Journal of Universal Computer Science, vol. 17, no. 11 (2011), 1529-1549
submitted: 17/9/10, accepted: 5/4/11, appeared: 1/7/11 © J.UCS

Expertise Recommender System for Scientific Community

Muhammad Tanvir Afzal
(Centre for Distributed and Semantic Computing
Mohammad Ali Jinnah University
Islamabad, Pakistan
mafzal@jinnah.edu.pk)

Hermann Maurer
(Institute for Information Systems and Computer Media
Graz University of Technology
Graz, Austria
hmaurer@iicm.edu)

Abstract: Finding experts in academics as well as in enterprises is an important practical problem. Both manual and automated approaches are employed and have their own pros and cons. On one hand, the manual approaches need extensive human efforts but the quality of data is good, on the other hand, the automated approaches normally do not need human efforts but the quality of service is not as good as in the manual approaches. Furthermore, the automated approaches normally use only one metric to measure the expertise of an individual. For example, for finding experts in academia, the number of publications of an individual is used to discover and rank experts. This paper illustrates both manual and automated approaches for finding experts and subsequently proposes and implements an automated approach for measuring expertise profile in academia. The proposed approach incorporates multiple metrics for measuring an overall expertise level. To visualize a rank list of experts, an extended hyperbolic visualization technique is proposed and implemented. Furthermore, the discovered experts are pushed to users based on their local context. The research has been implemented for Journal of Universal Computer Science (J. UCS) and is available online for the users of J. UCS.

Keywords: Expertise Finding, Hyperbolic Visualization, Digital Journals, Multi-faceted expert Profile, Spiral Visualization.
Categories: H.3.3, H.5.2

1 Introduction

The discovery of expertise is crucial in supporting a number of tasks. Finding appropriate experts is a key to unprecedented success in enterprises as well as in academia. Finding an appropriate expert is very helpful when one needs guidance on a subject matter, or needs to fill a vacancy based on relevant expertise, or needs to boost the overall productivity especially in enterprises, or needs to find research collaborators working in similar areas, or needs to find editors/reviewers in peer-review setting etc. Therefore, the expertise finding systems can increase overall productivity and can decrease critical delays due to ineffective work. There are different application areas like Software Engineering [Mockus and Herbsleb, 2002],

Enterprise [Balog et al., 2006], Medicine [Sun and Giles, 2007] and Research [Liu and Dew, 2004] which employ various techniques to find appropriate experts using both manual and automated approaches.

A variety of tools have been implemented within organizations 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. [Pipek et al., 2002].

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 [Postellon, 2008]. This also highlights the non-trivial nature of expertise mining and the difficulty faced in the disambiguation of individuals. Automated approaches normally use one facet to judge the overall expertise level of an expert. For example, in a discussion forum analyzing only the number of interactions of an individual is used to judge expertise level [Krulwich, 1995].

In the peer-review setting, appropriate and capable reviewers/committee-members/editors are discovered by computing their profiles, usually based on the overall collection of their publications [Cameron, 2007]. However, the publication quantity alone is insufficient to get an overall assessment of expertise. To incorporate the publication quality in the expertise profile, Cameron used the impact factor of publications' venues (journals, conferences etc.) [Cameron, 2007]. However, the impact factor in itself is arguable [Seglen, 1997] [Hecht et al., 1998]. All publications in a high impact venue do not necessarily get high number of citations. The impact factor of a journal is calculated by considering the number of citations received by all publications published in the journal for a typical period of time [Garfield, 1972]. However, Hirsch, a physicist, proposed another metric, the "H-Index", to rank individuals [Hirsch, 2005]. The H-Index of an author is calculated by considering the number of citations received by his/her most cited publications. To be precise, a scholar with an index h means that the author has published at least h papers each of which has been cited by others at least h times. However, this index works fine only for comparing scientists working in the same field because citation conventions differ widely among different fields [Hirsch, 2005]. Therefore, to measure the quality of one's work in the same field, it is better to calculate the number of citations a person receives rather than just considering the impact factors of journals/conferences where the publications were published. In our approach, additionally we incorporate the number of citations received by an author in a particular topic to make an overall assessment of expertise.

We propose an automated technique which incorporates multiple facets in providing a more representative assessment of expertise as explained in [Afzal et al., 2008]. To overcome automation errors during citation mining process as mentioned above. We introduced an innovative citation mining technique [Afzal et al., 2010]. 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 overall weight is further used to rank experts in the respective topic. 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 [Shneiderman, 2002]. 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. The aforementioned technique is useful especially for journal administration to find high profile authors (experts) who can be assigned as editors/reviewers for the respective topics. To facilitate users of J. UCS, the mined experts are further pushed to users by observing users' local context and task at hand. For example, when a user is viewing a paper, he/she will instantly know about assigned editors and highly active experts associated with topics of the paper. This helps users to establish collaborations in their respective area.

