system can lead users from what they know to what they don't know, which they don't know how to express or define.

Another model which describes the information seeking is Belkin's Anomalous

States of Knowledge (ASK) [5] [6]. Belkin calls “Anomalies” the factors that cause people to go to IR systems. ASK is a synthesis of several other ideas, and the process is described as follows: A person begins by realizing she has an information need. She then queries an information retrieval system, and evaluates the information returned from the system. She then decides if her information need has been completely satisfied, only partly satisfied, or not satisfied at all. This inherently raises the question of how to represent what is not known to the information seeker. Part of the ASK model is the concept of non-specifiability of need, which can be cognitive or linguistic. Cognitive non-specifiability can manifest itself as a low level of formedness of need where information seekers cannot define their needs exactly, or as a high level of formedness of need where information seekers know precisely what will satisfy their needs. Linguistic non-specifiability refers to the ability of users to describe the query in the appropriate language (e.g., if users know the kind of knowledge they need but not the keywords to describe it). According to Belkin, when the anomaly in the state of knowledge is at the lower end of the cognitive scale, the state of knowledge is ill-formed and the individual does not know what knowledge will be useful and how to modify his knowledge to meet the information need [5] [6].

2.2 Exploring Academic Literature

Several systems currently exist for exploring and summarizing academic literature, each with different strengths and features. The list below contains an appropriate sample of the available systems and features.

CiteSeerX is an online digital library for scientific literature [42]. It incorporates citation statistics (Figure 2.2), lists of citing papers, and the citation context (i.e. an excerpt of the citing paper which contains the citation, see Figure 2.1). These summaries and statistics provide insight into a paper and its contribution: the number of citations indicate the impact of the paper, the citing papers indicate the domains which are interested in the work, and the citation context indicates why the paper is being cited. CiteSeerX also allows users to view search results sorted by publication year or number of citations. This information is potentially highly valuable for exploring academic literature, but the information is spread apart in many different pages which potentially makes it difficult to see an overview of a domain; each view is essentially limited to each paper rather than an entire discipline. Thus, although CiteSeerX might be a useful system for summarizing individual papers, it is not ideal for summarizing an entire body of literature.

Google Scholar [22] is an application of the Google search interface to articles, patents, and court cases. Figure 2.3 shows a page with search results for the query “Support-Vector Networks”. Search results are similar to CiteSeerX and display the paper’s title, authors, publication venue and date, text excerpts highlighting keyword matches, and citation count. Again, these data can be useful for single-document search, but it may not be useful for multi-document exploration.

GoPubMed [66] shows publications by year and country (both in numeric and visualization form), by journal and keywords (numeric form only), and a network showing coauthorship of top authors. These data can be shown for authors, journals, and search queries. Figure 2.4 shows the top terms and publication counts over time

Home | Statistics | About | Bulletin | Submit Documents | Feedback | MetaCart | Sign in to MyCiteSeerX

CiteSeer^x beta

Documents | Authors | Tables

Support-Vector Networks Search

Include Citations | Advanced Search | Help

1 Searching for **Support-Vector Networks** – sorted by Relevance.

Order by: Citations | Year (Descending) | Year (Ascending) | Recency
 Try your query at: Scholar | Yahoo! | Ask | Bing | CSB

6,193 documents found, showing 1 through 10. Next 10 →

▼ ***Support-Vector Networks***
 by Corinna Cortes, Vladimir Vapnik — 1995 — Machine Learning
 ...**SUPPORT-VECTOR NETWORKS** Corinna Cortes 1 and Vladimir Vapnik 2 AT&T Labs-Research, USA...
 Cited by 1224 (15 self) – Add To MetaCart

▼ ***Identifying Speakers With Support Vector Networks***
 by Michael S. Schmidt — 1996 — In Interface '96 Proceedings
 ...Identifying Speakers with **Support Vector Networks** Michael S. Schmidt BBN Systems & Technologies...
 Cited by 13 (0 self) – Add To MetaCart

▼ ***A neural support vector network architecture with adaptive kernels***
 by Pascal Vincent, Yoshua Bengio — 2000 — In Proc. of the International Joint Conference on Neural Networks
 ...A Neural **Support Vector Network** Architecture with Adaptive Kernels Pascal Vincent & Yoshua Bengio...
 Cited by 3 (1 self) – Add To MetaCart

Figure 2.1: Query results page on CiteSeerX for “Support-Vector Networks”. (1) Search results can be ranked by search relevance, number of citations, publication year, or recency (which attempts to provide results that are both relevant and recent). (2) Excerpts from the documents which match the search query are shown in bold italics, and the citation count is shown at the bottom of each record.

Home | Statistics | About | Bulletin | Submit Documents | Feedback | MetaCart | Sign in to MyCiteSeerX

Documents | Authors | Tables

vapnik Search

Include Citations | Advanced Search | Help

Summary | Related Documents | Version History

Support-Vector Networks (1995) [1228 citations — 15 self]

by Corinna Cortes, Vladimir Vapnik
Machine Learning
Add To MetaCart

1

DOWNLOAD:
<http://www.research.att.com/~corinna/papers/support>
<http://www.cs.toronto.edu/~roweis/csc2515-2006/real>
 CiteULike | DBLP

CACHED:

Add to Collection | Correct Errors | Monitor Changes

Abstract:

The support-vector network is a new learning machine for two-group classification problems. The machine conceptually implements the following idea: input vectors are non-linearly mapped to a very high-dimension feature space. In this feature space a linear decision surface is constructed. Special properties of the decision surface ensures high generalization ability of the learning machine. The idea behind the supportvector network was previously implemented for the restricted case where the training data can be separated without errors. We here extend this result to non-separable training data.

Citations **2**

2155 [Learning Internal Representations by Error Propagation](#) - Rummelhart, Hinton, et al. - 1985

820 [A training algorithm for optimal margin classifiers](#) - Boser, Guyon, et al. - 1992

617 [The use of multiple measurements in taxonomic problems](#) - Fisher - 1936

179 [Theoretical foundations of the potential function method in pattern recognition learning, Autom. Remote Control 25](#) - Aizerman, Braverman, et al. - 1964

152 [Handwritten digit recognition with a back-propagation network](#) - Cun, Boser, et al. - 1990

83 [Learning internal representations by backpropagating errors](#) - Rumelhart, Hinton, et al. - 1986

60 [Learning Logic](#) - Parker - 1985

45 [Comparison of classifier methods: a case study in handwritten digit recognition](#) - Bottou, Cortes, et al. - 1994

41 [Une Procédure d'Apprentissage pour Réseau à Seuil Assymétrique. In Cognitiva 85: A la Frontière de l'Intelligence Artificielle des Sciences de la Connaissance des Neurosciences](#) - Cun - 1985

15 [Classification into two multivariate normal distributions with different covariance matrices](#) - Anderson, Bahadur - 1962

POPULAR TAGS

Add a tag: Submit

OCR

BIBTEX | ADD TO METACART

```
@INPROCEEDINGS{Cortes95support-
  vectornetworks,
  author = {Corinna Cortes and Vladimir Vapnik},
  title = {Support-Vector Networks},
  booktitle = {Machine Learning},
  year = {1995},
  pages = {273--297}}
}
```

BOOKMARKS

OPENURL

Figure 2.2: CiteSeerX’s detailed view of a paper, “Support-Vector Networks” by Corinna Cortes and Vladimir Vapnik. (1) Clicking the link for the number of citations shows all the documents which cite “Support-Vector Networks” and allows users to see the citation context of each citation. (2) The list of papers which “Support-Vector Networks” cites, with citation counts for each citation.

Web Images Videos Maps News Shopping Gmail more ▼ Scholar Preferences | Sign in

Google scholar Support-Vector Networks Search Advanced Scholar Search

Scholar Articles and patents anytime include citations Create email alert Results 1 - 10 of about 126,000. (0.07 sec)

Support-vector networks [\[PDF\] from psu.edu](#)
[Find it @ UMCP](#)
 C Cortes... - Machine learning, 1995 - Springer
 Abstract. The **support-vector network** is a new learning machine for two-group classification problems. The machine conceptually implements the following idea: input vectors are non-linearly mapped to a very high- dimension feature space. In this feature space a linear decision ...
 Cited by 6415 - [Related articles](#) - [BL Direct](#) - [All 39 versions](#)

Text categorization with support vector machines: Learning with many relevant features [\[PDF\] from psu.edu](#)
 T Joachims - Machine Learning: ECML-98, 1998 - Springer
 ... 1. C. Cortes and V. Vapnik. **Support-vector networks**. Machine Learning, 20:273- 297, November 1995. 2. T. Joachims. A probabilistic analysis of the rocchio algorithm with tfidf for text categorization. In International Conference on Machine Learning (ICML), 1997. 3. T. Joachims. ...
 Cited by 4233 - [Related articles](#) - [Library Search](#) - [BL Direct](#) - [All 34 versions](#)

[\[PDF\] LIBSVM: a library for support vector machines](#) [\[PDF\] from psu.edu](#)
 CC Chang... - 2001 - Citeseer
 Page 1. LIBSVM: a Library for **Support Vector** Machines Chih-Chung Chang and Chih-Jen Lin* Last updated: January 3, 2006 Abstract LIBSVM isa library for **support vector** machines (SVM). Its goal is to help users to easily use SVM as a tool. ...
 Cited by 3873 - [Related articles](#) - [View as HTML](#) - [All 33 versions](#)

A tutorial on support vector machines for pattern recognition [\[PDF\] from psu.edu](#)
[Find it @ UMCP](#)
 CJC Burges - Data mining and knowledge discovery, 1998 - Springer
 ... and to alert the reader to some of the literature we summarize some recent applications and

Figure 2.3: Query results page on Google Scholar for “Support-Vector Networks”. Search results can be filtered by recent years. Excerpts from the documents which match the search query are shown in bold text, and the citation count is shown at the bottom of each record.

for the author Antonella Tosti, and Figure 2.5 shows her collaboration network. Unfortunately, the visualizations and statistics are static, which forces the website to reload whenever the focus changes to a different keyword, author, or journal. Furthermore, the summaries listed above are not available on a per-paper basis, which potentially limits the analysis that can be performed.

Web of Knowledge [63] can create a citation map using hyperbolic trees which show a hierarchy of papers and the papers that cite them, as shown in Figure 2.6. These citation maps support a limited amount of dynamic interaction, mainly for changing the focal point of the map. Other interactions, such as filtering by publication year, require recreating the citation map which slows down the interaction and exploration process.

Of course, no system can work with documents that are not already part of its library. One of the limitations of the above systems is that there is no way to load additional documents, or to filter out all but a subset of the documents. The systems described below allow users to define custom libraries.

Reference managers such as JabRef [31], Zotero [9], EndNote [64], and Mendeley [35] also provide search features and some limited summary capabilities. For example, users can import a library of documents into Mendeley and then use it to search the full-text of the document, write notes on a document, and view limited statistics such as the number of publications per author and the most frequent publication venues. These statistics either can be aggregated across all documents in a user's own library or aggregated across all documents from all Mendeley users' libraries (see Figures 2.7 and 2.8). However, even though a custom-built library could

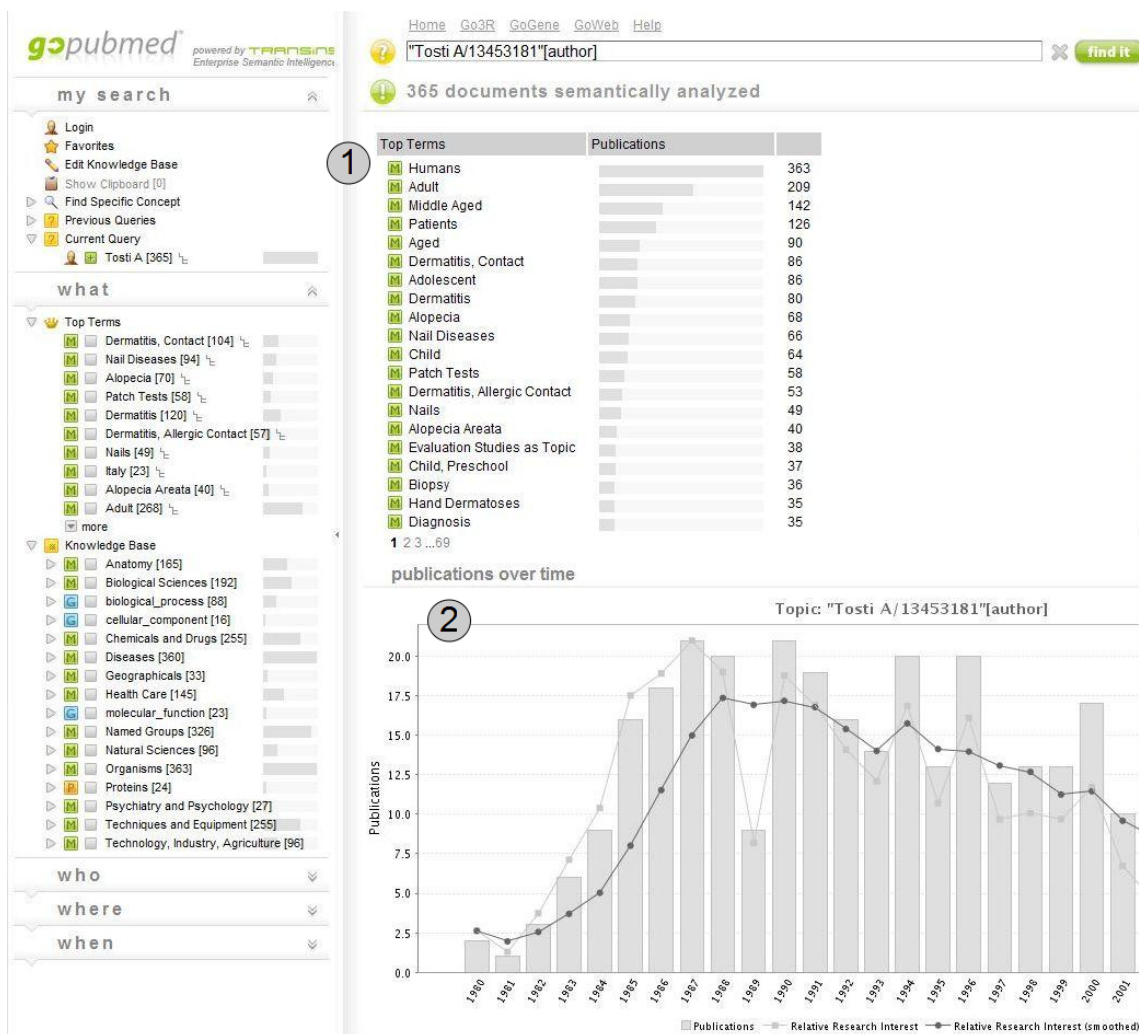


Figure 2.4: A view in GoPubMed of the author Antonella Tosti which shows (1) a summary of top terms, and (2) a bar chart of publication counts over time.



Figure 2.5: A view in GoPubMed of the author collaboration network for Antonella Tosti.

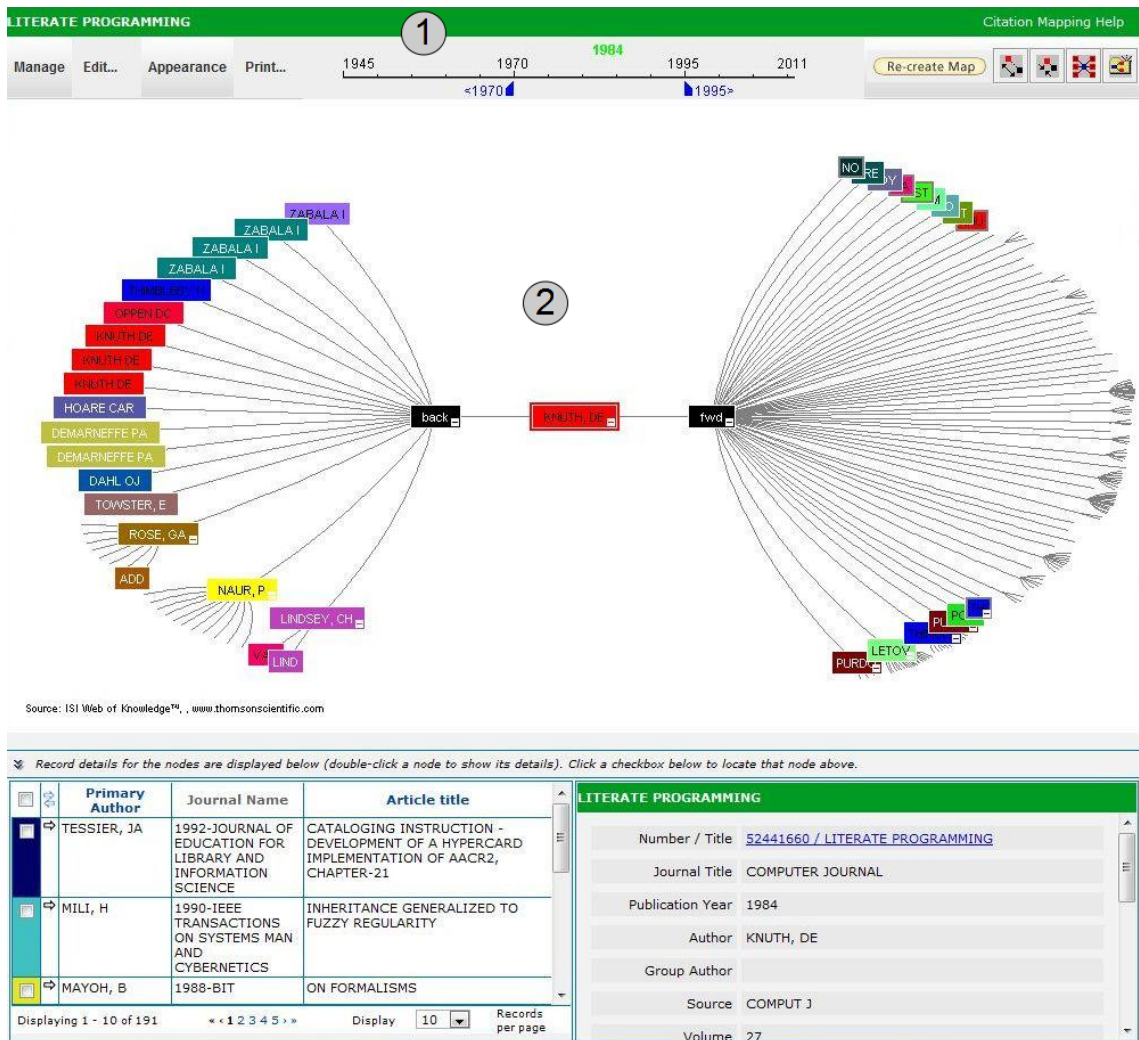


Figure 2.6: A citation map in Web of Knowledge of “Literate Programming” by Donald Knuth. This network can be filtered by publication year (1)—in this case, the network is filtered to show only papers published between 1970 and 1995. The citation map (2) shows the backward citations (i.e., papers cited by “Literate Programming”) and forward citations (i.e., papers that cite “Literate Programming”).

promote improved searching, many of these features in reference managers are not any better for summarizing groups of papers than digital libraries such as CiteSeerX and Google Scholar, and in some ways are worse because reference managers do not tend to provide lists of citing papers or citation context.

