Discovery and Visualization of Expertise in a Scientific Community

Muhammad Tanvir Afzal[1], Atif Latif[2], Anwar Us Saeed[2] , Philipp Sturm[3] , Salman Aslam[4]
Keith Andrews[1], Klaus Tochtermann[2], Hermann Maurer[1]

[1]Institute for Information Systems and Computer Media
Graz University of Technology, Austria
{mafzal,kandrews,hmaurer} @iicm.edu

[2]Institute for Knowledge Management, Know-Center
Graz University of Technology, Austria
{atif.latif, anwar.ussaeed}@student.tugraz.at, klaus.tochtermann@tugraz.at

[3]Graz University of Technology, Austria
psturm@sbox.tugraz.at

[4]COMSATS Institute of Information Technology, Islamabad, Pakistan
salman_aslam@comsats.edu.pk

ABSTRACT

In numerous contexts and environments, it is necessary to identify and assign (potential) experts to subject fields. In the context of an academic journal for computer science (J.UCS), papers and reviewers are classified using the ACM classification scheme. This paper describes a system to identify and present potential reviewers for each category from the entire body of paper's authors. The topical classification hierarchy is visualized as a hyperbolic tree and currently assigned reviewers are listed for a selected node (computer science category). In addition, a spiral visualization is used to overlay a ranked list of further potential reviewers (high-profile authors) around the currently selected category. This new interface eases the task of journal editors in finding and assigning reviewers. The system is also useful for users who want to find research collaborators in specific research areas.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval] *Information filtering.* H.5.2 [Information Interfaces and Presentation (e.g., HCI)] User Interfaces - graphical user interfaces (GUI).

General Terms

Algorithms, Design, Human Factors

Keywords

Expertise finder, reviewers, J.UCS, ACM classification scheme, user interface, hyperbolic, spiral.

1. INTRODUCTION

The discovery of expertise is crucial in supporting a number of tasks. A variety of tools have been implemented within organisations to find experts and expertise for different scenarios. Most related works make use of explicitly specified expert profiles constructed manually. The problem with such manually constructed profiles is that they tend to be developed for particular projects and constantly need to be updated e.g. [18].

Using an entirely automated mechanism for determining user expertise may also not be adequate in itself. As an illustration, Google Scholar employed an automated approach and wrongly identified names of places such as Ann Arbour, or Milton Keynes as cited authors [19]. This also highlights the non-trivial nature of expertise mining and the difficulty faced in the disambiguation of individuals.

We propose an automated technique which incorporates multiple facets in providing a more representative assessment of expertise as explained in Section 3. To overcome automation errors mentioned above, we introduced an innovative citation mining technique [4]. We see these facets as providing multiple sources of evidence for a more reflective perspective of experts. We present the combination of both tangible and intangible metrics to shed deeper insights into the intensity of expertise. The system mines multiple facets for an electronic journal and then calculates expertise' weights. The measures provided are, however, not absolute indicators of expertise as the discoveries are limited by the coverage of the database of publications and expert profiles used.

The system discoveries can be enhanced by visualizing the mined data [21]. In order to enhance the knowledge discoveries, we have visualized experts by using hyperbolic tree visualization technique. The proposed technique is based on focus plus context with extended focus to represent the statistical data as explained in section 6.

2. RELATED WORK

Expertise finder systems in the past, have been innovatively applied in helping PhD applicants in finding relevant supervisors [15] and also in identifying peer-reviewers for a conference [20]. The former made use of a manually constructed expertise profile database while the latter employed reference mining for all papers submitted to a conference. In the latter, a co-authorship network was constructed for each submitted paper making use of a measure of conflict-of-interest to ensure that papers were not reviewed by associates.