Visualizing a rank list of experts is not enough, users would have an option to explore more aspects of experts, for example short biographies, recent publications, contact information, and affiliations of experts. To support this task, editors are linked with their profiles represented in J. UCS. However, the actively emerging experts (potential reviewers) are further linked with FacetedDBLP. The FacetedDBLP provides a search interface of the huge repository of DBLP. The search interface provides different facets like publication years, co-authors, venues (journal/conference/book series etc) for the selected author. By this means, the users not only know about experts in the respective area, but they can also explore other recent publications of experts and their co-authors, indexed in DBLP.

2 Related Work

Expertise finder systems in the past, have been innovatively applied in helping PhD applicants in finding relevant supervisors [Liu and Dew, 2004] and also in identifying peer-reviewers for a conference [Rodriguez and Bollen, 2008]. 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 [Cameron et al., 2007a] employed a manually crafted taxonomy of 100 topics in DBLP [DBLP, 2009] covering the research areas of a small sample of User researchers appearing in DBLP. They proposed the need for automatic taxonomy creation as a key issue in finding experts. Mockus et al., [Mockus and Herbsleb, 2002] 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 [Yimam, 1999] have further shown that a decentralized approach can be applied for information gathering in the construction of expertise profiles. Tho et al., [Tho et al., 2007] 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., [Krulwich and Burkey, 1995] 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 [Card et al., 1999]. To visualize large hierarchal structures, the hyperbolic tree was developed by Xerox [Lamping and Rao, 1996]. 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 [Spoerri, 2004] 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 for mining the expertise for a digital journal [Afzal et al., 2008]. 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 thus incorporated the use of user-defined profiles, “experience atom” (as proposed by [Mockus and Herbsleb, 2002] to indicate fundamental experiential units), reference mining results and a characterization of expert participation as facets of an expert profile. In a comprehensive characterization of expertise, the following measurements have been proposed:

Number of publications: This describes the overall expertise areas of a person. The intensity of expertise, however, can be represented by the extent of publications. The number of publications can also be used to indicate the topic specific expertise intensity of researchers.

Number of citations received: Citations are indicative of the impact of publications and as a result can be applied to reflect the impact of expert.

Extent and proportion of citations within a particular area: This further indicates the actual interest of citing authors and the overall contribution in a specific area.

Expert Profile Records: J.UCS has expert profiles for its 300 members in its editorial board representing the specified area of expertise based on ACM categories.

This input can be useful as a source in identifying a person as an expert in the area. There are however a number of issues to be considered: areas of interest may change and the research area in itself may evolve.

Experience: Other experiential measures of a person can also be applied in representing one's expertise. Measures that can be acquired with regards to the assessment of experiences include: period of publishing in a particular area, list of projects participated in, assessments of mentoring activities, etc. In the current work, we have taken into consideration the publication age factor only.

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.

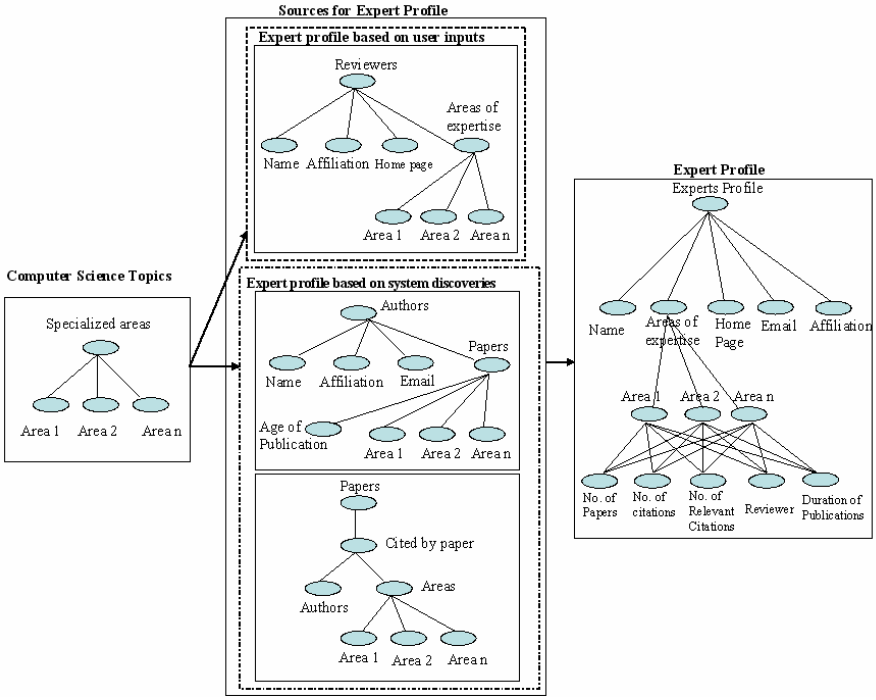

Figure 1: Expert profile.

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. The 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 [ACM-CCS, 1998]. 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, the number of

citations that 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 [Afzal et al., 2007] with a set of over 15,000 references [Afzal et al., 2008].

4 Experts Weights

4.1 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 stored inside the XML file.

Figure 2: A sample paper 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 [Afzal et al., 2010]. The 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.