XplorMed is a system for summarizing literature search results by keywords and word context [45]. Analyses are performed on abstracts, and results show keywords and keyword chains, plus their frequencies of occurrence in the abstracts along with the abstracts themselves. Users can iteratively refine a corpus by repeatedly filtering out documents with low keyword frequencies. Figure 2.9 shows screenshots of this process in Xplormed. XplorMed also allows users to supply their own database to analyze. However, the system does not allow users to visualize the data.

Recommender systems can also be used to aid the exploration of academic literature [65] [34] [23]. These systems work by recommending papers based on a given input, such as a paper or author name. Ekstrand et al. [18] use a novel approach that incorporates the influence of a paper when recommending introductory academic reading lists, but the authors assume that users already have a few papers in the field they want to learn about. Although these systems can recommend relevant papers, they do not necessarily give users an overview of a domain, nor do they freely allow users to explore the literature.

Boyack and Börner [8] used VxInsight [15] to visualize interconnections between research areas, institutions, grants, publications, and authors. The relationships were visualized as 3D landscape maps with points representing keywords, institutions, and other fields (see Figure 2.10 for an example). Boyack and Börner's

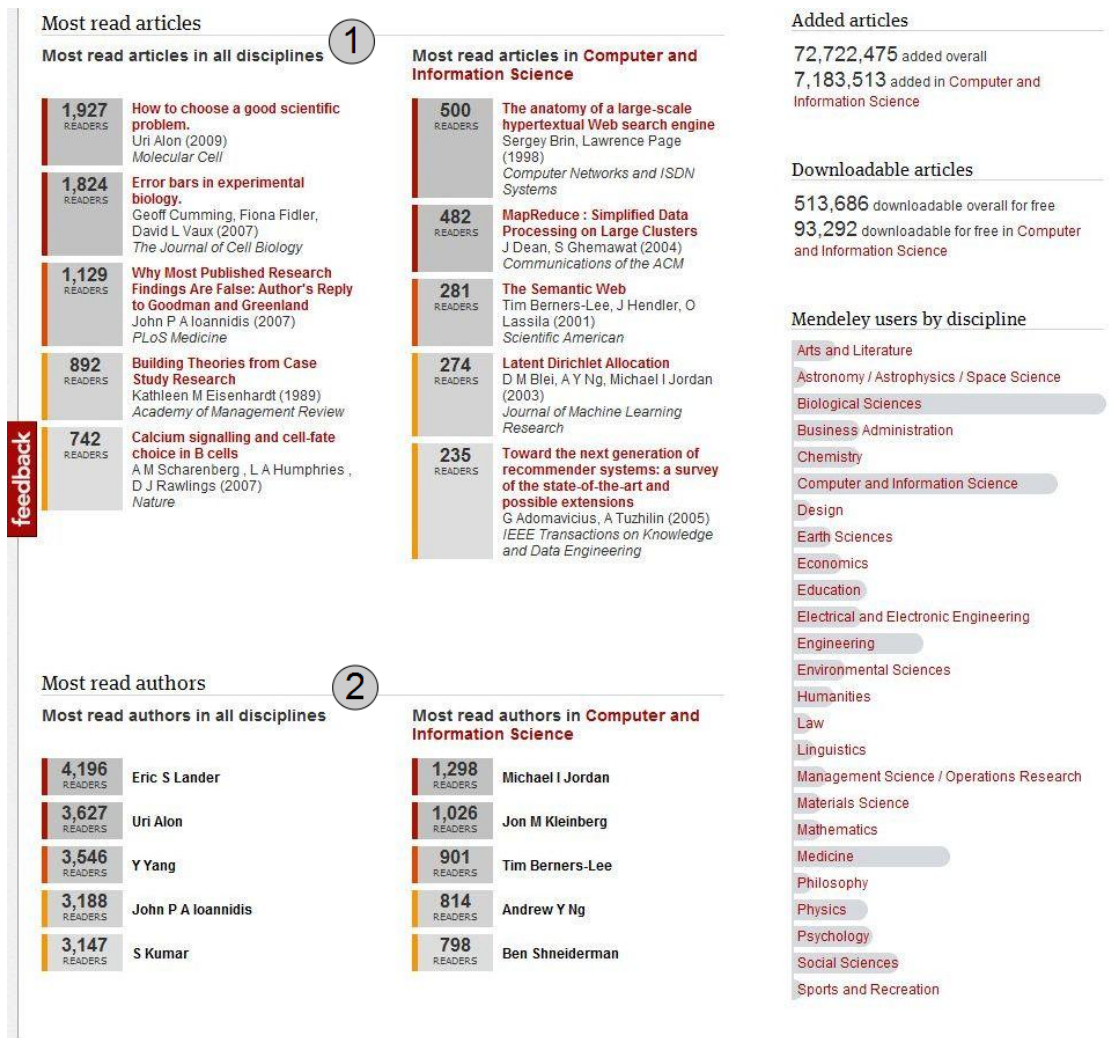


Figure 2.7: Mendeley can aggregate information from all users' libraries and show users (1) the top read articles in one or all disciplines, and (2) the most read articles in one or all disciplines.

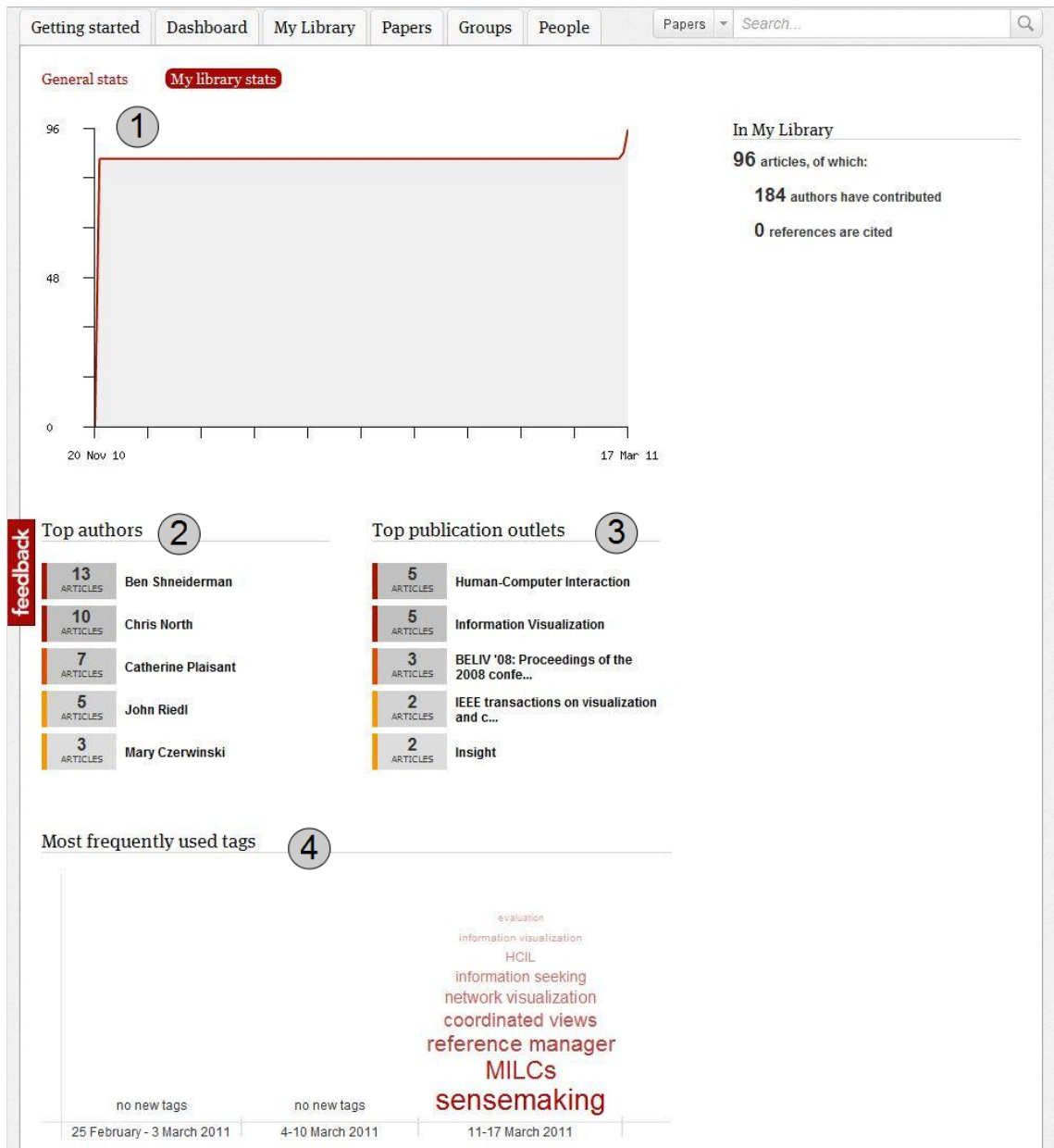


Figure 2.8: For users' own personal libraries, Mendely can show (1) the growth in users' libraries over time, (2) the top authors, (3) the top publication outlets, and (4) the most frequently used tags.

Query: 300226899
step 2: Select abstracts from one or more MeSH categories

MeSH Category	#	abstracts
Psychiatry and Psychology	0	none
Health Care	2	20981785, 20932923
Humanities	0	none
Analytical, Diagnostic and Therapeutic Techniques and Equipment	2	20981785, 20932923
Geographic Locations	0	none
Persons	0	none
No assigned MeSH	15	21093976, 21129400, 21134616, 21147619, 21194687, 21212321, 21242396, 21264858, 21288113, 21299445, 21315862, 21322384, 21330763, 21342171, 21345538
Diseases	0	none
Information Science	0	none
Anatomy	0	none
Technology and Food and Beverages	0	none
Organisms	0	none
Physical Sciences	0	none
Chemicals and Drugs	4	21044709, 20981785, 20932923, 20940690
Anthropology, Education, Sociology and Social Phenomena	1	20932923
Biological Sciences	1	20932923

Select abstracts by year: From To

(a) MeSH categories.

Query: 300226899
iteration 1 - step 3: Compute word-chains between the most related words

After the previous selection step, your set contains now 16 abstracts.
Warning: there may be not enough information for carrying on. Why?

Words ordered by association score

- 1.000 patient_irl
- 0.667 imaging_irl
- 0.667 artefact_irl
- 0.667 quality_irl
- 0.667 brain_irl
- 0.667 image_irl
- 0.667 resonance_irl
- 0.365 copy_irl
- 0.365 hiv_irl
- 0.365 xna_irl
- 0.365 resistance_irl
- 0.284 feature_irl
- 0.284 mark_irl
- 0.111 france_irl
- 0.111 italy_irl
- 0.111 immunization_irl

Click on the [R] to explore the relation of the corresponding word to other words

Control the number of relations that are going to be displayed in the next step by modifying the values of these parameters:

Alpha: (0.1-0.9)
Score: (0-0.9)

NEXT ACTION:

(b) Word associations.

Query: 300226899
iteration 1 - step 4: Choose word-chains to refine your query
Now you can make a selection on your current set of abstracts based on your topic(s) of interest.

List of word chains ranked by number of abstracts where they were found

- 2 t, resonance
- 2 ma, resistance
- 2 resonance, t, quality
- 2 resonance, t, imaging
- 2 resonance, t, image
- 2 resonance, t, brain
- 2 resonance, t, artefact
- 2 resonance, t
- 2 resistance, ma, hiv
- 2 resistance, ma, copy
- 2 resistance, ma
- 2 patient, hair
- 2 patient, feature
- 2 patient, alopecia
- 2 italy, immunization
- 2 immunization, italy

or type another set of words:

In the next step display links from abstracts to: OMDM, SwissProt, SMART domains

NEW! Make a summary of the MeSH terms of class:

NEXT ACTION:

(c) Word chains.

Figure 2.9: XplorMed users can filter their document set by categories and year (a), see word association scores (b), and filter documents by ranked word chains (c).

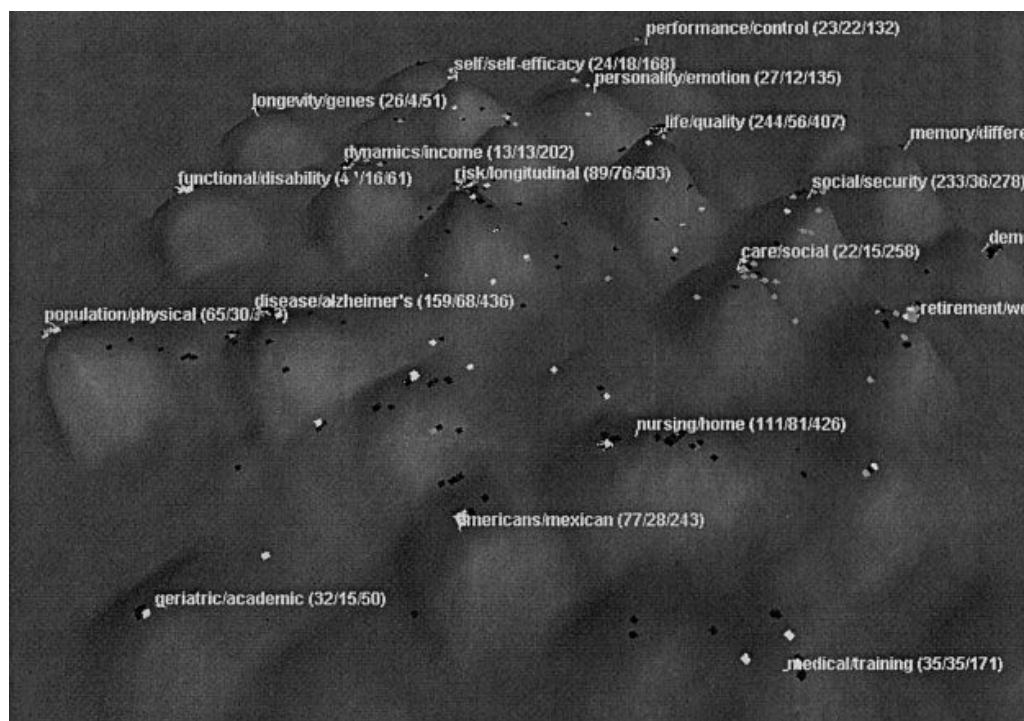


Figure 2.10: Boyack and Börner’s visualization in VxInsight of Behavioral and Social Research (BSR) Program grants and publications clustered by Medical Subject Heading (MeSH) terms. Labels show common keywords, and height indicates number of documents. Image is from [8].

primary goal was to explore these relationships, but the landscape maps also show the relative sizes of research areas and sub-areas, which could be useful for exploring growing sub-topics. The authors noted several difficulties in collecting and linking datasets, highlighting some of the issues with cleaning and merging data from disparate sources. Although their visualizations yielded interesting results, a purely visualization-driven interface would have difficulty revealing details of the research.

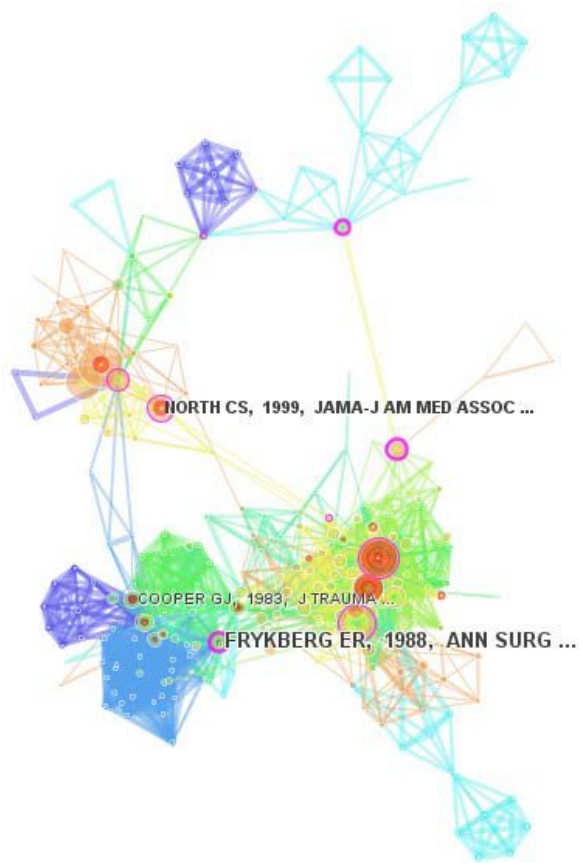
CiteSpace II uses purple rings and burst-detection to indicate sudden activity in networks (see Figure 2.11(a)), indicating breakthroughs and foundational

work in scientific literature networks [10]. Chen et al. [12] summarize author co-citation and document co-citation networks used for providing overviews of the field of information science. They go on to introduce a co-citation analysis technique in CiteSpace II using structural, temporal, and semantic patterns plus citations and automatic clustering and labeling. One aspect of this is a co-citation timeline shown in Figure 2.11(b). The authors then used these methods to identify specialties within the information science domain.