Cameron [6] employed a manually crafted taxonomy of 100 topics in DBLP [8] covering the research areas of a small sample of researchers appearing in DBLP. They proposed the need for automatic taxonomy creation as a key issue in finding experts. Mockus et al [17] employed data from a software project's change management records to locate people with desired expertise in a large organization. Their work indicated a need to explicitly represent experiential characterization of individuals as a means of providing insights into the knowledge and skills of individuals. Yimam [25] have further shown that a decentralized approach can be applied for information gathering in the construction of expertise profiles. Tho et al [23] employed a citation mining retrieval technique where a cross mapping between author clusters and topic clusters was applied to assign areas of expertise to serve as an additional layer of search results organization.

There are also expertise detection systems that were based entirely on an analysis of user activity and behavior while being engaged in an electronic environment. Krulwich et al [11] have analyzed the number of interactions of an individual within a discussion forum as a means of constructing an expert's profile. Although such an approach is useful in monitoring user participation, measures such as number of interactions on a particular topic is in itself not reflective of knowledge levels of individuals.

Information visualization techniques have been used to visualize large datasets to support exploration and in finding hidden patterns [5]. To visualize large hierarchal structures, the hyperbolic tree was developed by Xerox [13]. The principle of Focus plus Context is supported by a detailed view for the focused part of the data in the center of the display, while the overall hierarchal structure of data remains visible around the edges. In computer science, ACM categories are widely used to organize scientific work. ACM categories can be seen as a hierarchal taxonomy and can be visualized using a hyperbolic tree. To visualize experts in a proper ranking for a specific ACM category, spiral visualization is appropriate. The RankSpiral was used by [22] to maximize information density and minimize occlusions for large documents. We have applied a similar approach for the visualization of experts around a particular node in the ACM category tree.

3. A MULTI-FACETED EXPERT PROFILE

In exploring a comprehensive characterization of expertise, we proposed a multifaceted approach of mining the expertise for a digital journal [3]. The multiple facets are represented by the following measurements: number of publications, number of citations received, extent and proportion of citations within a particular area, expert profile records, and experience. We have
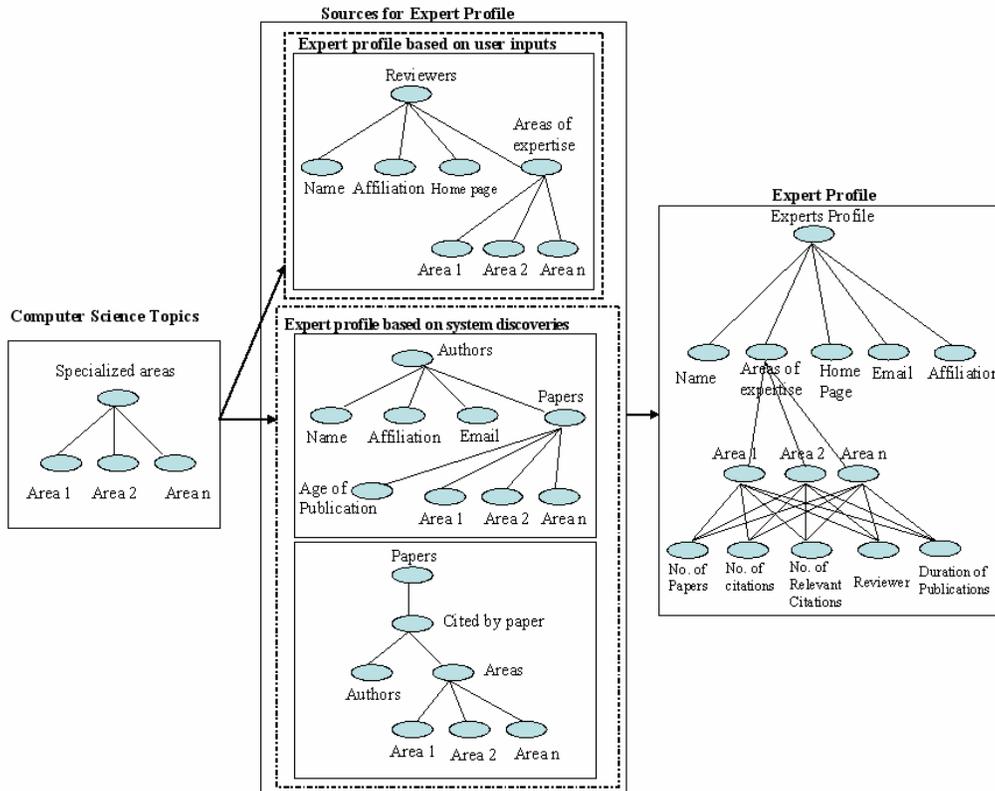