4.2 Weights assigned to Experts

There are different ways to calculate expertise for different tasks as explained earlier. Our focus is to measure expertise profile in a scientific community, more specifically for finding a program committee or for finding research collaborators. There is no standard and no absolute definition for calculating expertise. The debate for defining suitable scale for overall assessment of expertise is ongoing. Some argue that publication data alone is insufficient to accurately capture expertise [Seglen, 1997] [Hecht et al., 1998]. Others counter that bibliographic data is reasonable as experimental facts support their value [Cameron et al., 2007]. However, the belief that the quantity of publications is proportional to expertise is not universally true. In a very recent and related work, Cameron explained this problem with an example [Cameron, 2007]. He picked two experts in the field of databases in a scenario where one has a long list of publications in the field while the other has only fewer. In this scenario, on one hand, 'E.F. Codd', inventor of the relational database model, and recipient of the ACM Turing Award in 1981 and 1994, has only 49 articles in DBLP, on the other hand, 'Hector Garcia-Molina', an ACM Fellow too, recipient of the ACM SIGMOD Innovations Award in 1999, had 248 publications in DBLP until 2003 (the year of Codd's death). This example highlights a situation in which a researcher having a large list of publications, may by default, be ranked more prolific than his associates having fewer publications, in spite of publication quality. If one considers publication quantity alone as a measure of expertise, the statistics would conclude 'Garcia-Molina', as far more prolific in the field of databases. However, considering magnificent contributions of 'E.F. Codd' to the field, many may regard it astonishing. To measure a better rank of experts, Cameron employed 'publication impact' as an additional measure to incorporate the quality of the published manuscripts.

However, the impact factor in itself is arguable [Seglen, 1997] [Hecht et al., 1998] [PLoS Medicine Editors, 2006]. The impact factor does not work well since a small number of publications are cited much more than the majority of publication in a particular venue. For example, the well known journal Nature has analyzed the citations of individual papers in *Nature* and found that 89% of the impact factor was generated by just 25% of the papers [Nature Editorial, 2005]. Alternatively, if a publication is of great quality then it will receive a reasonable number of citations. Therefore, to rank experts in a field, it is better to calculate the number of citations of all publications of an individual [Hirsch, 2005]. This also applies the above defined scenario. As per Google Scholar database, the most cited paper by 'E.F. Codd' has received 5140 citations as of November 2009 while the most cited paper by 'Hector Garcia-Molina' has received 1408. Therefore, using citations of researchers' publications directly rather than using the impact factor as Hirsch did in calculating

H-Index [Hirsch, 2005] would be better. In our approach, we have applied the number of publications and citations in an innovative way to calculate overall expertise as explained in the next sections. Apart from publications and citations lists, there might be different measures that can be integrated into the overall weight of the experts. For example, one can use the fact that if a person is serving as a reviewer or on editorial board of some journals and conferences.

In our system, experts are grouped into one of two categories: 1) editors (persons currently manually assigned as reviewers for a particular ACM topic) 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 3 are used to assign the weights. The weights used in our system are publication_weight, citation_weight, and editor_weight.

4.2.1 Publication Weight

In a particular research area, the publication weight of an author is obtained by dividing the number of the author's publications by the number of publications' years (duration of publications). To find active experts, we consider the publications of an author that have been published in the last five years. The number of years is calculated from the year of a first publication (within last five years) until the current year. For example, if an author has published four contributions in the last four years then the publication weight of the author would be one. Consequently, for a specific research area, authors having a larger publication weight would get an edge over their counterparts having fewer publications.

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

4.2.2 Citation Weight

The citation weight reflects the author's impact in the growth of a particular research area. For example if all papers in a research area have received 1000 citations collectively and an author's papers in that specific area have received 100 citations, then the citation weight of this author would be 0.1.

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

4.2.3 Editor Weight

The editors' weight is calculated by dividing the number of J.UCS reviewers by the total number of J.UCS authors. This weight is assigned to only those authors who are also working as editor/reviewers. In this way, reviewers (already acclaimed experts) get an edge over the other authors.

Editor Weight = No. J.UCS editors / Total no. of J.UCS Authors.

The total weight is defined as the sum of the above defined weights:

Total weight = publication weight + citation weight + editor weight.

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

4.2.4 Algorithm to Construct experts' Profiles

The algorithm for calculating an expert profile is shown in Figure 3. The algorithm takes: topic, papers, citations, and reviewers as input and returns an expert profile for all topics.

5 Information Visualization

Two different visualizations were developed based on measured expertise, one for the journal administration and the other for users of this journal. The visualization for the journal administration is based on the assumption that all topics should be visible in one place where one can easily navigate to a particular topic and can see editors and potential experts belonging to that topic. To make it user-friendly, we have chosen a hyperbolic browser which is based on "focus+context" technique [Lamping and Rao, 1994] [Lamping et al., 1995] [Lamping and Rao, 1996]. The hyperbolic browser was further extended with a spiral representation of potentially ranked experts. This makes the job of administrator to focus on any particular topic while the overall context remains there. The details of hyperbolic visualization can be found in the next section. The second visualization was developed for users of the journal. This visualization is based on the assumption that the user should have an access to experts whenever he needs them. For the current implementation, when a user is looking on a particular paper and clicks 'Links into the Future', then he/she is shown active experts associated with the topics of the focused paper along with the similar papers written in the same area. The details of 'Links into the Future' can be found in [Afzal et al., 2007]. The remaining parts of this section explain both of the aforementioned visualizations.