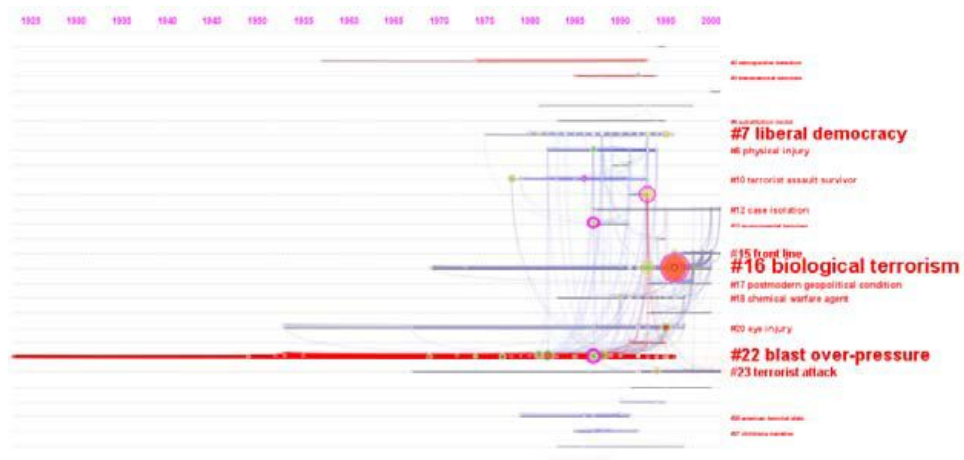
Trimmer is a system that produces text summaries of scientific literature in the context of the iOPENER project [71] [36] [72]. Several sources may serve as input to Trimmer, including full papers, abstracts, and citation texts (the set of sentences from other papers that refer to a given paper). Summaries can be helpful for gaining understanding about unfamiliar corpora, but Trimmer was not designed to explore literature networks.

The iOPENER project also explored the use of a network visualization of citations to reveal research fronts and their relationships on a semantic substrate that grouped related papers in separate screen regions [3].

News summarizers also attempt to summarize topics and corpora. One example is NewsInEssence [50] which summarizes news topics, but does not necessarily account for the corpus structures which are relevant for scientific literature domains. As such, these summaries might not be successful at telling the story of the evolution of a field. NewsCube [41] allows aspect-level news browsing and classifies articles by similarities of aspects, or viewpoints, which allows partitioning not only by topic but by author perspective. However, neither summarization nor aspect-level browsing



(a) Citation bursts are shown by purple rings.



(b) Co-citation cluster timeline.

Figure 2.11: CiteSpace II can highlight sudden increases in citations with bursts of color (a) and show co-citation clusters over the timeline of the network (b).

Functionality	Search Engines						Ref. Mangrs				Sum. & Rec.			
	CiteSeerX	Google Scholar	GoPubMed	Web of Knowledge	ACM Portal	IEEE Xplore	JabRef	Zotero	Mendeley	EndNote	XploreMed	Recommender Systems	NewsInEssence	Trimmer
Summary of textual excerpts													•	•
Custom database							•	•	•	•	•	•	•	•
Create notes							•	•	•	•				
Ranking	•			•	•	•	•	•	•	•				
Full-text search	•	•						•	•	•			•	
Document recommendations	•	•	•	•								•	•	
Document statistics	•	•	•	•	•	•								
“Cited by” list	•	•		•	•	•								
Search excerpts	•	•							•					
Corpus statistics			•						•		•			
Keyword summary			•						•		•			
Citation visualization			•	•										
Citation context	•													

Table 2.1: Table of capabilities by system, grouped by system type (search engines, reference managers, and summarizers and recommenders).

alone will provide both overviews and details on demand of a pool of articles.

Table 2.1 summarizes the above systems by the following capabilities, defined in terms of a given paper c , a set of papers C (that include c) loaded in the system, and a superset corpus S (e.g., the Association of Computational Linguistics Anthology):

Summary of textual excerpts Automatically generated summaries from textual excerpts from c or documents citing c .

Custom database The capability to load a custom document database C into the system.

Create notes Tool for creating and storing notes in the system about c or some subset of C .

Ranking A ranking of documents in C or S , using features such as publication year, author, citation count, etc.

Full-text search A query tool that searches text in c in addition to the metadata associated with c .

Document recommendations Recommendations of documents in C or S of likely interest to users.

Document statistics Statistical information about c .

“Cited by” list A list of documents in C or S that cite c .

Search excerpts Search query results that show text excerpts highlighting the query keywords in c .

Corpus statistics Statistical information about C or S .

Keyword summary A list of the most prominent keywords or phrases in c , C , or their metadata.

Citation context The citing text that shows the context of the citation of c .

Citation visualization A visualization showing a partial or complete citation network, possibly centered on a single paper c .

The digital libraries, such as CiteSeerX and Google Scholar, tend to support literature exploration and features for summarizing articles and authors, but are less likely to support functionality to summarize an entire field or corpus such as corpus-level statistics. Reference managers have the benefit of supporting custom-built databases and sorting references by various properties, but generally lack information on citation links between papers. Recommender and summarization systems provide automatic reports on a corpus, but the ones surveyed here do not support citation-level summaries or statistics. Overall, one can see a trend where each class of system tends to support a specific set of functionalities. In particular, most of these applications lack visualizations and the ability to see the citation context which give the context of citations. Each application has a different target audience, and across all applications there seems to be a lack of support for effectively summarizing and exploring academic literature.

Interestingly, the systems that support document statistics do not usually support corpus statistics, even though these seem to be similar and related capabilities. The systems that provide document statistics are more likely to also provide document recommendations and a “cited by” list instead.

2.3 Evaluation

Human-subjects evaluation of bibliographic systems and digital libraries is important to analyze the utility and acceptance of the systems. Xie [68] highlights many issues surrounding user evaluation of digital libraries. According to Xie, usability is the most investigated evaluation criterion. She performed an evaluation with graduate students in Information Studies, who may have a different perspective on digital libraries from non-professional information seekers. The study compared the use and perception of the American Memory and University of Wisconsin Digital Collections digital libraries, and found that users regarded usability and system performance as the most important criteria. Other studies have compared two or three IR systems or digital libraries [47] [70] [74].

Scatter/Gather [13] is a browsing technique for large corpora. Scatter/Gather clusters similar documents, and then allows users to iteratively select documents and clusters, and then re-cluster the selection. Pirolli et al. [47] compared Scatter/Gather to a word-based search interface in their abilities to retrieve documents and enhance users' understanding of the document collection. The evaluation methodology required participants to locate documents on a specific topic, and draw diagrams of the document collection structure. They found that Scatter/Gather alone is not as effective as other kinds of information retrieval tools for finding specific documents; however, Scatter/Gather can be useful for exploratory browsing.

Yuan et al. [70] compared user perception of retrieval results visualization sys-

tem and a text-based information retrieval system. They used CiteSpace and Web of Science as the visualization and text-based systems respectively. The researchers conducted a user study with search-based and browsing-based tasks, and then measured participants' perceptions. Participants perceived CiteSpace as more usable for the tasks, and users felt that the visualization system conveyed the information significantly better than the text-based system.

Zhang et al. [74] conducted usability studies to examine the usability of interaction design in the ACM, IEEE Xplore, and IEEE Computer Society digital libraries. As with the study in Yuan et al., the participants performed search-based and browsing-based tasks; however, this study was augmented with user feedback. Zhang et al. conclude by providing guidelines for the interface design of other digital libraries.

Hill et al. [26] conducted user evaluations of three versions of the Alexandria Digital Library. They used several evaluation methodologies, and the researchers used the results to iteratively improve the interface.

A major issue in user interfaces is that of evaluation: widely used quantitative techniques in tightly controlled laboratory conditions may not be adequate for capturing important aspects of system usage, such as creativity and discovery [7] [11] [59]. Common metrics in HCI research often focus on measuring the time to complete a task or the number of errors while performing the task [7]. One alternative to task completion time or number of errors is to quantitatively or qualitatively analyze the insights gained from using the system [54] [69] [60] [53]. Other alternatives have been proposed for evaluating effectiveness of different presentation approaches—

such as the application of ROUGE or Pyramid metrics for automatic evaluation of textual summaries of academic literature [36]—but these methods do not involve user evaluation nor do they assess visualization and exploration capabilities.

Qualitative evaluation methods for information visualization systems are becoming more common [59], which include cognitive walkthroughs [1], Grounded Theory [20], longitudinal studies [53] [55] [44], surveys [27] [55] [73], and case studies [55] [56] [44]. This reflects a growing shift that emphasizes the need for participants to use systems under naturalistic conditions similar to their everyday situations [49] [59].

Short-term studies—such as one- to two-hour evaluation sessions—when applied to data exploration systems are criticized because they are unable to capture the sort of insights gained by long-term data analysis [53] [49] [59] [67]. In contrast, longitudinal case studies have been successfully used to evaluate several information visualization systems [21] [55] [53].

Questionnaires are occasionally used to supplement other evaluation techniques with a quantitative component [55]. Questionnaires can be especially useful for measuring participant preferences and providing overviews [24]; however when used alone they may not provide sufficient depth to provide useful insights about the system under evaluation [25].

Several researchers highlight the need to have research participants analyze their own data, which keeps participants motivated and gives the system significance [53] [20] [59] [28] [49] [24]. They argue that systems to support discovery should be tested in more realistic situations where users will try to make discoveries, and

users will be more interested in making discoveries if they are important to the users. Thus, users may be less motivated to explore data and less likely to make discoveries if the discoveries and data are less significant to the users.

Although qualitative evaluations can allow more flexibility for studying certain aspects, such as long-term effects or allowing participants to use their own data, these methods have their own limitations: as evaluators reduce carefully controlled conditions they risk reducing the generalizability of their results. However, an emerging idea is that by “triangulating” with multiple evaluation methods these limitations can be minimized [59]. If multiple evaluation methods all indicate the same pattern of usage, then there is some reassurance in claims about the system’s usage.

The challenges of effective evaluation are some of the motivations for advocating Multi-dimensional In-depth Long-term Case studies (MILCs) as described by Shneiderman and Plaisant [59]. By using multiple evaluation methods in long-term case studies where participants analyze their own data, we can gain meaningful understanding of how the system is actually being used and whether it has the necessary functionality for users to perform the tasks they wish to perform.

Perer and Shneiderman used a methodology inspired by MILCs to evaluate SocialAction [44]. Their approach was to use interviews and long term case studies over 4–8 weeks with four participants who were each analyzing their own data. The case studies detailed the stages of users’ development of SocialAction’s usage, including training, early use, mature use, and outcomes. However, adding other dimensions to the study may have revealed insights such as which features are used

most commonly for certain types of analyses.

Seo and Shneiderman also used a MILC-like methodology for evaluating Hierarchical Clustering Explorer [55] by using questionnaires and three long-term case studies. The case studies consisted of interviews and participant observation which lasted for several months. In an attempt to increase the generalizability of the case studies the evaluators recruited participants from different fields. The questionnaires provided insight into frequency of use, feature usage, and impact on data analysis.

Multiple-window and multiple-view coordination, such as that in ASE, can be an important feature [40]. However, effective evaluation of multiple-coordinated views is not easy [52]. Kandogan and Shneiderman evaluated Elastic Windows using a quantitative experiment followed by interviews to debrief participants [32]. Namata et al. performed case studies with two participants of their dual-view of a network visualization [37], but there is no indication of the length of the case studies.

Because ASE incorporates multiple-coordinated views of multiple tools, it fits best on large high-resolution displays. Shupp et al. note the majority of research on large high-resolution displays has not focused on the impact and benefits of larger displays [60]. However, there is research supporting the benefits of large high-resolution displays, such as a preliminary comparative study performed by Czerwinski et al. [14] which showed significant performance improvement, and a study by Andrews et al. which showed that larger space can improve sensemaking [2]. We suspect that larger higher-resolution displays will benefit the usability of ASE, but there may be a point of diminishing returns.

Regardless of the system being evaluated, it is important to appropriately use exploratory, formative, and summative evaluations [19]. Due to the nature of MILCs, and considering the current state of ASE, I performed an extended usability study using qualitative methods as part of the iterative evaluation process to identify ASE's utility as well as potential improvements to the system.

2.4 Description of Action Science Explorer

Action Science Explorer¹ (ASE) is a system developed by Dunne et al. [16] [17] which can be used to explore scientific literature domains. A screenshot is shown in Figure 2.12. The system integrates bibliometrics, various summarization techniques, search features, and visualization tools. This is a combination of several features seen in other search engines and reference managers plus other features that, to the authors' knowledge, are novel in literature exploration and search tools.

Two of the main components are the JabRef reference manager [31] and the SocialAction network analysis tool [43]. JabRef contains many features found in typical reference managers: searching, grouping papers, sorting the bibliography, viewing abstracts, creating text annotations, viewing the full-text PDF, and importing and exporting many different formats.

SocialAction supplies a powerful suite of network analysis tools. Among these are a visualization of the citation network, paper rankings by various network metrics, filtering papers by metric values, automatic cluster detection, and scatter plots

¹More information and a video are available at the ASE website: <http://www.cs.umd.edu/hcil/ase/>

of paper attributes.

A third set of views provides several text views and summaries of the documents. If a paper or cluster of papers is selected, the In-Cite Text view shows a list of all sentences from other papers that are citing the selected paper(s). Each of these sentences is in turn a hyperlink. Clicking on one will open the full-text of the source paper in the Out-Cite Text view, which includes citation highlighting to show other hyperlinked citations to other papers. If the user has selected a cluster of documents, then the In-Cite Summary view will show an automatically generated summary of all the sentences shown in the In-Cite Text view.

All of these views are linked together, showing multiple coordinated views of the document set. When a paper is selected in the reference manager, it becomes highlighted in the citation network visualization. If the user clicks a node in the citation network, the citation text is shown in the In-Cite Text view. If a user clicks a link in the Out-Cite Text view, the corresponding paper becomes selected in the network visualization. Furthermore, each view can be detached into its own dockable window, allowing ASE to be rearranged and customized to fit a variety of displays (Figure 2.13 demonstrates the dockable window feature).

Below are detailed descriptions of the most prominent features of ASE, defined in terms of a given paper c , a set of papers C (that includes c) loaded in the system, and a superset corpus S (e.g., the Association of Computational Linguistics Anthology):

Attribute Ranking Ranking tool for sorting the nodes (papers) by network met-

The screenshot displays the Action Science Explorer system interface, which is divided into several functional areas:

- Top Panel:** A menu bar with options like File, Edit, View, BibTeX, Tools, Web search, Plugins, Options, and Help. Below it are icons for various actions and a search bar.
- Bibliography List (1):** A table listing bibliographic entries with columns for #, Act-ID, Entytype, Author, Title, Year, Booktitle, Bibbkey, and Owner. The first entry is circled in red.
- Search and Filter (3):** A search box with a magnifying glass icon and a 'Settings' button. Below it are radio buttons for 'Incremental', 'Float', 'Filter', and 'Global search'.
- Ranking and Visualization (5):** A 'Rank Nodes' section showing a list of nodes with their ranks and in-degree values. A 'Network Visualization' section shows a graph with nodes and edges, with a 'Find Communities' button.
- Full-Text Article (9):** A detailed view of a selected article, including its title, author, and abstract. The abstract text is highlighted in green.
- Summary and Text (7, 8):** A section titled 'In-Cite Text (48 entries)' and 'In-Cite Summary (Community 70-4...)' showing a multi-document summary of the cluster.

Figure 2.12: The Action Science Explorer system: (1) the bibliography list, (2) bibliographic entry details, (3) bibliographic search, (4) bibliographic groups, (5) the attribute ranking view, (6) the citation network visualization with Find Communities enabled, (7) the In-Cite Text, (8) the multi-document In-Cite Summary of a cluster, and (9) the full-text article with linked citations.

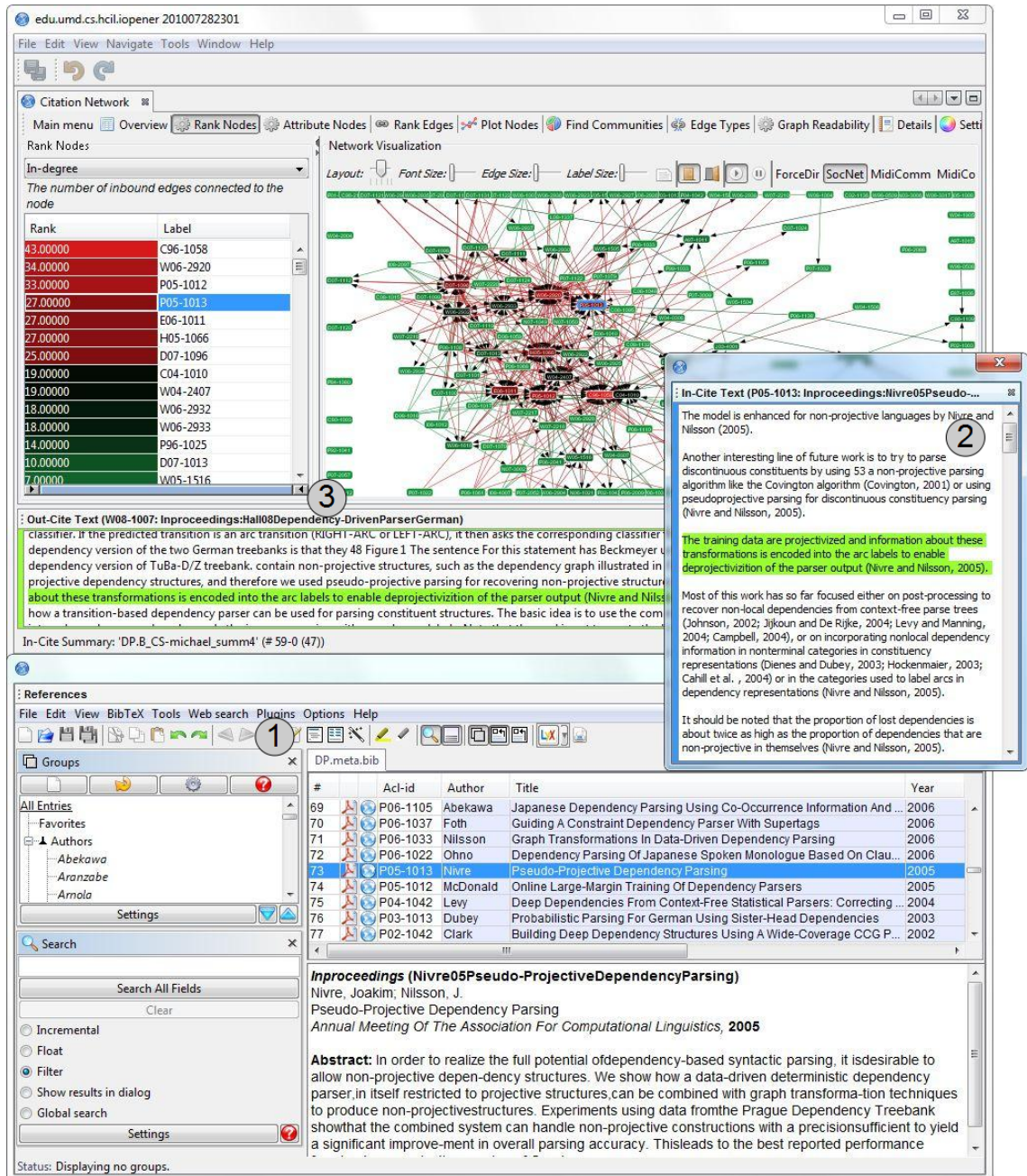


Figure 2.13: The different views in Action Science Explorer can be hidden or detached into separate windows. Here, the In-Cite Summary view is hidden, and the Reference Manager (1) and In-Cite Text (2) views have been detached from the main window (3) into their own windows and rearranged to fit a vertical display.

rics, such as document set in-degree (number of times c is cited in C), document set out-degree (number of papers in C cited by c), corpus in-degree (number of times c is cited in S), and corpus out-degree (number of papers in S cited by c).