Figure 1. Expert profile.

thus incorporated the use of user-defined profiles, "experience atom" (as proposed by [17] to indicate fundamental experiential units), reference mining results and a characterization of expert participation as facets of an expert profile. Combining all these factors provides a better indication of expertise with regards to a particular topic. Figure 1 shows the consolidated view of expert profile construction as applied in our research.

In our research, there are two main sources of information used to construct an expert profile 1) user inputs and 2) system discoveries. User inputs are taken from reviewers of the journal J.UCS [10]. J.UCS has over 300 reviewers on its editorial board. The expertise of these reviewers are specified and maintained according to the ACM classification scheme [1]. This information was extracted from J.UCS and used to populate the expert profile database.

The second source for constructing expert profiles is computed by the system. The computation considers the number of publications of an individual, number of citations a person receives, and the person's duration of publication in the respective area. The extraction of all publications (over 1,400) along with authors and co-authors of the publication is described in [2] with a set of over 15,000 references [4].

4. DATA EXTRACTION

Within J.UCS, ACM topics, editors, and every individual paper are represented in an XML notation, which needs to be parsed to extract metadata. A typical XML file for J.UCS papers can be seen in Figure 2. The metadata (paper title, authors, ACM topic, etc.) related to a paper is represented inside the XML file.

The extracted data was used to populate a relational database. The database presents a coherent view of all data with relationships (category, paper, authors, and citations). For citation extraction, a technique called Template-based Information Extraction using Rule-based Learning (TIERL) was developed [4]. TIERL outperformed existing citation extraction approaches (like ISI, Google Scholar, and CiteSeer). The data from this database was then used to calculate and visualize experts within the J.UCS environment.

5. WEIGHTS ASSIGNED to EXPERTS

In our system, experts are grouped into one of two categories: 1) *reviewers* (persons currently manually assigned as reviewers for a particular ACM topic category) and 2) *high-profile authors* (persons flagged automatically as experts in a particular topic). Reviewers are selected by the editor-in-chief based on their expertise in the respective ACM topical area. Reviewers for a particular ACM category are visualized without any further calculation. High-profile authors are calculated based on weights assigned to them. The facets defined in Figure 1 are used to assign the weights. Three weights called publication weight, citation weight, and reviewer weight are calculated as follows:

Publication Weight = No. of publications / duration (No. of years).

Citation Weight = No. of citations received by an author / total citations in an ACM topic

Reviewer Weight = No. of reviewers of J.UCS / Total no. of J.UCS Authors.

Figure 2. A sample paper XML File.

The total weight is defined as the sum of the three previous weights:

Total weight = publication weight + citation weight + reviewer weight

High-profile authors are then ranked according to their total weight.

6. Information Visualization

Reviewers are essentially attached to a node within the ACM classification hierarchy. For each node within the ACM classification hierarchy, a ranked list of high-profile authors (potential reviewers) can then be calculated. The hyperbolic browser [12] [13] [14] is an efficient visualization technique for large hierarchies. A hyperbolic browser is used to provide intuitive navigation within the ACM classification hierarchy. For any selected node in the ACM hierarchy, a spiral is used to visualize the ranked list of high-profile authors for that node. The spiral is simply superimposed upon and around the selected node. This builds on past work with GopherVR [16], PRISE [7], and RankSpiral [22] which all use spiral representations to display ranked search result lists.