5.1 Extended Hyperbolic 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) was calculated. The hyperbolic browser [Lamping and Rao, 1994] [Lamping et al., 1995] [Lamping and Rao, 1996] 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 [McCahill and Erickson, 1995], PRISE [Cugini et al., 1996], and RankSpiral [Spoerri, 2004] which both use spiral representations to display ranked search result lists.

```
Algorithm calculaeExpertiseProfile(Topics)
1.  create 'empty expertise profile'
2.  for each 'topic' do
3.      get 'papers' written in the last 5 years.
4.      get 'topic_cit_count' (get citation counts for all papers in a topic)
5.      get 'authors'
6.      for each author do
7.          get 'pub_count' (publication count of an author ).
8.          get 'pub_duration' (publication duration is defined as: current
            year - year of the first publication (within last 5 yrs) + 1).
9.          Get 'cit_count' (citation count of an author).
10.         publication_weight = pub_count / pub_duration.
11.         citation_weight    = cit_count / topic_cit_count.
12.         if 'author' is also a 'editor' do
13.             get 'editors_count' (number of editors in a topic).
14.             get 'authors_count' (number of authors in a topic)
15.             editor_weight = editors_count / authors_count
16.         else
17.             weight 3 = 0
18.         end
19.         weight = publication_weight + citation_weight + editor_weight
20.         add <'topic', 'author', 'weight'> to 'expert profile'
21.     end
22. end
    return 'expert profile'
```

Figure 3: Algorithm for Computing Expertise Profile

The user interface is shown in Figure 4. This is implemented in Java. A hyperbolic browser is used to visualize the ACM classification hierarchy, using the freely available Hypertree package [Hyperbolic Package, 2009]. Both categories of experts are visualized by superimposing upon the hyperbolic view. Reviewers are

shown in a simple list and high profile authors are shown in spiral visualization. To draw the spiral, a package called Turtle Graphics is used [Turtle Graphics, 2009]. 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.

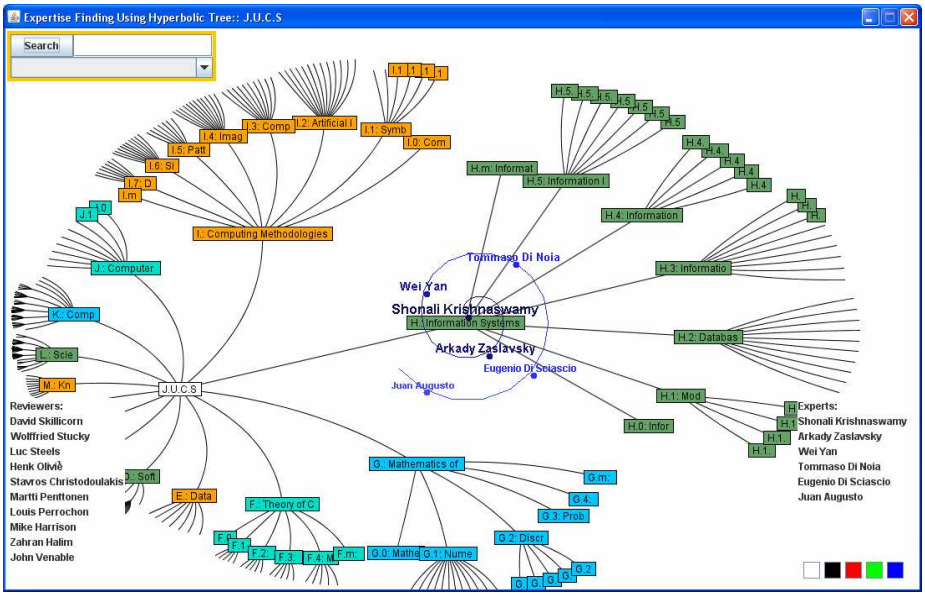

Figure 4: Hyperbolic Visualization

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 4. 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 coloured buttons. When clicked, the spiral is redrawn with the new colour. 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 a user drags a particular node, the spiral moves with the focused node.

Figure 4 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 centre of the spiral).

This visualization is useful for journal administering. For example, in J.UCS there are some 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 example, the topic 'M.8 Knowledge Reuse' has no reviewers at the

moment (this is a new topic added by J.UCS). Potential reviewers are easily found in the visualization, as shown in Figure 5. This type of discovery is very useful for administrators to locate potential reviewers for any selected area.