Find Communities Clustering tool that uses Newman’s community detection algorithm [38] to cluster documents in C or S based on citations. A user-movable slider sets clustering size, ranging from a single large loosely connected cluster to a single small tightly connected cluster (with several mid-sized clusters halfway in between).

Search Search tool in JabRef that supports basic regular expression and logical operators, and can search any fields in the bibliographic entry for c , including title, author, abstract, keywords, and user-added notes.

Sorting Sorting tool in JabRef that sorts bibliographic entries in C or S by any field, such as title, author, or abstract.

Groups Tool in JabRef for creating, and adding to, groups of papers in C or S based on authors, keywords or keyword pairs in the bibliographic entries. Some pre-computed groups of papers are created. These pre-computed groups are based on authors and keywords, allowing users to find papers from a particular author or papers which use the same keywords.

Import/Export Citations Exporting tool in JabRef for exporting citations to Word, LaTeX, and OpenOffice.org.

In-Cite Text Tool for displaying sentences from other papers in C or S that cite c . Each sentence is a hyperlink that opens the full-text of the citing paper in the Out-Cite Text view.

In-Cite Summary Tool for displaying an automatically generated summary of all sentences that cite c , using the Trimmer software [71]. This is a summary of the text show in the In-Cite Text view.

Out-Cite Text Tool for displaying the full-text of the source paper c , with sentence highlighting to show other hyperlinked sentences to other papers C .

Together, these features provide a multi-faceted approach to explore, search, and analyze a literature network. Of particular interest are the Attribute Ranking, Filtering, and Find Communities features, which, to my knowledge, are novel to literature search and exploration software. For example, Attribute Ranking allows users to rank documents several metrics, including network statistics not found in other systems. These features potentially allow users to quickly identify papers with key attribute values and groups of papers clustered around a common theme.

Early formative evaluations helped guide the development of ASE and provided a foundation for the evaluation presented in this paper. These early evaluations consisted of four participants over a period of about one year. Two of the participants used ASE individually over several 1- to 2-hour sessions which included assigned tasks and free exploration. The data were analyzed qualitatively, and the results informed ASE’s development and helped refine the ASE training methods. The other two participants used the system together to explore the dependency

parsing dataset (described in Section 3.1.2). These two participants used ASE to analyze the corpus and create a report which described several insights about the corpus as well as their comments and suggestions to ASE's developers.

Chapter 3

Evaluation and Improvements

Choosing an evaluation methodology was one of the first steps in evaluating ASE. One of the shortcomings in the current version of ASE is that it requires a non-trivial amount of data cleaning and processing to import a corpus into the program. This made it impractical to perform evaluations using the MILCs framework for ASE because the only data set that had been imported was a set of 147 papers on dependency parsing. Extended usability studies with computational linguistics students and postdocs offer a compromise evaluation strategy: participants are motivated to analyze the dataset because it is related to their work, but these usability studies do not require importing new datasets for each participant.

3.1 First Usability Study

In order to evaluate the usability and effectiveness of Action Science Explorer (ASE), I conducted several usability evaluation sessions.

3.1.1 Participants

The qualitative evaluation sessions were performed with four participants who will be referred to as P1, P2, P3, and P4. Two of the participants were current Computer Science Ph.D. students and two were recent graduates. The ages ranged

from 24 to 32 years old. Two of the participants were female.

All of the participants are currently working on computational linguistics. P2 and P4 are notable for already having some experience with dependency parsing, describing themselves as having “a little bit” of experience, and being “quite familiar” with dependency parsing respectively. P2 also had some previous experience using JabRef, but did not appear to have experience with anything except the most basic features.

3.1.2 Experimental Design

The evaluations were conducted on an Intel Core i3 2.26 Ghz laptop with 4 GB of RAM. The ASE system was displayed on an external 30-inch monitor at a resolution of 1920x1080.

I limited each session to two hours, and began by having participants read and sign the IRB-approved informed consent form. Next, participants went through a 30 minute training session. Participants were then asked to perform some predefined tasks, which usually lasted for one hour. For the last 30 minutes, participants were asked to perform tasks they thought would be useful. At the end of the session participants were asked to provide comments on their experience using the system. After the training session, the participants were asked to think-aloud while they performed the tasks, and I took notes about the participants’ thoughts and actions.

For the sessions, I setup a potential use case by loading the system with a set of 147 papers taken from the Association for Computational Linguistics (ACL)

Anthology Network [51] which were isolated by performing a keyword search for “dependency parsing”. Dependency parsing is a subfield of computational linguistics and is a way to analyze sentences according to which components depend on others. All papers were published between 1969 and 2008. The dataset includes common reference fields, such as author, title, publication year, and publication venue, as well as a field for the unique ACL ID. The ACL ID is used for the network visualization label, and the ID is based on the publication venue and year (e.g. W03-0501 is taken from the Human Language Technologies and North American Association for Computational Linguistics Text Summarization Workshop in 2003).

3.1.3 Procedure

Participants were recruited by announcing the evaluation at a computational linguistics seminar, and \$30 USD was offered for compensation. The participants signed informed consent forms and were given the opportunity to withdraw from the study at any time, as per IRB requirements.

The reason for recruiting participants with experience in computational linguistics was to address the motivation issue discussed in Section 2.3. Participants familiar with computational linguistics will be more likely to understand and want to explore documents on dependency parsing.

Each participant was in a separate evaluation session which consisted of two parts: a training session and an experimental session. During the training stage, the participants were shown a series of video clips describing ASE and demonstrating its

features. After each video, the participant was asked to practice by performing the tasks shown in the videos. The participants were also given opportunities during the training session to ask questions if they did not understand the system or its features.

The remainder of the session was divided into two parts. In the first part, the participants were asked to perform the following main tasks:

- Identify authors and papers which appear important to the participant and store them in groups.
- Select an important paper and collect evidence to determine why it is important.

This provided a common benchmark for performance of basic functionality in ASE, as well as observing variation in the depth of analysis that participants achieved.

For the second part, the participants were asked what they would want to find out about the dataset. One or more of these goals were selected as tasks for that individual participant, and the participant was asked to perform that task or tasks. This added some unique relevance to each participant in their evaluation session.

The data in Tables 3.1, 3.2, and 3.3 show the system capabilities that each participant used in each task. The tables were created by analyzing my observation notes from each participant. The system capabilities in the top row were selected by choosing the capabilities which were demonstrated in the training video. The system capabilities in the top row were chosen selecting all capabilities which met

two criteria: (1) the capabilities were demonstrated in the training video, and (2) the capabilities support exploring, viewing, summarizing, or grouping papers or meta data such as dates or abstracts. A system capability is marked in a participant’s row if the participant used that capability at least once during the task.

3.1.4 Results

There were two sets of tasks during the evaluation sessions: predefined tasks and user-defined tasks. For all participants the predefined tasks were performed first, followed by the user-defined tasks.

3.1.4.1 Predefined tasks

During the first predefined task (to identify important papers and authors), all participants created two groups in the reference manager for storing the most important authors and papers. Participants P1 and P3 relied mostly on node ranking to determine prominence, whereas P2 and P4 relied more on their prior knowledge of the field.

Participant P1 took the simplest approach and initially only used attribute ranking to select important papers and authors who wrote highly cited papers. Later P1 gained deeper insight by scanning the In-Cite Text and observing authors who were frequently mentioned in other papers. P1 also discovered an author who seemed to have done a lot of work on efficient algorithms, and this indicated to P1 that the author must be important, regardless of citation count (see Figure 3.1).

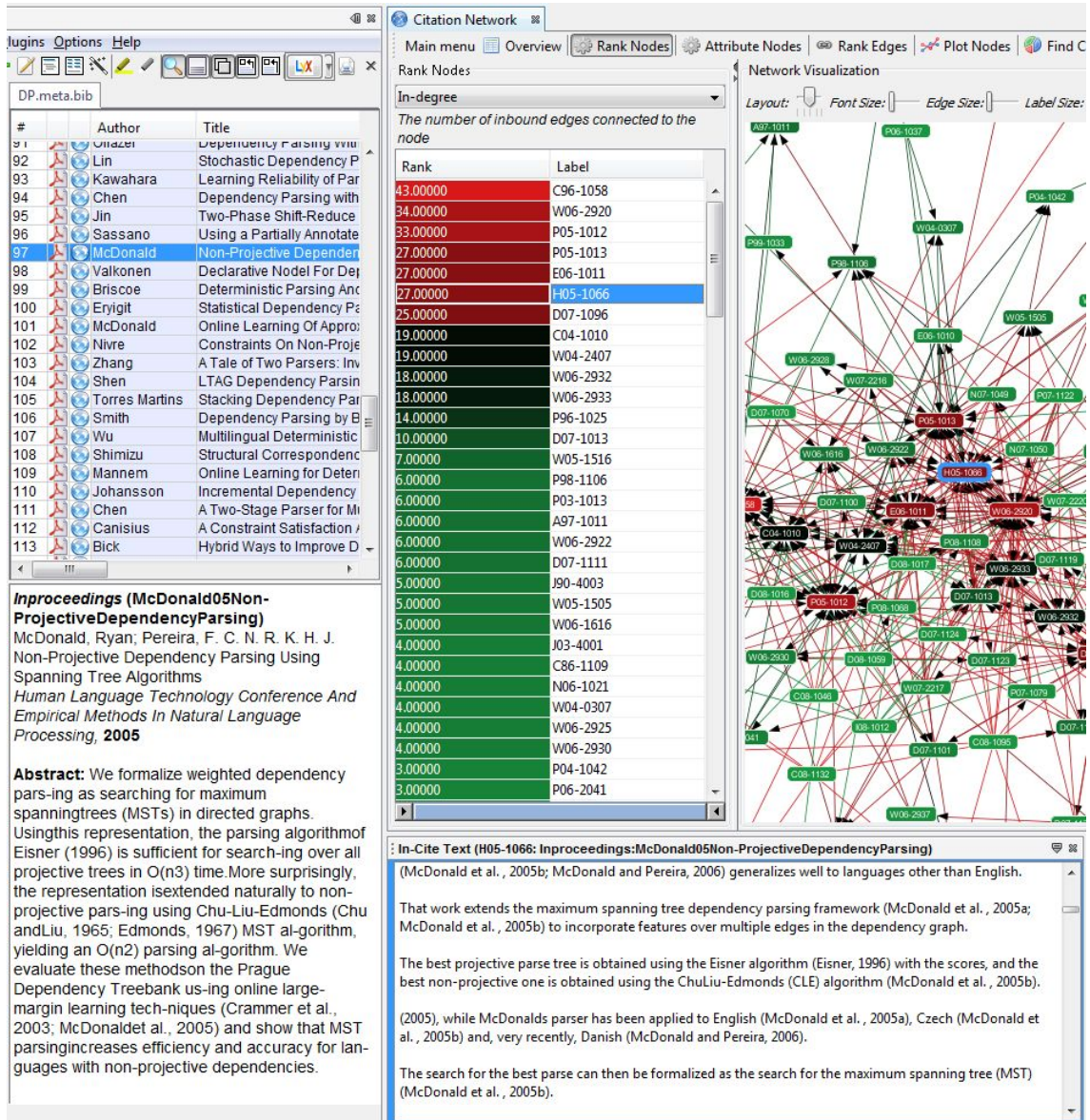


Figure 3.1: The In-Cite Text view (bottom right) shows several citations which make positive remarks about the paper H05-1066 by McDonald et al.

In order to identify important authors, P3 wanted to see rankings on a per-author basis to identify which authors had been cited most frequently. P3 described a method of using Find Communities and looking in each cluster for highly cited authors, but had trouble completing the task and chose not to finish it.

Initially, P2 only used the reference manager to view the list of authors and select ones P2 already knew were key. P2 tried to augment her approach by finding all papers of a certain author (Nivre) in the reference manager and selecting them to view their in-degree ranking in the network visualization. However, due to a bug in the software, the view was not coordinated, so her selection was not reflected in the node ranking which made it difficult to determine the ranking for each authors' papers. When identifying important papers, P2 wanted to rank papers both by in-degree and publication year in order to see which years had the most highly cited papers.

P4 had a similar approach to P2, but only used the reference manager to sort by author name and visually scan the list to select authors and papers that P4 already regarded as important. P4 did not use the network visualization or any other part of ASE to verify P4' notions of importance, indicating that P4 was confident in the selections and did not expect to make any additional discoveries.

Table 3.1 summarizes ASE's capabilities each participant used in the first pre-defined task.

For the second task, finding evidence to determine why a paper is important, P2 and P4 continued their pattern of relying on their prior knowledge. P4 did not use the system at all and described from memory why the selections were important.

Participant	Attribute Ranking	Find Communities	Plot Nodes	In-Cite Text	In-Cite Summary	Search	Sorting	Groups
P1	•			•				•
P2	•					•		•
P3	•							
P4							•	

Table 3.1: Table of capabilities in my first usability study used by each participant for the first predefined task (identifying important authors and papers).

P2 combined her existing knowledge with node ranking by corpus in-degree, and was able to disregard some well ranked papers because P2 felt that the papers were always cited out of obligation rather than because they are relevant. P2 ultimately chose a paper and justified her claim by filtering the node ranking to show only the highest corpus in-degree and then seeing in the visualization that this paper linked two of the most highly cited papers in this filtered subset (see Figure 3.2).

P3 used a similar approach to P2, and ranked nodes by in-degree. Then P3 selected a highly cited paper and justified her claim by pointing out in the visualization that it is cited by another highly cited paper.

P1 changed approaches several times while performing the task. P1 began by turning on Find Communities and viewing the In-Cite Summary of the largest community, but did not find anything that P1 thought was helpful. Then P1 opined

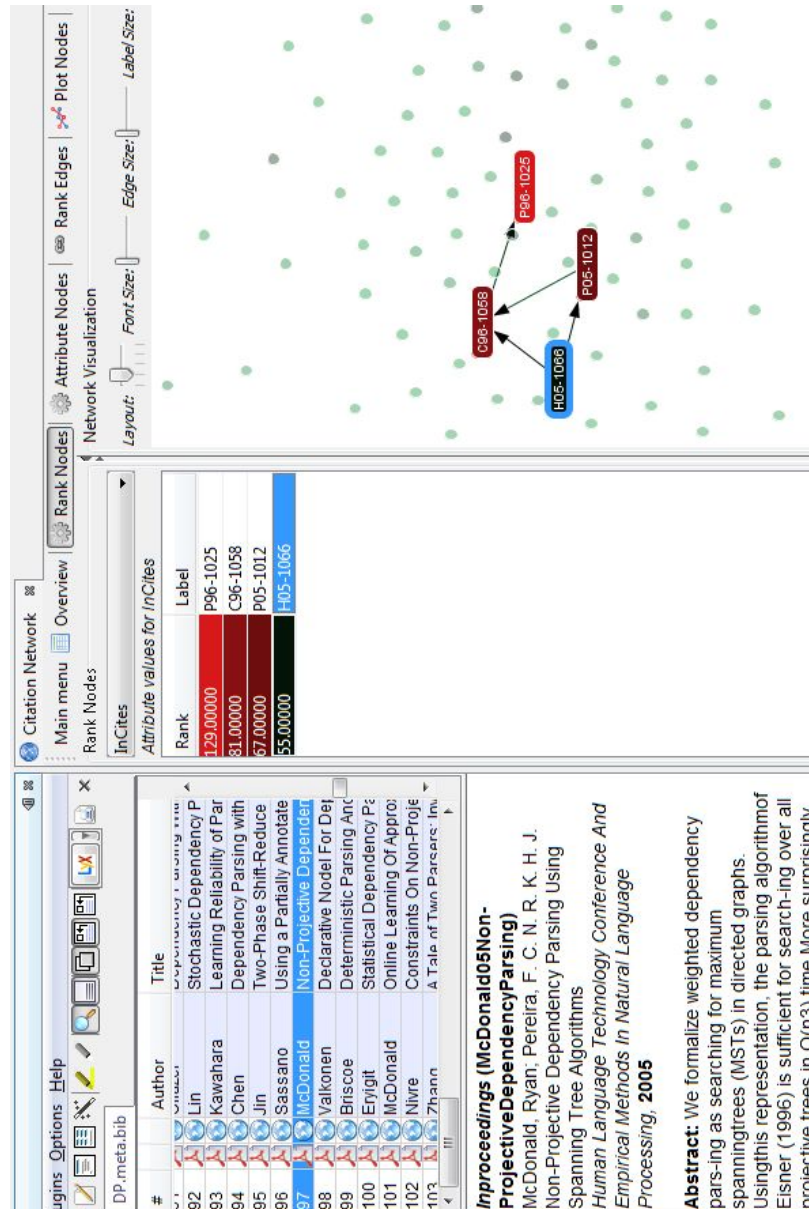


Figure 3.2: P2 chose H05-1066 as an important paper after ranking papers by corpus in-degree filtering to show the top most highly cited papers. This revealed that H05-1066 (shown highlighted in blue) is highly cited and cites two of the most highly cited papers. The paper’s details are visible on the left, including authors, title, and abstract.