7. VISUALIZATION of EXPERTS

The user interface is implemented in Java. A hyperbolic browser is used to visualize the ACM classification hierarchy, using the freely available Hypertree package[9]. Both categories of experts. are visualized by superimposing upon the hyperbolic view. Reviewers are shown in a simple list and high profile authors are
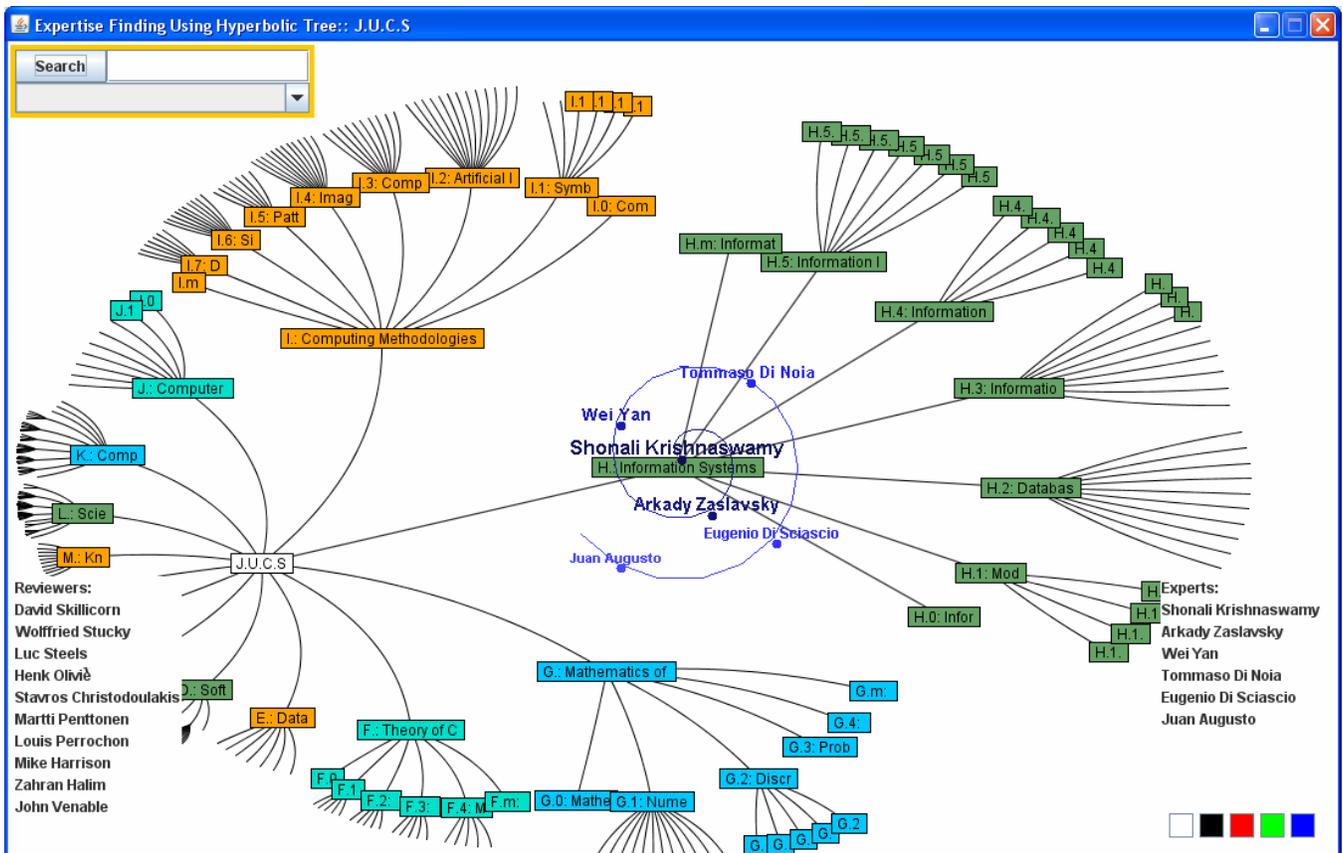
Figure 3. User Interface