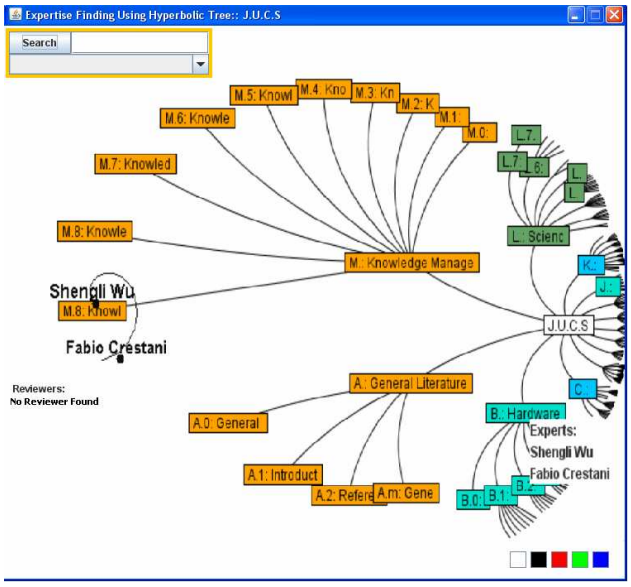

Figure 5: Discovery of Potential Reviewers

Table 1 shows a case of one author "Hermann Maurer". The author is already a reviewer for ACM topics: A., H.5.1, K.3, and K.4, in addition, he can be considered as a reviewer for ACM topics: H.1, H.3, H.4, H.5 based on author's contributions to these ACM topics.

Name: Hermann Maurer		**Total publications in J.UCS** = 29		
Already reviewer for:	A. (7%)	H.5.1 (10%)	K.3 (21%)	K.4 (21%)
Can be considered for:	H.1 (14%)	H.3 (17%)	H.4 (24%)	H.5 (14%)

Table 1: A comparison between user defined and system discovered expertise

Although it is convenient to explore the topical hierarchy with the hyperbolic tree, users sometimes know the name of a topic and want to navigate directly to it. The search facility in the top left corner of the main interface (see Figure 6) supports this task. For example, if a user searches for the term "Information", then a combo box is filled with all topics containing the term "Information" as a substring. The 13 topics containing the term "Information" are shown in Figure 6. The user can select any ACM topic from the search result list and the hyperbolic tree is redrawn to show the selected topic centred in the window.

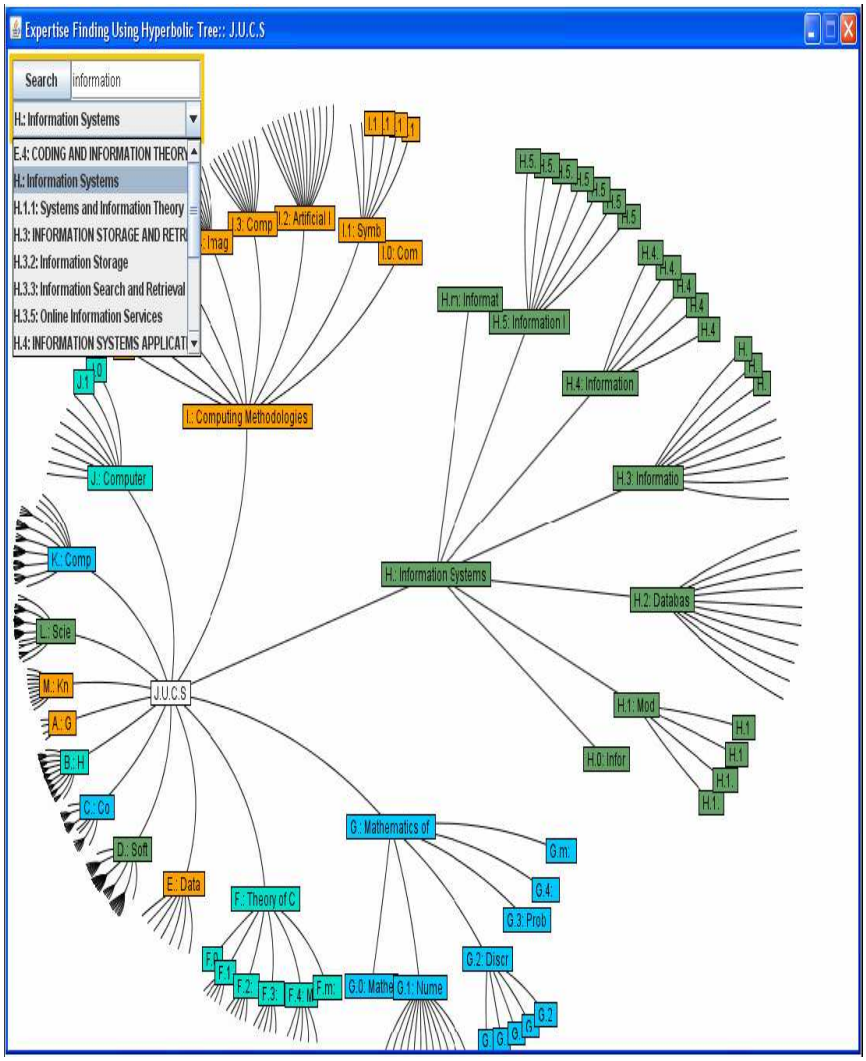

Figure 6: ACM Topic Search Facility

5.2 Visualization for Users of J. UCS

The measured experts for topics of the paper are pushed to users by looking at user's local context. For example a user is viewing a paper titled 'The Transformation of the Web: How Emerging Communities Shape the Information we Consume' as shown in Figure 7.