Participant	Attribute Ranking	Find Communities	Plot Nodes	In-Cite Text	In-Cite Summary	Search	Sorting	Groups
P1		•						•
P2	•							
P3	•							
P4								

Table 3.2: Table of capabilities in my first usability study evaluation used by each participant for the second predefined task (collecting evidence).

that the most recent paper that gave an efficient algorithm for dependency parsing would be the most important paper. P1 wanted to find a list of all papers that cited a given paper and then sort the results by publication year, but P1 had difficulty with this because there is no automated way to do it. P1 then went back to Find Communities and began searching through the largest community by looking at the incoming citation sentences. P1 eventually chose C96-1058 and H05-1066 based on the sentences shown in the In-Cite Text view which talk about efficient algorithms, but P1 could not satisfy the original search for a most recent paper on efficient algorithms.

Table 3.2 summarizes ASE’s capabilities that each participant used during the second predefined task.

3.1.4.2 User-defined tasks

Users were asked what kinds of questions they would want to answer while exploring this dataset. Below is a summary of these questions, grouped by the participant who was interested in that question:

- **P2** Identify the historical foundations of the field.
- **P2** Identify the big breakthroughs in the field, and the state of the art.
- **P1,P4** Find relationships between communities and clusters of papers, such as papers that use the same research methods.
- **P1** Discover frequent co-authors.
- **P3** Find review papers.
- **P3** Search for specific aspects (e.g. models or applications).
- **P3** Identify different research methods used in the literature.
- **P3** Identify student-advisor relationships in papers.

The participants mostly gave unique lists of questions, except P1 and P4 both were interested in identifying relationships among papers in communities. However, some participants identified similar questions, such as discovering frequent co-authors and discovering student-advisor relationships.

To test whether users are able to answer their own questions using the system, the facilitator chose some of the tasks identified by each participant and asked that

participant, except for P1, to perform the task(s). P1 was an exception because the allotted time for the session had expired by this point.

P2's task was to try to search for papers that the participant would want to start reading to identify the state of the art in dependency parsing, since this was one of the tasks P2 had identified. P2 first tried ranking papers by publication year and trying to look at the In-Cite Text for the most recent papers, but they did not have In-Cite Text because they had no citations. P2 then continued this approach without using the Attribute Ranking by selecting papers in the network visualization and examining their publication year and In-Cite Text. P2 appeared to be selecting papers randomly. P2 then chose a paper from 2007 and commented that finding the state of the art can be difficult if it is too new because it will not yet be cited by many papers. P2 created a group in the reference manager to store her results, and examined the paper titles and abstracts in the reference manager for relevance, and eventually opened the full PDF version of candidate papers to scan them. In the end, P2 successfully identified papers that the participant wanted to investigate later.

P3 was assigned two additional tasks. First, P3 was asked to determine if the most important author was still active. To accomplish this, P3 wished to rank papers both by corpus in-degree and by publication year. This action is not directly supported in the system, so after a few attempts to do this, P3 changed her approach because the participant felt that recent publications might not be highly cited. P3 then sorted the references by author and looked for authors with many publications who also have a recent publication, allowing her to identify two authors.

P3’s second task was to find review papers. Her approach was to rank nodes by corpus out-degree because a review paper will cite lots of other papers. After looking at the top papers and not finding a review paper, P3 also used the search feature to search for the keywords “tutorial” and “survey” but did not find any matches. After the evaluation session I discussed the dataset with another graduate student in computational linguistics who works with dependency parsing, and he said it is likely that there is no survey paper on dependency parsing in this dataset.

P4’s task was to find any unifying themes in any of the communities. P4 used the Find Communities feature and noticed that most of the papers in the biggest community cited the same central papers. P4 proceeded to look at the titles of the papers in the community, and P4 then mentioned that it was hard to find a unifying theme by looking at the paper titles. Eventually P4 said “It doesn’t jump out at me why it clustered these things [this way].” At this point P4 had looked as much as possible P4 felt was possible, and gave up on the task.

Table 3.3 summarizes the capabilities in ASE that participants used for the user-defined tasks.

3.1.4.3 Additional tasks

Due to P4’s familiarity with dependency parsing, I gave this participant four additional tasks.

The first was to determine if there has been any interesting work using any of the most important papers. P4 began by selecting a paper, P05-1013, in the

Participant	Attribute Ranking	Find Communities	Plot Nodes	In-Cite Text	In-Cite Summary	Search	Sorting	Groups
P1								
P2	•			•				•
P3	•					•	•	
P4		•						

Table 3.3: Table of capabilities in my first usability study used by each participant for the user-defined tasks.

visualization and looking at recent citations: P4 would examine the In-Cite Text, click on the hyperlink, and then look at the publication at the top of the plain text (out-cite text). P4 expressed frustration that in order to open the PDF for any of these citing papers that focus had to be removed from P05-1013. After using the procedure for a few papers and opening a few PDFs to see if authors were citing P05-1013 in a way indicating something interesting, P4 isolated D07-1123 as an interesting advancement of P05-1013.

P4’s second task was to use the system to support P4’s previous claims about the most important authors from the first pre-defined task. P4 first ranked the nodes by document set in-degree and corpus in-degree, and mostly P4 used a single paper from each author as a representative for the author’s body of work; however, P4 wanted to group papers by author and look at rankings just for a specific author’s

papers. P4 only looked at the document set in-degree and corpus in-degree rankings to decide that P4's choice of Schank is not supported; P4 felt the paper is classic, but it has a low document set in-degree and corpus in-degree ranking. For another author, P4 felt that the paper was actually ranked higher than it should have been, but nonetheless P4's previous selection was justified. For the other three papers—P05-1013, C96-1058, and P96-1025—P4 felt that the attribute rankings confirmed P4's previous choices, and that overall the previous choice of authors was confirmed.

For the third task, P4 needed to use the system to support P4's previous claims about important papers. This time P4 used only attribute ranking by corpus in-degree, and felt that eight of the eleven choices were confirmed by the attribute rankings.

P4's final task was to summarize the literature using the system. P4 first commented that ranking the nodes and finding the top paper could be a good place to begin. However, P4 decided to use the Find Communities feature and look at the In-Cite Summary for the largest community. P4 noticed several things in the automatically generated summary that were factually incorrect, and decided that the summary was not helpful. At this point the session time had expired and the participant did not have time to complete the task.

3.1.5 Discussion

Overall, participants were able to quickly grasp the basics of the reference manager and network visualization components of ASE. Some participants immediately

began using somewhat more advanced features, such as P1 who began using searches almost from the beginning. However, participants did not, in general, appear to improve much from the beginning of the session through the end; the functionality they used at the beginning of the sessions tended to be the same functionality they used at the end.

By examining the functionality used per task shown in Tables 3.1, 3.2, and 3.3 it is clear that by far the most used feature was attribute ranking. Participants almost ubiquitously used it to determine author and paper importance. Even when tasks did not directly involve finding an important paper, participants still relied heavily on attribute ranking, such as when searching for review papers. This may be partly because review papers often focus on the most salient aspects of the corpus and users may believe those papers will be cited most often. On the other hand, participants relied almost exclusively on corpus-level and document set-level degree ranking, and occasionally mentioned that they were not sure how to interpret some of the other metrics such as betweenness centrality.

As Hearst discusses [24], there are theories describing how information seekers make navigation decisions based on information “scent.” One example of this is when P2 used papers’ In-Cite Text to guess which papers might describe state-of-the-art research, and P2 then examined promising-looking papers in more detail by opening the PDF full-text versions. It is possible that participants were so eager to use the Attribute Ranking feature because it provided very powerful quantitative clues to begin their navigation process and it offered information seekers a simple way to direct their search during the Exploration stage of Kuhlthau’s [33] description

of information seeking behavior.

It could also be possible that Attribute Ranking offered a fast way for participants to filter out potentially unimportant papers and allow the participants to focus their exploration on portions of the literature which would be considered most important according to some Attribute Ranking metric. This is similar to the Area Scan stratagem described by Bates [4] where users browse in a general area after identifying an area of interest.

In general, participants did not seem to use multiple features in the system to drill down and find specific answers to questions; they tended to use a single system capability, such as Attribute Ranking, to support their claims. For example, when participants selected an important paper most of the participants only used Attribute Ranking by document set in-degree to justify their claim. It is possible that participants found Attribute Ranking to be very accessible because it is simple to use and easy to interpret. However, participants did sometimes use other capabilities, so they did appear to understand ASE's many capabilities. However, it is possible that the participants were not familiar enough with ASE to use multiple capabilities in conjunction with one another.

Although the Out-Cite Text view is potentially useful, it was not without its problems. For example, participants had difficulty understanding what the Out-Cite Text is and which paper it comes from. When P2 asked for clarification, the participant was shown a node-linked diagram which seemed to improve her understanding, but P2 still made a few comments which indicated the participant did not fully understand the Out-Cite Text view. Furthermore, there were issues

with changing focus in the network diagram to a different paper when using the Out-Cite Text view and not being able to easily return focus to the previously selected paper. When a participant wanted to investigate further a different paper linked in the Out-Cite Text view, P2 had to change focus onto that paper, but there was no convenient way to return focus to the previously selected paper. If the view had been filtered to show only the selected node and all of its citing papers it might have made the task easier.

We also see the importance of multiple coordinated views in the example when P2 tried to select papers in the reference manager and see their ranking in the Attribute Ranking view. In that version of ASE, the view was not coordinated and the paper did not become selected in the Attribute Ranking view. This bug constrained the participant's use of ASE and limited her analysis.

Many of the analyses performed by the participants may not have been possible, or would have been much more difficult, if the system did not allow users to use a custom database of documents. Even so, participants frequently expressed interest in analyzing only a subset of the papers in the network. Frequently participants wanted to do this to show a given node and all of its neighbors. This underscores the importance of allowing users to select a corpus of documents and drill down to perform fine-grained analysis.

The In-Cite Summary is a feature that participants wanted to use. In the version used during the evaluation sessions, the output quality of the summarization algorithm was limited. This is an avenue of on-going research, and I received new versions of the summaries which were used during the next set of usability evaluation

sessions.

Three of the four participants said that they liked the system and thought it was useful, in particular for searching for important papers. P4 was skeptical about the system’s utility, however: P4 felt that the interface is too “busy”, and that the system might not be useful unless P4 had a goal in mind and already knew a few papers as a starting point. P2 also commented that the interface is complicated, but the participant had an overall positive response, enjoyed using the network visualization, and was very enthusiastic about the system.

One of the limitations of the user study is the size of the corpus, which only contains 147 papers. This demonstrates a particular use case for a narrow research domain, but may not be indicative of exploring a broader research domain such as network visualization.

A further limitation is the number of participants who took part in the study. This prompted additional evaluation sessions which are described in the next sections of this chapter.

3.2 Other Evaluations

My first set of usability studies revealed that ASE is well suited for some tasks, such as searching for facts like the top ten most highly cited papers. It was less clear whether ASE is useful for long-term exploration projects seeking deeper insights. In an attempt to answer this question, a participant with experience in the grant approval process volunteered to work with ASE over a period of a few weeks

to analyze the dependency parsing data set used in the previous usability study. During the first session, the participant appeared disinterested and unmotivated in the data analyses. The participant was unresponsive to requests for follow-up data analysis sessions and in the end did not work with ASE beyond the initial session. This anecdote underscores the importance of having real people analyze real data, as emphasized by Shneiderman and Plaisant [59].

An information retrieval expert also volunteered to use ASE for 45 minutes and then to offer feedback. The participant began by watching the three training videos, but found them to be too long and it was too difficult to remember all of the functionalities and how to access them. Following the training phase, the participant used ASE briefly and mentioned several points for improvement:

- The force-directed layout algorithm continuously animates, which causes a distracting visual effect.
- ASE does not allow users to choose how a certain functionality is performed. For example, ASE has a pre-defined way of performing clustering (based on edges in the citation network) which may not be natural to users or could be inappropriate for certain tasks (e.g., clustering to identify research institutions). Another example is the In-Cite Summary of communities, where sometimes users may prefer to see a different type of community summary instead.
- The interface is complex and could provide a steep learning curve.

However, the participant felt that ASE appears to be useful for identifying salient

features of the scientific literature network, such as important papers and authors, and groups of documents that comprise the core papers in the field.

3.3 Improvements Made to Action Science Explorer

Based on the results of the previous evaluations, I implemented three improvements to ASE. One was a temporary improvement to the In-Cite Summary system which was later replaced by a new version of the In-Cite Summary text created with a new version of Trimmer.

Multiple coordinated views. One of the usability problems that arose from my first usability study evaluations was that the multiple views were not fully coordinated: When users selected a paper in the reference manager it would become selected in the citation network visualization, but not in the Attribute Ranking view. I fixed the system so that selections were fully coordinated across all views.

Multiple-selection across views. The Attribute Ranking view was limited in that only one paper could be selected at a time. Although this was not a significant usability concern during my first usability study evaluations, this problem came up during several demonstrations of ASE. Because users expressed a desire to be able to select a subset of the corpus for analysis, I enhanced the Attribute Ranking view to allow multiple selections that would be correctly coordinated across all views. This capability is illustrated in Figure 3.3. I also set out to add the capability for users to select a group of documents in the reference manager and then filter the visualizations and text views to show only the selected papers. However, it became

clear that such a feature would require a significant re-engineering of both JabRef and SocialAction; neither system is designed to be used in conjunction with the other which makes complex interactions difficult. To illustrate this point, refer to Figure 3.3 which shows that JabRef has at least three different ways for a document to be selected: (1) a document can be highlighted in blue indicating that its details are being displayed in the bibliography entry preview, (2) a document can be selected in white and moved to the top of the document list while the others are grayed out, or (3) a document can be shown in the bibliography list because it matches a search query while other documents which do not match the query are hidden from view. This makes it difficult to coordinate complex interactions between JabRef and SocialAction, and is one of the limitations of this style of software architecture.

In-Cite Summaries. As discussed in Section 2.4, the Find Communities capability incorporates a slider to adjust community size. In the earlier version of ASE, summaries only existed for every tenth increment of the slider, forcing users to carefully adjust the slider or else summaries would not be displayed when users clicked on a community. I explored a work-around in the software to ensure that a summary was always shown, but eventually I obtained a complete set of citation summaries which solved the problem and obviated my software work-around. These summaries were produced by Michael Whidby, David Zajic, and Bonnie Dorr from the University of Maryland Computational Linguistics and Information Processing lab. This new complete set of summaries also used an updated version of Trimmer, the summarization software. One goal of the updated summaries was to improve readability and syntactic content of the summaries.

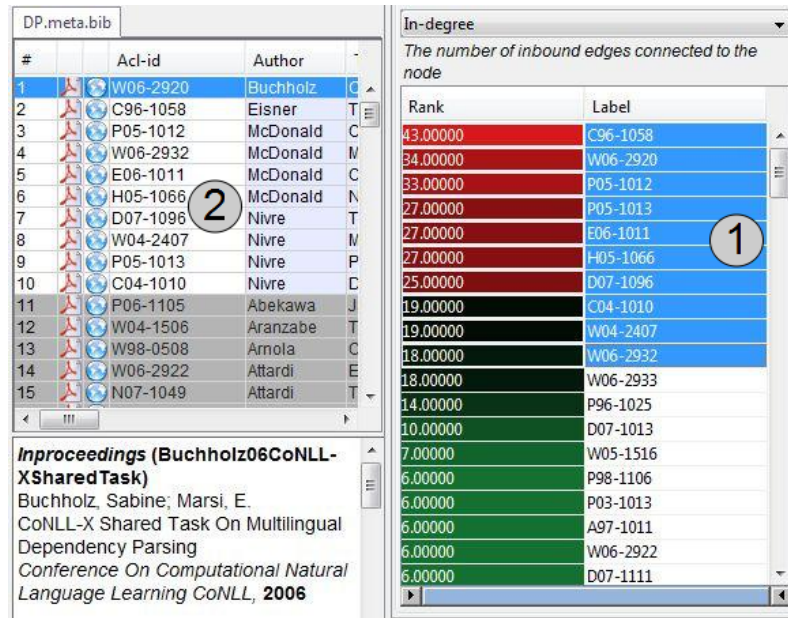


Figure 3.3: The Attribute Ranking view in ASE was fixed to allow multiple selections that would correctly coordinate selection across all views. Here, the top ten most highly cited papers in the dependency parsing set were selected in the Attribute Ranking (1), which then move to the top of the reference manager and are shown in white and blue (2), allowing us to see that McDonald and Nivre authored eight of the top ten papers.

To incorporate these changes, I first had to learn two things: (1) ASE's software architecture, which is based on the NetBeans version of the Rich Client Platform, a software framework for writing modular applications, and (2) the system design of SocialAction, which I also had to modify to implement modifications in ASE. Next, I implemented the new functionality in ASE, which resulted in 74 lines of new and modified Java code. I tested the modifications through a set of GUI test cases which tested selecting papers in the three views (the network visualization, the Attribute Ranking view, and the reference manager). The test cases were created using methodology based on the category-partition method. In total, these modifications to ASE required close to one month of time.

These updates to both the functionality of ASE and the content of the summaries prompted a second set of usability study evaluations.

3.4 Second Usability Study

After addressing the usability problems discussed in Section 3.3, four participants took part in an additional evaluation session to evaluate the modifications to ASE. Two of the participants from my first usability study participated in these new evaluations, which provided additional insight into usage patterns of users with more experience. This new evaluation took place about six months after my first one.

3.4.1 Participants

My second set of evaluation sessions were performed with four participants: two new participants (referred to as P5 and P6), and two participants from my first evaluation sessions (P1 and P3). All of the participants were current Computer Science Ph.D. students. The ages ranged from 23 to 29 years old, and one of the participants was female.

Two of the participants are currently working in computational linguistics. P5 and P6 described themselves as understanding the concept of dependency parsing, but were not familiar with the literature on dependency parsing. P1 and P3 said that they are familiar with dependency parsing and have read a small number of papers on the topic.

3.4.2 Experimental Design

The experimental design was identical to the usability study evaluation in Section 3.1.2.

3.4.3 Procedure

The procedure for the evaluation was identical to the usability study in Section 3.1.3, except that I recorded a video capture of the participants' monitors plus audio during the evaluation sessions.

