iOpener Workbench: Tools for Rapid Understanding of Scientific Literature

Cody Dunne, Ben Shneiderman, Bonnie Dorr & Judith Klavans
Department of Computer Science
Contact: {cdunne, ben, bonnie}@cs.umd.edu, jklavans@umd.edu
tangra.si.umich.edu/clair/iopener

Scholars and scientists today have the challenging task before them of keeping up with the advances within their rapidly expanding fields. The growing number of publications each year combined with ever-increasing cross-disciplinary research make following all the branches of research in a field and identifying key papers challenging. Rapid generation of accurate survey articles can help ease these difficulties, providing readers with concise overviews tailored to their needs. Students, educators, scientists, and government decision makers all have vastly different levels of expertise and requirements for learning about scientific fields.

The goal of iOpener (Information Organization for PENning Expositions on Research) is to generate readily-consumable surveys of different scientific domains and topics, targeted to different audiences and levels. We've created an infrastructure for automatic summarization of research domains that links bibliometric lexical link mining, summarization techniques, and visualization tools. Part of this is the *iOpener Workbench*, a new tool which presents the academic literature for a field using many different modalities: lists of articles, their full texts, automatic textual summaries, and visualizations of the structure of the citation network – all shown in Figure 1.
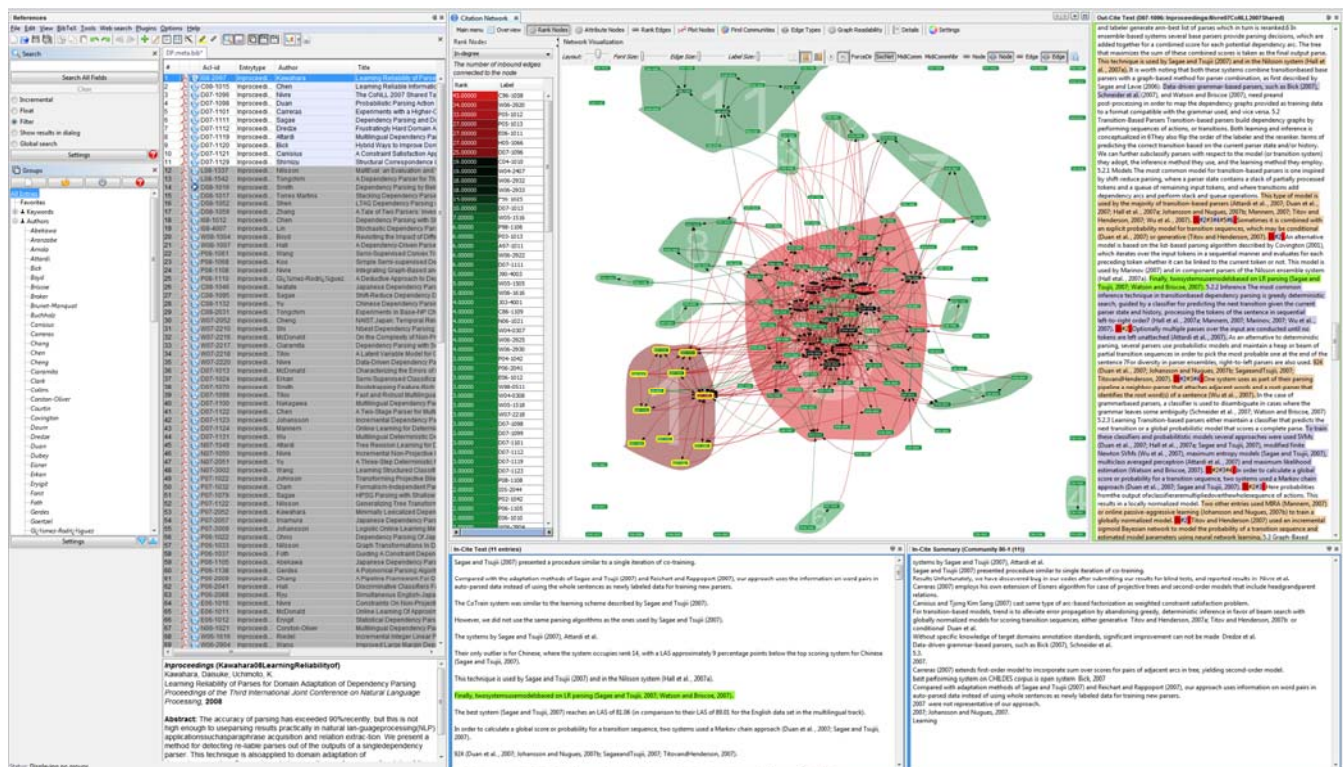