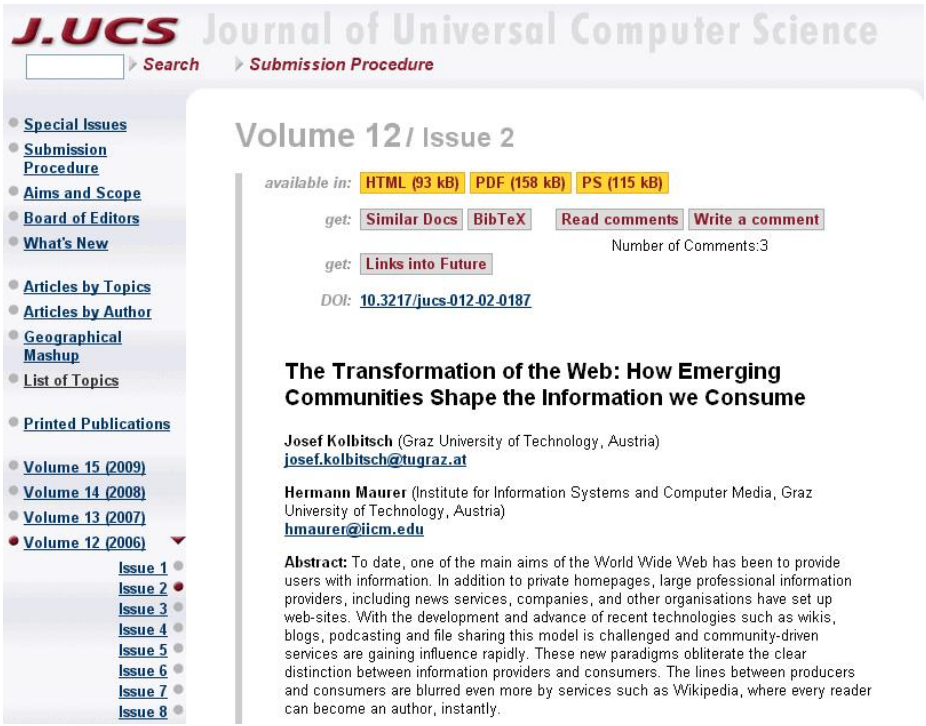

Figure 7: J. UCS interface for viewing a paper

When the user clicks on the 'Links into Future' button, he is redirected to a screen as shown in Figure 8. This was implemented using a java servelet. The servelet receives a reference of the viewing paper as a parameter. The servelet subsequently fetch data from different database tables. On the top of this visualization, the focused paper and its metadata are shown. The lower part of the screen is divided into two columns, the left part is dedicated to visualize 'Links into the Future' i.e. related papers written in the same area in future dates as compared to the publication date of the focused paper as explained in [Afzal et al., 2007], while the right part of the screen visualizes the experts associated with the topics of the focused paper. As already mentioned, experts are categorized into two categories: 1) the editors (reviewers) assigned by the editor-in-chief, and 2) the potential experts flagged by the system. Both categories of experts are shown in this visualization. To find more information about experts, the experts are further linked within J. UCS and with FacetedDBLP [Diederich et al., 2007]. The current section gives details about reviewers' linkage within J.UCS while section 5.2.2 explains how discovered experts are linked with FacetedDBLP.

Figure 8: Visualization of experts

There are more than 300 editors serving as reviewers for J. UCS. There is a many-to-many association between editors and topics. Every editor is usually assigned to multiple topics and each topic is assigned to multiple editors. According to current statistics, there are 45 topics in J. UCS which have more than 10 editors associated. If we visualize all editors for all topics of the paper at one place, then it would become a problem to locate required information for users. To avoid such a situation, an indirect way was used. Initially, the topics of the paper are visualized as shown in Figure 9. The user can follow any topic to look for all associated editors and published papers in the focused topic. For example from Figure 9, a user is interested in finding editors of 'H.3.5: Online Information Services'. On click, the user is redirected to the screen as shown in Figure 9.

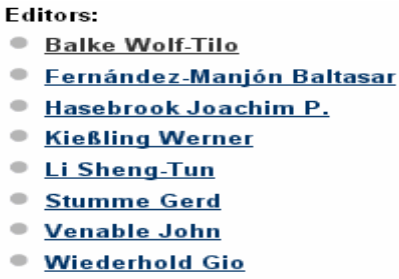

Figure 9: Editors associated with topicH.3.5

5.2.1 Linking Editors Profiles in J. UCS

The information about editors is maintained by the J. UCS administration in a highly structured way. This information normally includes: short biography, assigned topics for review, institution, address, email and homepage of editors. For example when a user clicks on "Balke Wolf-Tilo" in Figure 9 to view the details, he is redirected to the screen as shown in Figure10. The user is then able to read a brief biography of the expert and can follow to expert's homepage for recent contributions in the area.