3.4.4 Results

As with the my first usability study, there were two sets of tasks during the evaluation sessions: predefined tasks and user-defined tasks. For all participants the predefined tasks were performed first, followed by the user-defined tasks. See Section 3.1.3 for a description of the predefined tasks.

3.4.4.1 Predefined tasks

During the first predefined task (see Section 3.1.3), P1 began by creating authors and papers groups, and then ranking papers by document set in-degree. Next P1 explored the pre-computed groups of authors (shown in Figure 3.4), scanned for authors P1 already knew about and added them to the group of important authors. P1 then performed a search for particular authors and added them to the group of authors. To add important papers, P1 ranked the papers by document set in-degree, filtered out all but the five most highly cited papers, and added them to the important papers group. P1 continued by using Plot Nodes by corpus in-degree and publication year, and filtered out papers with low citation counts (see Figure 3.5). P1 chose a recent paper which is highly cited. P1 identified this as an important paper, but it was among the top five papers P1 already added to the important papers group.

P3 began by creating groups in the reference manager for important papers and important authors. P3 used plot nodes to compare relationships between corpus in-degree and document set in-degree, and then selected papers which where had

The screenshot displays a software interface for managing a bibliography. On the left, a search bar is active, and a list of authors is shown. A group of authors, including 'Eisner', is highlighted and numbered '1'. This group is selected, and the corresponding papers are moved to the top of the bibliography list, numbered '2'. The main window shows a table of papers with columns for Rank, Label, and Title. The top of the list includes 'Three New Probabilistic Models For Dependency Parsing: An Exploration' by Eisner, J. M. On the right, a network visualization shows nodes representing authors and edges representing relationships. A tooltip for node 'C96-1058' provides details about the paper and its author, Eisner (1996).

Rank	Label	Title	Author
13,00000	C96-1058	Three New Probabilistic Models For Dependency Parsing: An Exploration	Eisner, J. M.
34,00000	W06-2920	Computational Linguistics, 1996	Eisner, J. M.
33,00000	P05-1012	Computational Linguistics, 2005	Yamada and Matsumoto, 2003; Nivre and Scholz, 2004
27,00000	P05-1013	Computational Linguistics, 2005	Yamada and Matsumoto, 2003; Nivre and Scholz, 2004
27,00000	E06-1011	Computational Linguistics, 2006	Collins (1996)
27,00000	H05-1066	Computational Linguistics, 2005	Yamada and Matsumoto, 2003; Nivre and Scholz, 2004
25,00000	D07-1096	Computational Linguistics, 2007	Esner, 1996a
19,00000	C04-1010	Computational Linguistics, 2004	Hodermaier and Steedman (2002b)
19,00000	W04-2407	Computational Linguistics, 2004	Esner, 1996
18,00000	W06-2932	Computational Linguistics, 2006	Esner, 1996
14,00000	P96-1025	Computational Linguistics, 1996	Esner, 1996
10,00000	D07-1013	Computational Linguistics, 2007	Esner, 1996
7,00000	W05-1516	Computational Linguistics, 2005	Esner, 1996
6,00000	P98-1106	Computational Linguistics, 1998	Esner, 1996
6,00000	P03-1013	Computational Linguistics, 2003	Esner, 1996
6,00000	A97-1011	Computational Linguistics, 1997	Esner, 1996
6,00000	W06-2922	Computational Linguistics, 2006	Esner, 1996
6,00000	D07-1111	Computational Linguistics, 2007	Esner, 1996
5,00000	P90-4003	Computational Linguistics, 1990	Esner, 1996
5,00000	W05-1505	Computational Linguistics, 2005	Esner, 1996
5,00000	W06-1616	Computational Linguistics, 2006	Esner, 1996
4,00000	P03-4001	Computational Linguistics, 2003	Esner, 1996

Figure 3.4: The pre-computed groups of authors (1) on the left can be selected, which then moves the papers in that group (2) to the top of the list of bibliography entries.

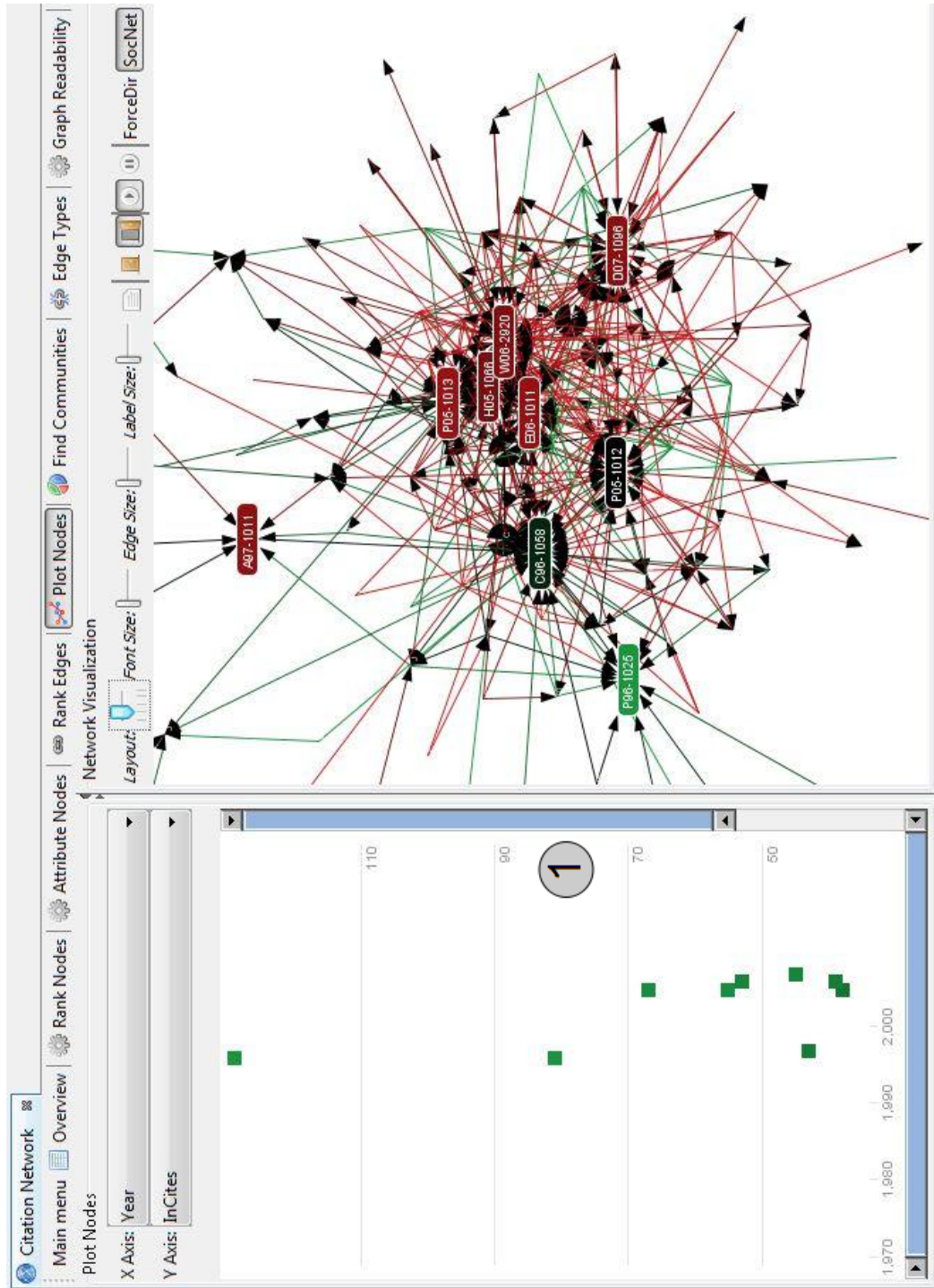


Figure 3.5: The Plot Nodes view shows a scatter plot of the papers in the citation network with one variable on the X axis (e.g., publication year) and a second variable on the Y axis (e.g., corpus in-degree). Here, the vertical double-ended slider (1) has been adjusted to filter out papers with a low corpus in-degree.

high values for both metrics and P3 added them to the important papers group. For selecting important authors, P3 wanted to compute the number of citations that authors have for all of their papers. In lieu of that, P3 opted to choose the authors who ranked in the top 10 as measured by the corpus in-degree ranking.

P5's approach was to use Attribute Ranking to rank the documents by corpus in-degree which P5 felt would give a "larger sample" of citations and give a more accurate picture of importance than document set in-degree. P5 noted that Collins had the most highly cited paper. To find all of Collins' papers, P5 initially used the search functionality but then chose to sort by author name in order to find all papers written by Collins. P5 added the search results to the two groups for the most important authors and most important papers. P5 went through a process of choosing the next most highly cited paper, adding it to the group for important authors, and then putting other highly cited papers by that author into the important papers group. One problem that P5 encountered was that the reference manager will only sort by first author, so sorting by author name may not group together all papers from a particular author. Nonetheless, P5 preferred sorting by author name instead of searching for an author because it was easier.

P6 had a simple approach of creating two groups for important authors and papers, and then using Attribute Ranking to see the papers with the highest corpus in-degree. P5 chose the top ten papers for the important papers group. P5 wanted to see the total number of citations for each author, but since this was not directly supported by ASE P5 instead sorted the papers by author name and ranked them by corpus in-degree to determine which authors had the most highly cited papers.

Participant	Attribute Ranking	Find Communities	Plot Nodes	In-Cite Text	In-Cite Summary	Search	Sorting	Groups
P1	•		•			•		•
P3	•		•					•
P5	•					•	•	•
P6	•						•	•

Table 3.4: Table of capabilities in my second usability study used by each participant for the first predefined task (identifying important authors and papers).

P5 then chose the authors of the top ten papers as the important authors.

Table 3.4 summarizes ASE’s capabilities that the four participants used in the first predefined task.

For the second predefined task, P1 explored the use of Find Communities to search for evidence about important papers, but in the end P1 did not collect evidence from it. Instead, P1 used Attribute Ranking by corpus in-degree to identify candidate papers, viewed their In-Cite Text to see if other authors were saying good things about the paper, read their abstracts, and then placed selected papers into a group in the reference manager for candidate papers. P1 then used Plot Nodes by year and corpus in-degree to search for recent highly cited papers, and continued to read the In-Cite Text and abstracts for candidate papers. Eventually P1 selected McDonald’s 2005 paper because of its research contribution to dependency parsing,

it was highly cited, published recently, and other authors said good things about it in the In-Cite Text. This was the same paper P1 selected from my first evaluation session.

P3 used a similar approach as P6, but used Plot Nodes to create a scatter plot with the two variables document set in-degree and corpus in-degree. P3 explained that a good paper should be highly cited both by papers within the dependency parsing subdomain as well as the whole domain of computational linguistics.

P5 chose the paper with the highest document set in-degree. P5 examined it in the reference manager and justified this claim by saying the paper was slightly older and thus should be a more canonical paper rather than a trendy paper which may quickly fade.

P6 chose to use only Attribute Ranking by corpus in-degree to determine which paper was the most important; P6 felt that was sufficient for determining importance.

Table 3.5 summarizes ASE's capabilities that the four participants used in the second predefined task.

3.4.4.2 User-defined tasks

Users were asked what kinds of questions they would want to answer while exploring this dataset. Below is a summary of these questions, grouped by the participant who was interested in that question:

- **P1** Find major topics being investigated by this collection.

Participant	Attribute Ranking	Find Communities	Plot Nodes	In-Cite Text	In-Cite Summary	Search	Sorting	Groups
P1	•	•	•				•	•
P3			•					
P5	•							
P6	•							

Table 3.5: Table of capabilities in my second usability study used by each participant for the second predefined task (collecting evidence).

- **P1** Identify major researchers by overall importance to other researchers.
- **P3** Search for review papers.
- **P3** Find papers that are easy to read.
- **P3** Gather opinions of papers as a basis for paper importance.
- **P3** Determine who founded the area.
- **P3** Determine if a given domain is an author’s main research area, or if it is more of a side project for the author.
- **P3** Search for authors who collaborate together often and analyze collaboration patterns.

- **P3** Choose a paper and view a list of all papers that cite it and were cited by it.
- **P3** Identify the state of the art.
- **P5** Identify old papers with a high number of citations.
- **P5** Follow an important paper's citations backwards or forwards in time to examine the evolution of the field.
- **P1,P5** Examine the automatic clusters to search for subtopics.
- **P6** Determine if there is a correlation between different paper metrics, such as corpus in-degree and HITS Authority.
- **P6** Find the author with the most papers in the document set.
- **P6** Examine how a paper was cited over time, both through its In-Cite Text and the number of citations over time.

Four of these questions are equivalent to questions identified during my first usability study. Three of these overlapping questions come from P3.

As with my first usability study, the facilitator chose some of the tasks identified by each participant and asked that participant to perform the tasks.

P5's first user-defined task was to start with Eisner's 1996 paper and work forwards in time through the citations. P5 began by looking at the In-Cite Text for Eisner's paper and wanted to be able to filter the In-Cite Text by categories such as the paper or author the text is coming from, but ASE does not support any filtering

or sorting for the In-Cite Text.c P5 followed a few In-Cite Text links, but felt that P5 was not getting a good feel for how the field evolved since 1996.

P5's second user-defined task was to examine the automatically detected communities found by Find Communities and determine subtopics in the papers. P5 chose a large community and looked at the In-Cite Summary. Then P5 looked at the community's papers in the reference manager to look at the paper titles to search for a common theme. (See Figure 3.6 for a screenshot of these actions.) Overall, P5 did not feel like it was possible to determine the subtopics in the cluster.

For P5's final user-defined task, P5 was asked to identify old papers that have a high citation count. P5 was easily able to do this using the Plot Nodes functionality, and quickly spotted several papers that P5 felt were both older and highly cited.

P1's first task was to identify major topics in the document set. P1's process was to enable Find Communities, go from community to community, and view the In-Cite Summary. P1 commented that the summaries didn't provide a clear theme for the larger communities, but they did give a coherent summary of small communities' themes. Overall, P1 was satisfied with the linguistic structure of the summaries. P1 also wanted to find communities based on keywords, which P1 thought might be more helpful for identifying themes in the document set, but ASE does not support sorting or filtering communities by keyword content.

P1's second task was to identify the authors who are important globally, even to those not doing dependency parsing research. To accomplish this, P1 wanted to see co-citation and collaboration networks of the authors in the network. Since that is not supported, the participant sorted papers by author name and eventually

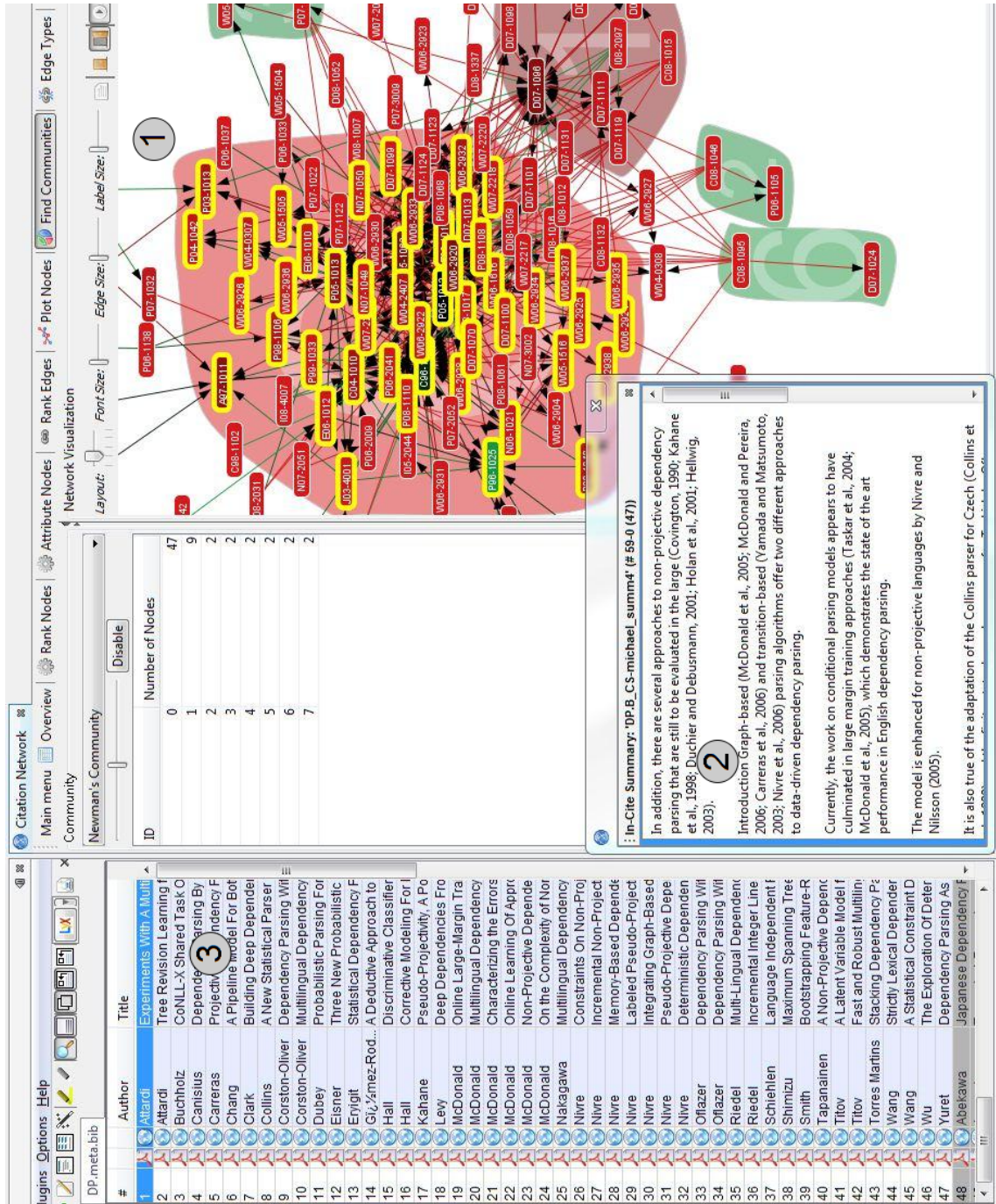


Figure 3.6: When the Find Communities function is enabled, users can select a cluster of papers (1) which become highlighted in yellow, the In-Cite Summary (2) is shown, and the documents in the community are moved to the top of the reference manager (3).

chose McDonald and Nivre. P1's basis for choosing Nivre was that P1 saw that Nivre authored a large number of papers.