shown in a spiral visualization. To draw the spiral, a package called Turtle Graphics is used [24]. With Turtle Graphics, simple commands are used to move and draw on the graphical surface. With these commands, the spiral is drawn and the names of the experts are written at constant angular steps. To visualize the reviewers of a specific ACM topic, a simple JList is used. A maximum of 10 reviewers are shown in the JList.

The JList, spiral, and Hypertree are placed in JPanels inside a frame, and are ordered with a JLayeredPane. One can arrange the JPanels horizontally and vertically and even manipulate the z-order. The Hypertree is drawn in the back. When an ACM topic is clicked, the list of reviewers is shown in the bottom left and the spiral of high-profile authors is overlaid over the ACM topic in the top layer, as shown in Figure 3. When there are neither reviewers nor high-profile authors, no list or spiral is drawn. In the bottom right of the window, there are five colored buttons. When clicked, the spiral is redrawn with the new color. It is possible to choose black, red, green, or blue. Users can hide both the spiral and the reviewers list if required by clicking white button. When users drag a particular node, the spiral moves with the focused node.

Figure 3 shows the visualization for ACM category H. Information Systems. The reviewers are shown in the bottom left corner. When a user clicks on the node "H. Information Systems", a spiral is drawn around the selected node. The high-profile authors are placed in the spiral in descending order of their total weight (the highest weighted in the center of the spiral).

This visualization is useful both for users who want to find potential collaborators and for J.UCS administrators. For example, in J.UCS there are some ACM topics with very few assigned reviewers. J.UCS administration can instantly find potential reviewers based on the high-profile authors shown by the system. For instance, in Figure 4 the topic 'M.8 Knowledge Reuse' has no reviewers at the moment (this is a new category
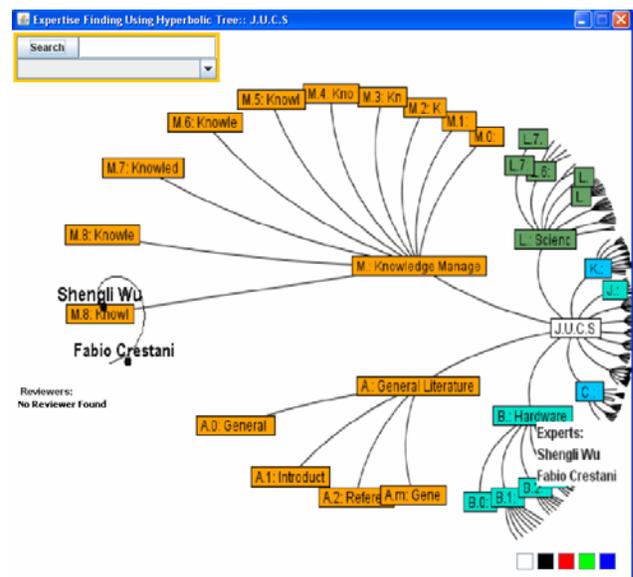

Figure 4. Discovery of potential reviewers

added by the J.UCS). Potential reviewers are easily found in the visualization, as shown in Figure 4. This type of discovery is very useful for administrators to locate potential reviewers for any selected area.

Although it is convenient to explore the topical hierarchy with the hyperbolic tree, users sometimes know the name of a category and want to navigate directly to it. The search facility in the top left corner of the main interface (see Figure 5) supports this task. For example, if a user searches for the term "Information", then a combo box is filled with all ACM topics containing the term "Information" as a substring. The 13 topics containing the term "Information" are shown in Figure 5.