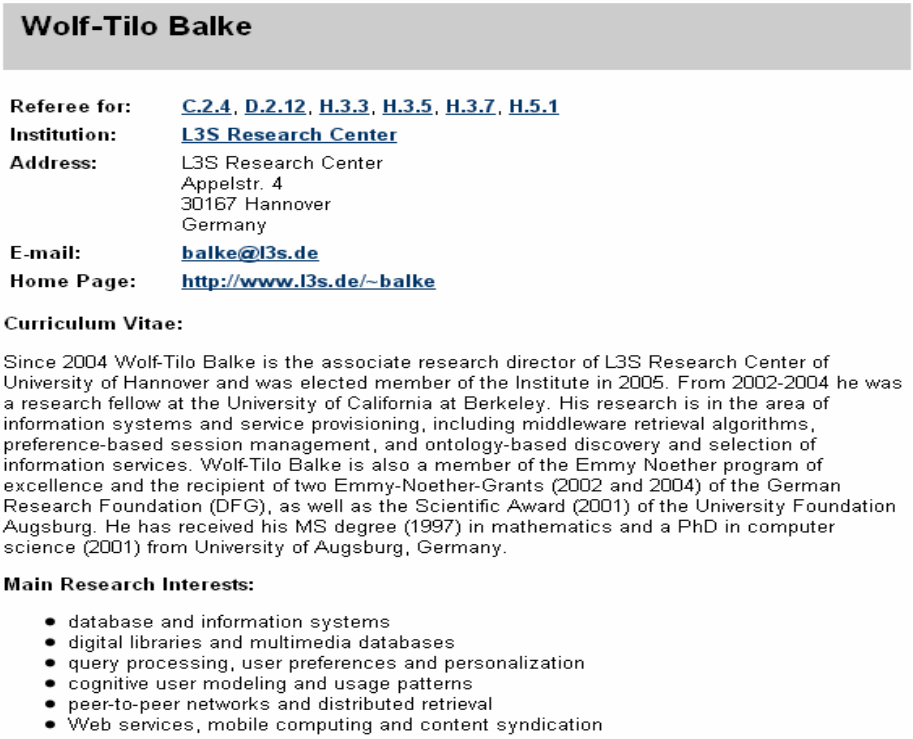

Figure 10: Editor profile

5.2.2 Linking Discovered Experts' Profiles to FacetedDBLP

The potential experts (discovered by the system) are shown after the aforementioned visualization as can be seen in Figure 8. Only the active research areas (having contributions in the last five years) and their experts are visualized. For example the focused article in Figure 8 belongs to five topics and all of them remain active research areas in the last five years in J. UCS. The top 10 ranked experts are visualized for each topic of the paper. To gain deeper insights into the experts' contributions, these experts are further linked with FacetedDBLP [Diederich et al., 2007]. This FacetedDBLP is build upon the large collection of DBLP dataset.

For example a user clicks on the author's name "Jong Hyuk Park" (topic H.5.1) in Figure 8. The user is redirected to the screen as shown in Figure 11. This was achieved by querying FacetedDBLP http://dblp.l3s.de by adjusting author's first, middle, and last names using some heuristics. The FacetedDBLP is based on the large repository of DBLP (DBLP currently index more than 1.3 million computer science publications). In the figure, there are 65 publications of the author "Jong Hyuk Park" found in FacetedDBLP. A user can search using different facets as shown on the left side of the figure like: publication years, publication type (article, proceedings, etc), venues (journals, conferences etc), authors, and the GrowBag graph. Based on the user selection, search results are shown on the right side of the figure. For example a user can search all papers of the focused author which were co-authored with any of the authors shown on the left side. The user can restrict the search results to find papers which appeared only in any of the venues (like: computer communication, The Journal of Supercomputing, etc). The user can characterize the result set in terms of the main research topics and filter it according to certain subtopics. The GrowBag terms may be very useful for the user. For example a user can restrict the result set to see only papers of the focused author which deal with any of the shown GrowBag terms (like security, pervasive computing, privacy protection etc). Therefore, a user may find required information more efficiently and accurately using this interface and instantly becomes aware of the research areas of the authors, his collaborators list, the venues where the author has published, etc.

Figure 11: Adapted from FacetedDBLP

6 Conclusion and Future Work

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, however, it can easily be 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.
- A visualization of experts for users of J. UCS which is further linked with expert's profiles within J. UCS and in FacetedDBLP.

The discovered experts in this research are based on their contributions within J. UCS. This does not reflect the overall ranking of experts for an area. In future, we will extend the system to include other datasets such as DBLP and CiteSeer to find experts. The incorporation of these datasets will provide overall better ranking of experts for a focused area in Computer Science.

References

[Afzal et al., 2007] Afzal, M. T., Kulathuramaiyer, N., and Maurer, H.: Creating Links into the Future, published in Journal of Universal Computer Science, vol. 13 issue. 9, 2007.

[Afzal et al., 2008] Afzal, M. T., Kulathuramaiyer, N., Maurer, H. (2008). Expertise Finding for an Electronic Journal, In: Proceedings of International Conference on Knowledge Management and Knowledge Technologies, pp. 436-440, Graz, Austria, 3-5, Sep. 2008.