P1 also wanted to explore the document set in an undirected search for insights. (Free exploration was not a task or question P1 gave in the list of questions for literature exploration, so this portion of the study is not included in Table 3.6.) P1 enabled Find Communities and used Attribute Ranking to sort papers according to the community to which they were assigned. P1 then sorted these papers in the reference manager by publication year and then by publication venue. P1 found that a majority of papers in the two largest clusters come from the 2006 CoNLL conference and the 2007 EMNLP-CoNLL conference; in fact, P1 found that these papers accounted for roughly 23% of all the papers in the document set (Figure 3.7 shows this insight). P1 commented that this explains the explosion of papers on dependency parsing in 2006 and 2007. Using Google, P1 searched for the conference websites from those two years and determined that the shared task in both years was on dependency parsing.

P6's first user-defined task used plot nodes to compare HITS Authority and corpus in-degree. P6 saw a correlation, and noticed an outlier (shown in Figure 3.8) where the paper with the highest corpus in-degree but did not follow the trend (a monotone increase in HITS Authority with corpus in-degree) because it had a relatively low HITS Authority.

The second user-defined task for P6 was to look for the author with the highest number of publications. But P6 quickly decided ASE doesn't support this task in any scalable way that could work for larger datasets, and the participant opted to

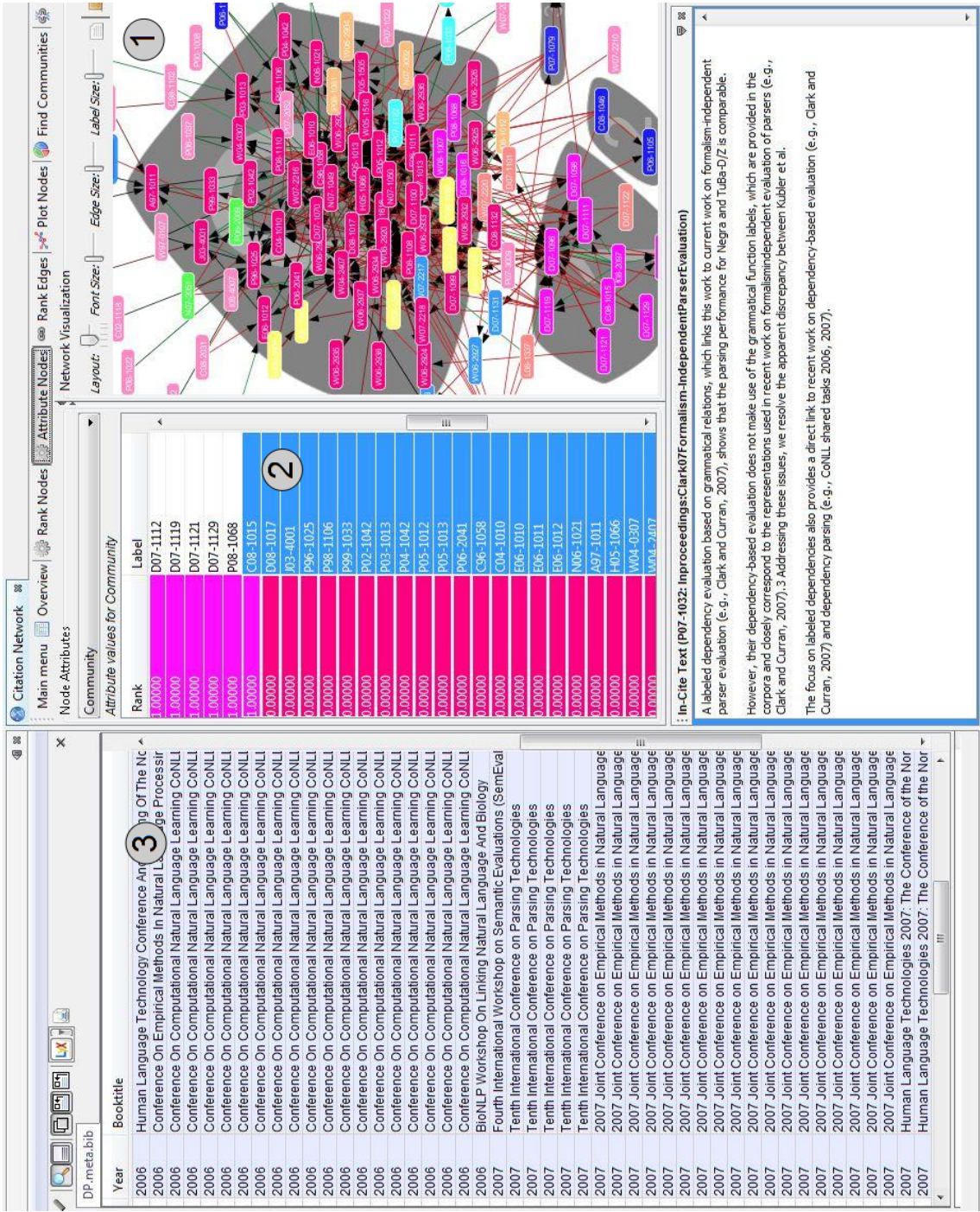


Figure 3.7: While exploring the documents in the large cluster (1) and viewing them in the attribute ranking (2) and reference manager (3), P1 was able to discover that a large number of papers in this dataset were published in two conferences from 2006 and 2007 (3).

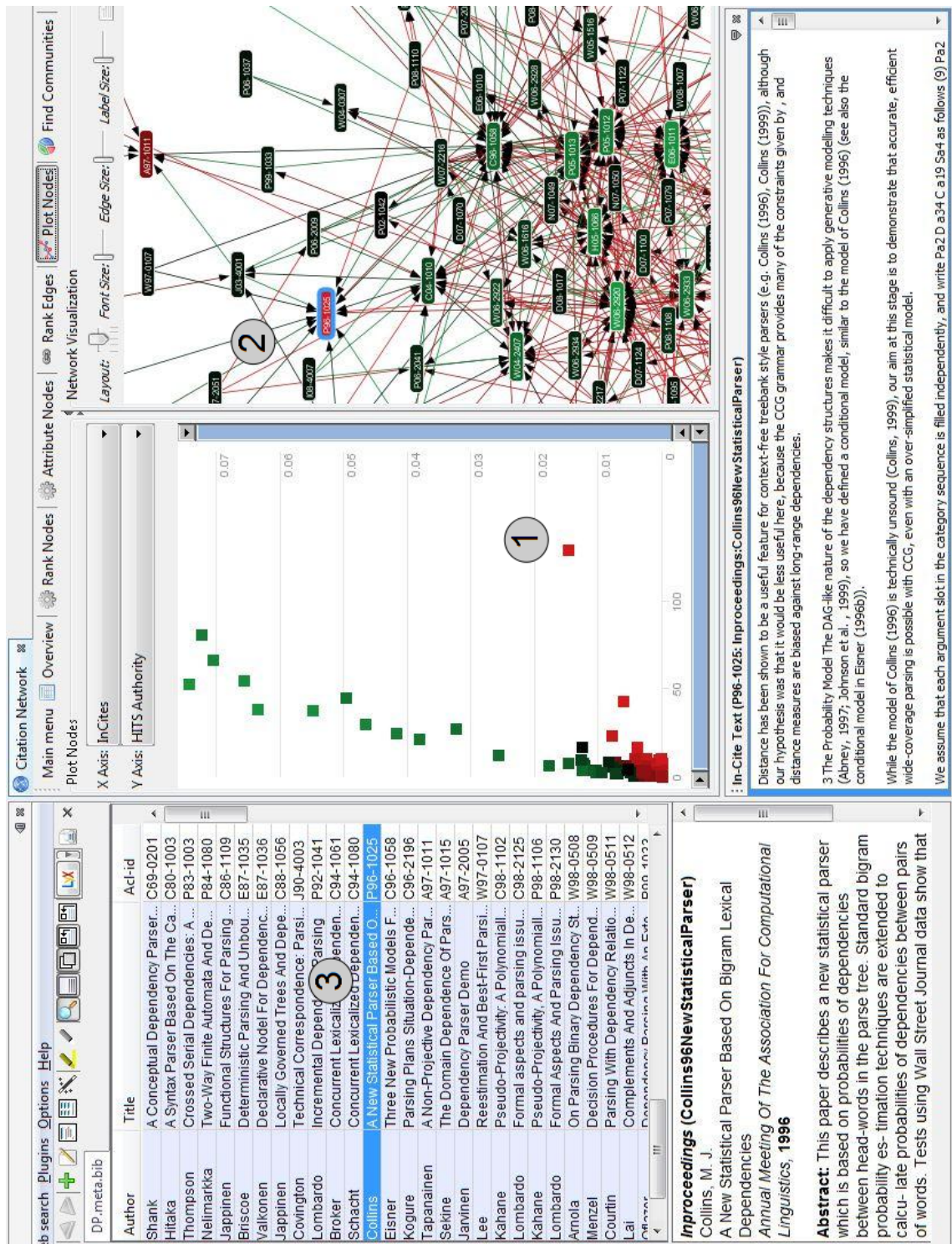


Figure 3.8: A scatter plot of corpus in-degree and HITS Authority reveals an outlier (1) (also highlighted in the citation network (2)) which is shown to be a paper by Collins (3).

stop the task.

P6's third task was to look at the most highly cited paper and how the citations of that paper changed over time. P6 thought that the system had some capability to do that with the In-Cite Text view, but that it was not supported well enough because it was not clear the In-Cite Text was sorted and the participant could not customize the sorting.

P3 performed one additional task, which was to examine Jason Eisner's collaboration patterns; specifically, to see if Jason Eisner made the transition from being a first author as a graduate student to no longer being the first author as a researcher who is advising graduate students. P3 did this by using the search capability to search for "Eisner" and then look at the author field in the query results. P3 felt that this was true, but there were not a lot of papers written by Eisner which made it hard to see.

Table 3.6 summarizes the capabilities in ASE that participants used during the user-defined tasks.

At the end of their two-hour sessions, both P1 and P3 asked if they could have a copy of ASE to use in their own research, indicating that they think ASE offers a package of useful capabilities not available in other systems. P3 gave more details and expressed interest in the ability to rank by citation count, because this is what is most important to P3 when exploring a completely new topic. P6 also said "I would love to try [ASE] for my own research." P5 said that ASE "seems pretty handy" and both P5 and P6 thought it would be useful for following citations across several papers. P5 also thought it would be easier to "keep things straight"

Participant	Attribute Ranking	Find Communities	Plot Nodes	In-Cite Text	In-Cite Summary	Search	Sorting	Groups
P1	•	•			•		•	
P3						•		
P5		•	•	•	•			
P6			•	•				

Table 3.6: Table of capabilities in my second usability study used by each participant for the user-defined tasks.

than it would when using a system like Google Scholar.

With respect to the In-Cite Text and In-Cite Summaries, P3 thought that citations would give a partial idea of what other authors think of a particular paper, but P3 felt that citations can be negative and downplay results of other authors in order to emphasize the need for their own paper. P6 also commented on the In-Cite Summary: P6 wanted to generate a summary for an arbitrary group of papers.

P6 said that it would be good to stop the constant animation rather than animating and having it flicker all the time.

3.4.5 Discussion

As with my first usability study, the new participants were able to quickly learn how to use ASE.

The general approach to the two predefined tasks was the same in both my first and second set of usability studies, although both P1 and P3 began using Plot Nodes in my second usability study when they had not used it during the my first usability study.

The participants had some problems where they wanted to perform an action which was not supported by ASE. Some examples include viewing a co-citation or collaboration network instead of a citation network, and choosing a different type of community summary such as topic modeling instead of citation summaries.

By comparing Tables 3.4 and 3.5 to Tables 3.1 and 3.2 we see that participants in my second usability study used more capabilities of ASE per task than the participants during my first usability study. Attribute ranking continued to be very highly used, but now P1 and P3 augmented it by using Plot Nodes which allowed the participants to rank by two variables simultaneously. This is interesting because none of the participants used the Plot Nodes capability during my first usability study. Participants mostly focused on degree ranking metrics, just as they did in my first usability study.

There is some overlap between the two usability studies in questions the users said they wanted to ask when exploring literature. Some of this overlap comes from P3, who listed several identical questions both in my first and second usability studies.

Participants used the new multi-document selection capability several times during my second usability study. P3 and P5 in particular used it frequently, possibly because it gave the ability to select an arbitrary set of papers in one view and

then perform some types of analysis in the other views. Nonetheless, P6 wanted this behavior to be more robust and be able to generate summaries of arbitrary groups of papers, too. This emphasizes the point that users want to be able to filter documents at many levels and to drill down onto a smaller document set to perform additional analysis.

Participants tried to use clustering to complete several tasks. This indicates that automatic clustering may be a useful capability for literature exploration. However, clustering was limited by two things: (1) the document set only contained 147 papers, and (2) automatically detected communities are only determined by how papers cite each other. The limited document set reduces the possibilities of clustering (fewer possible clusters). By only computing clusters based on citations, users lack control over choosing a clustering technique that is useful for their particular tasks and process models. Consequently, literature exploration system developers should consider adding automatic clustering functionalities, and implement a flexible clustering technique that allows users to cluster documents in the most appropriate way for a given task.

P1 was also able to make a powerful insight into the document set by discovering that a large number of the papers come from two conferences whose themes were on dependency parsing. This is an insight no one else had mentioned, even though more than a dozen other people have examined this specific dataset. P1 was able to make this insight even after using ASE for less than four hours. This emphasizes that insights can occur at any time, and it is necessary to have several participants use the system under evaluation for a long period of time.

Even though the usability studies have a low number of participants, P1’s insight combined with P1 and P3’s increase in capability usage give us evidence that users expand their analysis skills over time. Furthermore, ASE does support users in developing both simple insights as well as deeper insights into scientific literature.

We also see that ASE’s interface can be distracting. Both the information retrieval expert and P6 commented that the continuous network animation is distracting, which echoes P4’s comment that the interface is too busy.

3.5 Summary of Evaluations

Between my first usability with four participants and my second usability study with two new and two previous participants, the evaluations reveal several strengths and weaknesses of ASE. Furthermore, they found that ASE offers novel and useful features to information seekers. In particular, Attribute Ranking was very popular and provided users with a simple method to begin their searches, and it became clear that users wish to specify, at a fine-grained level, which documents to explore and analyze. Although corpus-refinement is present in Scatter/Gather, it does not appear to be a feature supported in consumer bibliography literature exploration systems.

Repeating the evaluations with two participants provided insight into the trajectory of use for maturing users. Participants who used ASE for a second time used more system features and were able to make deeper insights into the literature

network.

In the two usability studies, participants gave a set of questions they wish to ask when exploring scientific literature networks. There is some overlap in the questions given in my first and second usability studies, which indicates that users may find some questions more important than others. The questions have some common themes:

- Identify the foundations, breakthroughs, state-of-the-art, and evolution of a field.
- Find collaborators and relationships between communities.
- Search for review papers and easily understandable papers.

Nonetheless, the evaluation of ASE revealed a few shortcomings of the evaluation protocol: participants did not always remember ASE’s capabilities or how to operate them, the participant who aborted the evaluation did not appear to be motivated, the evaluation used a small dataset because ASE does not readily support importing new data, and the studies may not have been long enough. Applying MILCs in the novel context of literature exploration systems would overcome many of these shortcomings.

In summary, these usability studies provided clear guidance for work in developing ASE and specific recommendations to designers of literature exploration systems:

- Allow users control over which documents to explore, both to select an initial document set as well as drill down on a subset of documents.

- Provide easy-to-understand metrics for ranking documents.
- Give users a way to see overviews of the corpus using multiple coordinated views.
- Incorporate undo/redo functionality to navigate the history.
- Include capabilities to create groups of documents and annotations for documents.

Chapter 4

Conclusions and Future Work

4.1 Information Seeking Behavior in Action Science Explorer

In the context of the ASK model, ASE’s developers assume that information seekers have solved the linguistic non-specifiability of need—i.e., users are able to collect a corpus of potentially relevant documents. ASE may not be well suited for cognitive non-specifiability at the upper extreme where users know exactly what will satisfy their information need; in this case a digital library or search engine may be better suited to satisfying the users’ information needs. However, systems like ASE which support multiple methods of exploration may be well-suited for non-specifiable needs where information seekers’ needs are not well-formed and users cannot define their needs thoroughly. All of the tasks in the evaluations had well-defined information needs, but the results illustrate that some capabilities such as attribute ranking can provide an effective filter to target a search in a set of documents where the information is most likely to be found. Furthermore, attribute ranking seems to provide an alluring information scent which directs users who are beginning their search, and it seems to provide a simple way for users to focus their exploration on portions of the literature that are most important according to some ranking metric. Plot Nodes works in a similar way, but in cases where users want to rank by two variables it is more powerful.

4.2 Recommendations

Below are a list of recommendations that came from the evaluations of ASE.

4.2.1 User Requirements for Understanding Scientific Literature Networks

Allow users some level of control over which documents to explore.

The first step to analyze the field of dependency parsing was to create a custom set of documents, thus an important capability of ASE is to allow users to define a custom database. Users also frequently expressed the desire to choose a subset of the document set and then perform further analysis on the subset. This allows users to drill down onto a set of documents relevant to the users' current tasks and gives users the sense that they are in control of the analysis.

Give users an overview of the document set. Although participants did not gain insights about the document set from the network diagram alone, the participants frequently examined the network. It's possible that overviews are helpful to orient users in the document set and allow them to quickly get details-on-demand by clicking on a paper in the network diagram. However, further study about what kinds of overviews are most helpful is needed.

Provide users with easy-to-understand metrics for ranking documents. Attribute ranking was a nearly ubiquitous part of the analysis process. However, participants avoided metrics that were difficult to interpret in citation networks (e.g., betweenness centrality). Therefore, this thesis recommends that sys-

tem designers focus on providing users with easy-to-understand metrics for ranking documents.