The user can select any ACM topic from the search result list and the hyperbolic tree is redrawn to show the selected topic centered in the window.
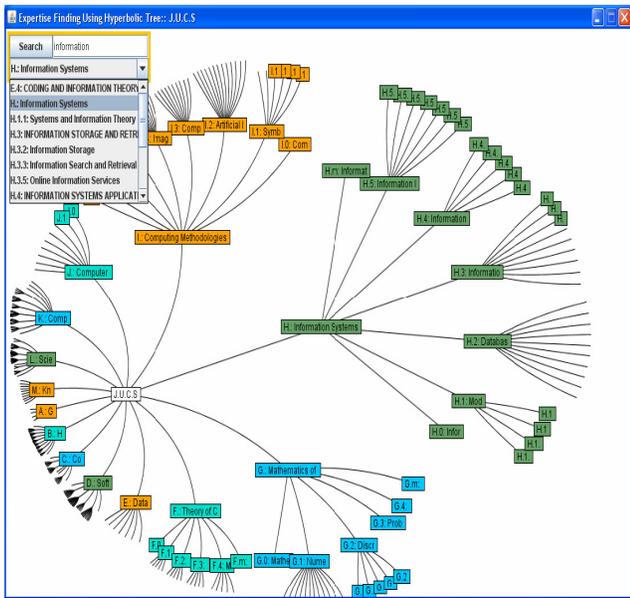

Figure 5. ACM keyword search facility

8. CONCLUDING REMARKS

This paper presented a new system to identify and visualize current and potential experts in topical areas of a scientific discipline. It is used in the context of a computer science journal to identify and assign reviewers to areas of computer science, but is easily generalized to other scientific communities.

The main contributions of this paper are:

- A methodology for automatically identifying potential experts from assembled profiles.

- A combined visualization of a topical classification hierarchy and a ranked list of potential experts at each level in the hierarchy.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] ACM Classification Scheme, http://www.acm.org/about/class/1998/

[2] Afzal, M.T., Kulathuramaiyer, N., and Maurer, H. 2007. Creating Links into the Future. Journal of Universal Computer Science. 13, 9 (Sep. 2007), 1234-1245. doi:10.3217/jucs-013-09-1234

[3] Afzal, M.T., Kulathuramaiyer, N., and Maurer, H. 2008. Expertise Finding for an Electronic Journal. In proceedings of I-Know (Graz, Austria, Sep. 3- 5 2008). I-Know'08. 436-440.

[4] Afzal, M.T., Maurer, H., Balke, W.T., Kulathuramaiyer, N. 2009. Improving Citation Mining. In proceedings of Networked Digital Technologies (Ostrava, Czech Republic, Jul. 29-31 2009). NDT'09. 122-127. doi:10.1109/NDT.2009.5272186

[5] Card, S. K., Mackinlay, J., Shneiderman, B.1999. Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, San Francisco, CA. ISBN-13: 978-1558605336.

[6] Cameron, D., Aleman-Meza, B., Arpinar, I.B. 2007. Collecting Expertise of Researchers for Finding for Relevant Experts in Peer-Review Setting. In the proceeding of 1st International Expert Finder Workshop (Berlin, Germany, Jan 16 2007).

[7] Cugini, J. V., Piatko, C. D., and Laskowski, S. J. 1996. Interactive 3D Visualization for Document Retrieval. In Proceedings of the Workshop on New Paradigms in Information Visualization and Manipulation at CIKM (Rockville, Maryland, USA, Nov. 12-16, 1996). NPIVM'96. ACM Press.

[8] Digital Bibliography and Library Project, http://www.informatik.uni-trier.de/~ley/db/index.html

[9] Hyperbolic Tree Library http://hypertree.cvs.sourceforge.net/viewvc/hypertree/hypertree/

[10] Journal of Universal Computer Science, J. UCS, http://www.jucs.org

[11] Krulwich, B., Burkey, C. 1995. ContactFinder: Extracting Indications of Expertise and Answering Questions with Referrals. Technical Report.In the Working Notes of the Symposium on Intelligent Knowledge Navigation and Retrieval. The AAAI Press, 85-91.