Figure 1 The *iOpener Workbench* combines (clockwise from top) a citation network analysis tool (SocialAction), full-text with linked citations, multi-document summaries of clusters, the text of incoming citations, and reference manager (JabRef),

The iOpener Workbench is partially an integration of two powerful existing tools – the SocialAction network analysis tool[1] and the JabRef reference manager[2]. SocialAction provides us with powerful network analysis capabilities including force-directed citation network visualization, ranking and filtering papers by statistical measures, scatterplots of paper attributes and statistics, categorical and numerical range coloring, and automatic cluster detection (shown using convex colored hulls in Figure 1). Using visualizations of the citation network we can easily find unexpected trends, clusters, gaps and outliers. Additionally, visualizations can immediately identify invalid data that is easily missed in tabular views.

JabRef supplies all the features one would expect from a reference manager, including searching using simple regular expressions, automatic and manual grouping of papers, DOI and URL links, PDF full text with annotations, abstracts, user generated reviews and text annotations, and many ways of exporting. It integrates with Microsoft Word, OpenOffice.org, and BibTeX, which allows quick adding of citations to discovered articles when writing survey papers.

These tools are linked together to form multiple coordinated views of the data. Clicking on a node in the citation network selects it and its corresponding paper in the reference manager, displaying its abstract, review, and other data associated with it. Moreover, when clusters of nodes are selected their papers are floated to the top of the reference manager for easy perusal. The inverse is true as well, with any paper, group, or search term selected in the reference manager highlighting the corresponding nodes in the network.

There are other coordinated views that provide the user with other aspects of the field. When any node or cluster is selected, the In-Cite Text window displays the text of all incoming citations to the paper(s), i.e. the whole sentences from the citing papers that include the citation to the selected paper(s). These are displayed in a hyperlinked list that allows the user to select any one of them to show their surrounding context in the Out-Cite Text window. This window shows the full text of the paper citing one of the selected papers, with highlighting showing the selected citation sentence as well as any other sentences that include hyperlinked citations to other papers. The last view is the summary window, which can contain various multi-document summaries of a selected cluster. Using automatic summarization techniques, we can summarize all of the incoming citations to papers within that cluster, hopefully providing key insights into that research community.

The iOpener Workbench integrates these many components in order to provide a tool that supports rapid understanding of scientific literature. Users can analyze the network of citations between papers, identify key papers and research clusters, automatically summarize them, dig into the full text of articles to extract context, make annotations, write reviews, and finally export their findings in many of document authoring formats. We hope this infrastructure will enable users to generate readily-consumable surveys of scientific fields.

Acknowledgements

PAPERS[3]

1. Aris, Aleks, Ben Shneiderman, Vahed Qazvinian, Dragomir Radev. Visual Overviews for Discovering Key Papers and Influences Across Research Fronts. JASIST 60(11): 2219-2228 (2009).

2. Mohammad, Saif, Bonnie Dorr, Dragomir Radev, Melissa Egan, Ahmed Hassan, Pradeep Muthukrishnan, Vahed Qazvinian, David Zajic. Using Citations to Generate Surveys of Scientific Paradigms. In NAACL-HLT 2009.

3. Qazvinian, Vahed and Dragomir R. Radev. Scientific paper summarization using citation summary networks. In COLING 2008, Manchester, UK, 2008.

4. Radev, Dragomir R. and Joseph, Mark and Gibson, Bryan and Muthukrishnan, Pradeep. A Bibliometric and Network Analysis of the Field of Computational Linguistics. Journal of the American Society for Information Science and Technology (JASIST).

5. Shneiderman, B., Research Agenda: Visual Overviews for Exploratory Search, Position paper for National Science Foundation Workshop on Information Seeking Support Systems, Chapel Hill, NC, June 26-27, 2008.

---

[1] www.cs.umd.edu/hcil/socialaction

[2] jabref.sourceforge.net

[3] This work is not yet published. Related papers are listed below.