Incorporate undo/redo functionality to navigate through users' exploration history. Shneiderman and Plaisant [58] discuss the easy reversal of actions in user interfaces and how it encourages users to explore. The participants in my first usability study had more difficulties than the participants in my second usability study with respect to going back to a previous navigation state, but it was a problem nonetheless. Since one of the main goals of ASE is to aid exploration, it is important to make users feel comfortable in all aspects of exploration.

Include capabilities to create groups of documents and annotations for documents. Users created named groups to annotate documents for reference, but doing so also helps them to drill down into these document subsets. If users can easily create custom documents groups, then they can later analyze these groups. Although the usability study participants did not create document annotations, their creating named groups is equivalent to annotating them.

4.2.2 Evaluation Strategies

The evaluation of digital libraries and literature exploration systems often focuses on usability issues, time to task completion, and errors while performing the task [68], but many users don't have well-defined search queries and use the systems to explore topics and identify important documents. Thus, evaluation techniques should account for individual users' goals and measure whether users can make useful

insights. With these goals in mind, MILCs [59] provide an appropriate framework for evaluating literature exploration systems. To my knowledge, MILCs have not been applied to the evaluation of literature exploration systems.

For longer term studies, it would be better to follow the suggestions of Plaisant and Shneiderman [48] and have several short clips about each type of functionality instead of a three videos that last one and a half minutes or more. During the one-and-a-half hour post-training portion of the evaluation sessions, users sometimes forgot about functionality that was introduced in the training videos. Having several short video clips would allow users to easily view tutorials about functionality whenever they need to refresh their memory. An additional possibility is to support embedded training, where the software system uses pop-ups, call out boxes, and animations to guide users in using the system's capabilities.

Plan the design of your literature exploration system to import data easily from at least one source, and identify a set of target evaluation participants early on. From repeating the usability studies with P1 and P3 it became clear that users can make interesting insights using ASE, but a MILCs-style evaluation that lasted two weeks or more may have revealed additional details about the insight-making process. Longer ethnographic studies also require researchers who are motivated to use the system for their own research goals.

Participants should analyze their own data. It became clear that motivation is a real problem for longer-term studies; the participant who volunteered to participate over several evaluation sessions eventually decided to stop participating, apparently due to motivation issues. Participants should have their own data and

problems they wish to solve. Although participants may not already have a well-formatted corpus to import into the system under evaluation, participants who have a research goal should be able to produce a corpus from other databases such as by performing a keyword search in a digital library to obtain a set of potentially interesting documents.

Utilize longer-term evaluations of your system. Participants in the usability studies described in Chapter 3 were able to make deeper insights the longer they used ASE—for example, after two usability evaluation sessions P1 used multiple capabilities in the system to learn something about the publication venues that no other participant had seen. It is unknown how long it may take to develop an insight, which is one goal of literature exploration systems. So evaluation protocols should have participants use the systems over a longer period of time, such as 2-4 weeks.

4.3 Limitations

One limitation of the evaluations is that the data set contained only 147 papers. These papers were chosen to demonstrate a particular use case for exploring the search results from a larger corpus. Participants occasionally commented that they thought some capabilities in ASE would work better with a larger set of documents, but nonetheless the participants were still able to make insightful comments on the papers.

A further limitation is that both sets of evaluations only had six participants

total. However, two of the participants used ASE in two sessions, which gives insight into the trajectory of use as users mature. Nonetheless, having more participants from other domains would give a broader perspective on how ASE is used. Extending the usability studies into case studies over a longer period of time would also give better insight into domain experts' working methods and users' abilities to effectively explore scientific literature networks.

Additionally, participants need to be appropriately motivated to use the system. In the usability studies described in this thesis, participants were motivated by offering compensation and recruiting computational linguists to analyze a literature network from a sub-field of computational linguistics. Although this worked well for extended usability studies, participants in longer-term case studies will need to be motivated by analyzing their own data and producing insights that are meaningful to their own investigations. An important measurement of the success of ASE will be whether expert users can achieve their own goals through using ASE.

4.4 Conclusion

This thesis makes five contributions, outlined below:

- A taxonomy of capabilities in literature search and exploration systems.
- A formative evaluation of ASE and its capabilities.
- A list of questions that information seekers ask when exploring new scientific literature domains.

- A list of recommendations to literature exploration system designers.
- A set of recommendations for evaluators of literature exploration systems.

Bibliography

- [1] Kenneth Allendoerfer, Serge Aluker, Gulshan Panjwani, Jason Proctor, David Sturtz, Mirjana Vukovic, and Chaomei Chen. Adapting the cognitive walk-through method to assess the usability of a knowledge domain visualization. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, Washington, DC, USA, 2005. IEEE Computer Society.
- [2] Christopher Andrews, Alex Endert, and Chris North. Space to think: large high-resolution displays for sensemaking. In *CHI '10: Proceedings of the 28th international conference on Human Factors in Computing Systems*, pages 55–64, New York, NY, USA, 2010. ACM.
- [3] Aleks Aris, Ben Shneiderman, Vahed Qazvinian, and Dragomir Radev. Visual overviews for discovering key papers and influences across research fronts. *Journal of the American Society for Information Science and Technology*, 60:2219–2228, November 2009.
- [4] Marcia J. Bates. Where should the person stop and the information search interface start? *Information Processing and Management*, 26:575–591, October 1990.
- [5] N.J. Belkin. Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5:133–143, 1980.
- [6] N.J. Belkin, R.N. Oddy, and H.M. Brooks. ASK for information retrieval: Part I. background and theory. *Journal of Documentation*, 38(2):61–71, 1982.
- [7] Enrico Bertini, Adam Perer, Catherine Plaisant, and Giuseppe Santucci. BELIV '08: Beyond time and errors: novel evaluation methods for information visualization. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 3913–3916, New York, NY, USA, 2008. ACM.
- [8] Kevin W. Boyack and Katy Börner. Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information Science and Technology*, 54(5):447–461, 2003.
- [9] Center for History and New Media. Zotero. <http://www.zotero.org/>. March, 2011.
- [10] Chaomei Chen. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci. Technol.*, 57:359–377, February 2006.

- [11] Chaomei Chen and Mary P. Czerwinski. Empirical evaluation of information visualizations: an introduction. *Int. J. Hum.-Comput. Stud.*, 53(5):631–635, 2000.
- [12] Chaomei Chen, Fidelia Ibekwe-SanJuan, and Jianhua Hou. The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *J. Am. Soc. Inf. Sci. Technol.*, 61:1386–1409, July 2010.
- [13] Douglass R. Cutting, David R. Karger, Jan O. Pedersen, and John W. Tukey. Scatter/gather: a cluster-based approach to browsing large document collections. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '92, pages 318–329, New York, NY, USA, 1992. ACM.
- [14] Mary P. Czerwinski, Greg Smith, Tim Regan, Brian Meyers, George G. Robertson, and Gary Starkweather. Toward characterizing the productivity benefits of very large displays. In *Interact*, pages 9–16. IFIP, 2003.
- [15] George S. Davidson, Bruce Hendrickson, David K. Johnson, Charles E. Meyers, and Brian N. Wylie. Knowledge mining with vxinsight: Discovery through interaction. *J. Intelligent Info Sys*, 11:259–285, 1998.
- [16] Cody Dunne, Ben Shneiderman, Bonnie Dorr, and Judith Klavans. iOpener Workbench: Tools for rapid understanding of scientific literature. In *Human-Computer Interaction Lab 27th Annual Symposium*, University of Maryland, College Park, MD, 2010.
- [17] Cody Dunne, Ben Shneiderman, Robert Gove, Judith Klavans, and Bonnie Dorr. Action Science Explorer: tools for rapid understanding of scientific literature. Human-Computer Interaction Lab Tech Report HCIL-2011, University of Maryland, 2011.
- [18] Michael D. Ekstrand, Praveen Kannan, James A. Stemper, John T. Butler, Joseph A. Konstan, and John T. Riedl. Automatically building research reading lists. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 159–166, New York, NY, USA, 2010. ACM.
- [19] Geoffrey Ellis and Alan Dix. An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pages 1–7, New York, NY, USA, 2006. ACM.
- [20] Sarah Faisal, Brock Craft, Paul Cairns, and Ann Blandford. Internalization, qualitative methods, and evaluation. In *Proceedings of the 2008 conference on BEyond time and errors: novel evaluation methods for Information Visualization*, pages 5:1–5:8, New York, NY, USA, 2008. ACM.

- [21] Victor Gonzalez and Alfred Kobsa. A workplace study of the adoption of information visualization systems. In *I-KNOW'03: 3rd International Conference on Knowledge Management*, pages 92–102, 2003.
- [22] Google. Google Scholar. <http://scholar.google.com/>. March, 2011.
- [23] Qi He, Jian Pei, Daniel Kifer, Prasenjit Mitra, and Lee Giles. Context-aware citation recommendation. In *Proceedings of the 19th international conference on World wide web, WWW '10*, pages 421–430, New York, NY, USA, 2010. ACM.
- [24] Marti A. Hearst. *Search User Interfaces*. Cambridge University Press, 2009.
- [25] Tom Hewett, Mary Czerwinski, Michael Terry, Jay Nunamaker, Linda Candy, Bill Kules, and Elisabeth Sylvan. Creativity support tool evaluation methods and metrics. In *NSF Workshop Report on Creativity Support Tools*, September 2005.
- [26] Linda L. Hill, Larry Carver, Ron Dolin, Terence R. Smith, James Frew, and Mary-Anna Rae. Alexandria digital library: user evaluation studies and system design. *J. Am. Soc. Inf. Sci.*, 51:246–259, February 2000.
- [27] Orland Hoeber and Xue Dong Yang. Exploring web search results using coordinated views. In *Proc. Int. Conf. Coordinated & Multiple Views in Exploratory Visualization*, pages 3–13, 2006.
- [28] John Hughes, Val King, Tom Rodden, and Hans Andersen. The role of ethnography in interactive systems design. *Interactions*, 2(2):56–65, 1995.
- [29] W. J. Hutchins. On the problem of “aboutness” in document analysis. *J. of Informatics*, 1(1):17–35, 1977.
- [30] W. J. Hutchins. The concept of “aboutness” in subject indexing. In *Aslib Proceedings*, pages 172–181, 1978.
- [31] JabRef Development Team. *JabRef*. JabRef Development Team, 2011.
- [32] Eser Kandogan and Ben Shneiderman. Elastic windows: evaluation of multi-window operations. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 250–257, New York, NY, USA, 1997. ACM.
- [33] Carol C. Kuhlthau. Inside the search process: Information seeking from the user’s perspective. *J. Amer. Soc. Inform. Sci.*, 42(5):361–371, 1991.
- [34] Sean M. McNee, Istvan Albert, Dan Cosley, Prateep Gopalkrishnan, Shyong K. Lam, Al Mamunur Rashid, Joseph A. Konstan, and John Riedl. On the recommending of citations for research papers. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work, CSCW '02*, pages 116–125, New York, NY, USA, 2002. ACM.

- [35] Mendeley Ltd. Mendeley. <http://www.mendeley.com/>. March, 2011.
- [36] Saif Mohammad, Bonnie Dorr, Melissa Egan, Ahmed Hassan, Pradeep Muthukrishnan, Vahed Qazvinian, Dragomir Radev, and David Zajic. Using citations to generate surveys of scientific paradigms. In *Proc. Human Language Technologies: The 2009 Ann. Conf. North Amer. Chapter Ass. for Computational Linguistics*, NAACL '09, pages 584–592, Morristown, NJ, USA, 2009. Association for Computational Linguistics.
- [37] Galileo Mark Namata, Brian Staats, Lise Getoor, and Ben Shneiderman. A dual-view approach to interactive network visualization. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 939–942, New York, NY, USA, 2007. ACM.
- [38] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6):066133, June 2004.
- [39] Chris North. Toward measuring visualization insight. *Computer Graphics and Applications, IEEE*, 26(3):6–9, 2006.
- [40] Chris North and Ben Shneiderman. A taxonomy of multiple window coordinations. Technical Report HCIL-1997-18, University of Maryland, 1997.
- [41] Souneil Park, SangJeong Lee, and Junehwa Song. Aspect-level news browsing: understanding news events from multiple viewpoints. In *Proc. int. conf. Intelligent user interfaces, IUI '10*, pages 41–50, New York, NY, USA, 2010. ACM.
- [42] Pennsylvania State University, College of Information Sciences and Technology. CiteSeerX. <http://citeseerx.ist.psu.edu/>. March, 2011.
- [43] Adam Perer and Ben Shneiderman. Balancing systematic and flexible exploration of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12:693–700, September 2006.
- [44] Adam Perer and Ben Shneiderman. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In *Proc. SIGCHI conf. Human factors in computing syst.*, CHI '08, pages 265–274, New York, NY, USA, 2008. ACM.
- [45] Carolina Perez-Iratxeta, Antonio J. Perez, Peer Bork, and Miguel A. Andrade. Update on XplorMed: a web server for exploring scientific literature. *Nucleic Acids Research*, 31(13):3866–3868, 2003.
- [46] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. Int. Conf. Intelligence Anal.*, 2005.

- [47] Peter Pirolli, Patricia Schank, Marti Hearst, and Christine Diehl. Scatter/gather browsing communicates the topic structure of a very large text collection. In *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground*, CHI '96, pages 213–220, New York, NY, USA, 1996. ACM.
- [48] C. Plaisant and B. Shneiderman. Show me! guidelines for producing recorded demonstrations. In *2005 IEEE Symposium on Visual Languages and Human-Centric Computing*, pages 171 – 178, 2005.
- [49] Catherine Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '04, pages 109–116, New York, NY, USA, 2004. ACM.
- [50] Dragomir Radev, Jahna Otterbacher, Adam Winkel, and Sasha Blair-Goldensohn. NewsInEssence: summarizing online news topics. *Commun. ACM*, 48:95–98, October 2005.
- [51] Dragomir R. Radev, Pradeep Muthukrishnan, and Vahed Qazvinian. The ACL anthology network corpus. In *Proc. ACL Workshop on Natural Language Process. and Inform. Retrieval for Digital Libraries*, Singapore, 2009.
- [52] J.C. Roberts. State of the art: Coordinated multiple views in exploratory visualization. pages 61 –71, July 2007.
- [53] P. Saraiya, C. North, V. Lam, and K.A. Duca. An insight-based longitudinal study of visual analytics. *IEEE Trans. Vis. Comput. Graphics*, 12(6):1511–1522, November 2006.
- [54] Purvi Saraiya, Chris North, and Karen Duca. Comparing benchmark task and insight evaluation methods on timeseries graph visualizations. In *BELIV '10: Proc. 2010 AVI workshop on BEyond time and errors*, 2010.
- [55] Jinwook Seo and B. Shneiderman. Knowledge discovery in high-dimensional data: case studies and a user survey for the rank-by-feature framework. *IEEE Trans. Vis. Comput. Graphics*, 12(3):311–322, 2006.
- [56] M.H. Shimabukuro, E.F. Flores, M.C.F. de Oliveira, and H. Levkowitz. Coordinated views to assist exploration of spatio-temporal data: a case study. In *Proc. Int. Conf. Coordinated & Multiple Views in Exploratory Visualization*, pages 107–117, 2004.
- [57] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, pages 336 –343, September 1996.
- [58] Ben Shneiderman and Catherine Plaisant. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley, 2004.

- [59] Ben Shneiderman and Catherine Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pages 1–7. ACM, 2006.
- [60] Lauren Shupp, Christopher Andrews, Margaret Dickey-Kurdziolek, Beth Yost, and Chris North. Shaping the display of the future: The effects of display size and curvature on user performance and insights. *Human-Computer Interaction*, 24:230–272, 2009.
- [61] R.S. Taylor. Process of asking questions. *American Documentation*, 13:391–396, 1962.
- [62] R.S. Taylor. Question-negotiation and information seeking in libraries. *College and Research Libraries*, 28:178–194, 1968.
- [63] Thomson Reuters. ISI web of knowledge. <http://www.isiwebofknowledge.com/>. March, 2011.
- [64] Thomson Reuters. EndNote. <http://www.endnote.com/>, March 2011.
- [65] Roberto Torres, Sean M. McNee, Mara Abel, Joseph A. Konstan, and John Riedl. Enhancing digital libraries with techlens+. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '04, pages 228–236, New York, NY, USA, 2004. ACM.
- [66] Transinsight. GoPubMed. <http://www.gopubmed.org/>. March, 2011.
- [67] Eliane R. A. Valiati, Carla M. D. S. Freitas, and Marcelo S. Pimenta. Using multi-dimensional in-depth long-term case studies for information visualization evaluation. In *BELIV '08: Proc. 2008 Conf. BEyond time and errors*, pages 1–7, New York, NY, USA, 2008. ACM.
- [68] Hong Iris Xie. Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment. *Inf. Process. Manage.*, 44:1346–1373, May 2008.
- [69] Ji Soo Yi, Youn-ah Kang, John T. Stasko, and Julie A. Jacko. Understanding and characterizing insights: how do people gain insights using information visualization? In *BELIV '08: Proc. 2008 Conf. BEyond time and errors*, pages 1–6, New York, NY, USA, 2008. ACM.
- [70] Xiaojun Yuan, Xiangmin Zhang, and Alex Trofimovsky. Testing visualization on the use of information systems. In *Proceeding of the third symposium on Information interaction in context*, IiiX '10, pages 365–370, New York, NY, USA, 2010. ACM.
- [71] D. Zajic, B. Dorr, R. Schwartz, C. Monz, and J. Lin. A sentence-trimming approach to multi-document summarization. In *Proc. Workshop Text Summarization*, pages 151–158, 2005.

- [72] David Zajic, Bonnie J. Dorr, Jimmy Lin, and Richard Schwartz. Multi-candidate reduction: Sentence compression as a tool for document summarization tasks. *Inf. Process. Manage.*, 43:1549–1570, 2007.
- [73] Pengyi Zhang, Lynne Plettenberg, Judith L. Klavans, Douglas W. Oard, and Dagobert Soergel. Task-based interaction with an integrated multilingual, multimedia information system: a formative evaluation. In *JCDL '07: Proc. Joint Conf. Digital libraries*, pages 117–126, New York, NY, USA, 2007. ACM.
- [74] Xiangmin Zhang, Jinjing Liu, Yuelin Li, and Ying Zhang. How usable are operational digital libraries: a usability evaluation of system interactions. In *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*, EICS '09, pages 177–186, New York, NY, USA, 2009. ACM.