[12] Lamping, J., Rao, R. 1994. Laying out and Visualizing Large Trees Using a Hyperbolic Space. In ACM Symposium on User Interface Software and Technology (Marina del Rey, California, USA, Nov. 2-4, 1994) UISTP'94. 13-14. doi:10.1145/192426.192430.

[13] Lamping, J., Rao, R. 1996. The Hyperbolic Browser: A Focus+Context Technique for Visualizing Large Hierarchies.

Journal of Visual Languages and Computing. 7, 1 (Mar. 1996), 33-55. doi:10.1006/jvlc.1996.0003.

[14] Lamping, J., Rao, R., Pirolli, P. 1995. A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Denver, Colorado, May 7 - 11, 1995). CHI '95. ACM Press, New York, NY, 401-408. doi:10.1145/223904.223956

[15] Liu, P., Dew, P. 2004. Using Semantic Web Technologies to Improve Expertise Matching Within Academia. In Proceedings of I-Know (Graz, Austria, Jun. 30- July 02 2004). I-Know'04. 370-378.

[16] McCahill, M. P., Erickson, T. 1995. Design for a 3D Spatial User Interface for Internet Gopher. In Proceedings of World Conference on Educational Multimedia and Hypermedia (Graz, Austria, Jun. 17-21, 1995). ED-MEDIA 95. AACE, 39-44.

[17] Mockus, A., Herbsleb, J. D. 2002. Expertise Browser: A Quantitative Approach to Identifying Expertise. In Proceedings on the International Conference on Software Engineering (Florida, USA, May 19-25 2002). ICSE'02. 503-312. doi:10.1145/581339.581401.

[18] Pipek, V., Hinrichs, J., Wulf, V. 2002. Sharing Expertise Challenges for Technical Support. In Ackerman, M./Pipek, V./Wulf, V. (eds): Beyond Knowledge Management: Sharing Expertise, MIT-Press, Cambridge MA. 111-136.

[19] Postellon D.C. 2008. Hall and Keynes join Arbor in the Citation Indices. Nature, 452, 282. doi:10.1038/452282b

[20] Rodriguez, M.A., Bollen, J. 2008. An Algorithm to Determine Peer- Reviewers. In the Proceeding of the 17th ACM conference on Information and Knowledge Management (Napa Valley, California, USA, Oct. 26-30 2008) CIKM'08. ACM Press. 319-328. doi:10.1145/1458082.1458127.

[21] Shneiderman, B. 2002. Inventing Discovery Tools: Combining Information Visualization with Data Mining. . Information Visualization, 1,1 (march 2002), 5-12. doi:10.1007/3-540-45650-3_4.

[22] Spoerri, A. 2004. RankSpiral: Toward Enhancing Search Results Visualizations. In Posters Compendium, IEEE Symposium on Information Visualization (Austin, Texas, USA, Oct. 10-12 2004) InfoVis'04. ACM Press. 39-40. doi:10.1109/INFVIS.2004.56.

[23] Tho, Q.T., Hui, S.C., and Fong, A.C.M. 2007. A Citation Based Document Retrieval System for Finding Research Expertise, Information Processing and Management, 43, 1 (January 2007), 248-264. doi:10.1016/j.ipm.2006.05.015.

[24] Turtle Graphics http://www.gkrueger.com/java/aufgaben/loesung/TurtleGraphics.java

[25] Yimam, D. 1999. Expert Finding Systems for Organizations: Domain Analysis and the DEMOIR Approach. In ECSCW 99 Beyond Knowledge Management: Management Expertise Workshop (Copenhagen, Denmark, Sep. 12-16 1999). 276-283.