[Afzal et al., 2010] Afzal, M. T., Balke, W. T., Kulathuramaiyer, N., Maurer, H.: Rule based Autonomous Citation Mining with TIERL, Journal of Digital Information Management, 8 (3), 196-204. June, 2010.

[Balog et al., 2006] Balog, K., Azzopardi, L., de Rijke, M.: Formal Models for Expert Finding in Enterprise Corpora. Special Interest Group on Information Retrieval, Seattle, pp. 43-50. Washington, 6-11. Aug, 2006.

[Garfield, 1972] Garfield. E.: Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. Science, 178, 471-479. 1972.

[Cameron et al., 2007] Cameron, D., Aleman-Meza, B. Decker, S. L., Arpinar, I. B.: SEMEF: A Taxonomy based Discovery of Experts, Expertise and Collaboration Networks. University of Georgia, LSDIS Lab, Technical Report, July. 2007.

[Cameron, 2007] Cameron, D. L.: SEMEF: A Taxonomy-based Discovery of Experts, Expertise and Collaboration Networks. MS thesis, The University of Georgia, ATHENS, GA. 2007.

[Cameron et al., 2007a] Cameron, D., Aleman-Meza, B., Arpinar, I.B.: Collecting Expertise of Researchers for Finding for Relevant Experts in Peer-Review Setting. In: Proceeding of 1st International Expert Finder Workshop, Berlin, Germany, 16, January. 2007.

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

[Cugini et al., 1996] Cugini, J. V., Piatko, C. D., Laskowski, S. J.: 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. 1996.

[DBLP, 2009]Digital Bibliography and Library Project, http://www.informatik.uni-trier.de/~ley/db/index.html.

[Diederich et al., 2007] Diederich, J., Balke, W-T., Thaden, U.: Demonstrating the Semantic GrowBag: Automatically Creating Topic Facets for FacetedDBLP, In: Proceedings of JCDL'07, 17–22, Jun. 2007, Vancouver, British Columbia, Canada. 2007.

[Hecht et al., 1998] Hecht, F., Hecht, B. K., Sandberg, A.A.: The Journal Impact Factor: A Misnamed, Misleading, Misused Measure, Cancer Genet Cytogenet 04: 77-71 (1998), Elsevier Science Inc. 1998.

[Hirsch, 2005] Hirsch, J. E.: An index to quantify an individual's scientific research output. PNAS 102 (46): 16569–16572. doi:10.1073/pnas.0507655102. 2005.

[Hyperbolic Package, 2009] Hyperbolic Tree Library,http://hypertree.cvs.sourceforge.net/viewvc/hypertree/hypertree/

[Krulwich and Burkey, 1995] Krulwich, B., Burkey, C.: 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. 1995.

[Krulwich, 1995] Krulwich, B., Burkey, C.: 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. 1995.

[Lamping and Rao, 1994] Lamping, J., Rao, R.: 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. 1994.

[Lamping and Rao, 1996] Lamping, J., Rao, R.: 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. 1996.

[Lamping et al., 1995] Lamping, J., Rao, R., Pirolli, P.: 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. 1995.

[Liu and Dew, 2004] Liu, P., Dew, P.: Using Semantic Web Technologies to Improve Expertise Matching within Academia. In: Proceedings of the 2nd International Conference on Knowledge Management, pp. 370-378. Graz, Austria. June 30-July 2, 2004.

[McCahill and Erickson, 1995] McCahill, M. P., Erickson, T.: 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. 1995.

[Mockus and Herbsleb, 2002] Mockus, A., Herbsleb J. A.: Expertise Browser: A Quantitative Approach to Identifying Expertise. In: Proceedings of International Conference on Software Engineering, pp. 503-512. Orlando, Florida, 19-25, May. 2002.

[Nature Editorial, 2005] Nature Editorial.: Not-so-deep impact, Nature 435, 1003-1004 (23 June 2005) | doi:10.1038/4351003b; Published online 22, June. 2005.

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

[PLoS Medicine Editors, 2006] The PLoS Medicine Editors. The impact factor game. It is time to fi nd a better way to assess the scientifi c literature. PLoS Med. 3:e291. doi:10.1371/journal.pmed.0030291. 2006.

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

[Seglen, 1997] Seglen, P. O.: Why the impact factor of journals should not be used for evaluating research", BMJ 1997; 314(7079):497. 1997.

[Spoerri, 2004] Spoerri, A.: 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. 2004.

[Sun and Giles, 2007] Sun Y., Giles C. L.: Popularity Weighted Ranking for Academic Digital Libraries, In: Proceedings of 29th European Conference on Information Retrieval Research, pp. 605-612. Rome, Italy, 5-7, April. 2007.

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

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

[Tho et al., 2007] Tho, Q.T., Hui, S.C., Fong, A.C.M.: 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. 2007.

[Turtle Graphics 2009] Turtle Graphics,
http://www.gkrueger.com/java/aufgaben/loesung/TurtleGraphics.java

[Yimam, 1999] Yimam, D.: 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. 2007.