Automatic Support in the Review Process

Harald Krottmaier
Institute for Information Processing and Computer
Supported new Media (IICM)
Inffeldgasse 16c, A-8010 Graz, Austria
hkrott@iicm.edu

Abstract

One of the most complicated problems of peer-reviewed journals, regarding the quality of publications, is the right selection of referees for submitted papers. In computer science related journals referees express their interest and main research areas in terms of the ACM classification system. Interests change over time and therefore the referee has to update the “table of interest” from time to time. This paper will give an overview of the current refereeing situation in the Journal of Universal Computer Science (J.UCS) where referees are able to select papers for refereeing from a broad range of topics. In the future, we want to narrow the range of potential papers per referee. We investigate a new approach for the selection of appropriate referees assuming that referees’ recently published papers express their current main interests.

1. Introduction

The reputation of peer-reviewed journals is very much dependent on the reputation of the referees of the editorial board. Referees are usually invited by the editor(s)-in-chief to join the editorial board of the journal and then usually sign-up for a topical category of papers. In computer science theses categories are often expressed by the ACM Computing Classification System (ACM-CCS, [ACM, 1998]).

Submitted papers should be reviewed by the most relevant referees to guarantee steady quality of published papers. In this paper we are going to illustrate the current situation related to the communication between the referees and the editorial team in the Journal of Universal Computer Science ([Calude et al., 1994], [Maurer and Schmaranz, 1994]). J.UCS is a peer reviewed journal and was founded in 1994 by Springer Pub. Co. ([Springer Pub. Co., 2002]) in cooperation with the Institute for Information Processing and Computer Supported new Media ([IICM, 2002]). Graz University of Technology in 1994. As of 2001 J.UCS is a publication of the Know-Center ([Know-Center, 2002]) in cooperation with Springer Co. Pub., JOANNEUM RESEARCH ([JOANNEUM RESEARCH, 2002]) and the IICM.

We introduce ideas to improve the communication and reduce the workload for both referees and the editorial team. Thereafter we show some further applications of the presented ideas.

2. Current Situation and Improvements

Authors submit their papers via email or via web-server application in the appropriate electronic format. The editorial team then extracts the abstract out of the documents and sends the abstracts to the referees by email. If at least 3 referees sign-up for the paper, the paper is accepted for refereeing, otherwise the submission is withdrawn and the authors are free to submit their papers to other journals.

There are several drawbacks in this procedure and much effort in communication is spent by the editorial team and the referees as well. The editorial team has to extract the abstracts and send emails to the referees on a regular basis. Referees must sign-up for reviews based on abstracts. Then the editorial team must collect the reviews of the papers and inform the authors about the review.

It is obvious that a web-application may solve some of the problems concerning this email-based communication. A simple first approach, currently under development, is to sign-up specific referees for particular topics described in terms of the ACM-CCS. The application then decides based on the author given categories or keywords to which referees the abstracts or a notification should be sent. Although the authors are requested to state the appropriate keywords and categories, many authors do not supply us with this information. A simple submission form with mandatory fields may also solve this problem. Many conference supporting systems work like the approach described above.
In the current submission and refereeing procedure we have to deal with two problems and have to solve them with modern knowledge management features:

**Submitted Papers** do not follow formatting guidelines and do not provide the right information.

**Referees** do not update their area of interest and will therefore receive unwanted emails.

In the following sections we are going to illustrate solutions to these two problems.

### 3. Improvements in the Paper Submission Process

As described above papers are submitted to J.UCS by either email or via web-form (see figure 1). At the moment communication with the corresponding author is done exclusively via email with the editorial team. To reduce the communication workload for the editorial team we are going to assign a username and password at a first submission of an author’s paper. Thereafter the author will be automatically informed about any change in the status of the submitted article, i.e. the author knows exactly when the review process starts and when to expect results.

**Figure 1. Current Paper Submission**

Due to the anonymous review of the papers, authors do not know which referees are going to review or have reviewed their papers. The review form is partitioned in author-visible and editor-visible parts. A highly sophisticated access system with integrated user- and group-management is responsible for displaying the appropriate parts of the review to the author.

Some meta data (especially about authors information) is collected by the form of the paper upload. The following data is collected in the web-form: 1.) title of the paper, 2.) the file (either PDF or PostScript) itself, 3.) information about the author like first/last name, institution, country and email. Additional meta data is required for the distribution to the referees: 1.) abstract, 2.) keywords and 3.) categories.

Automatic extraction of the abstract is quite easy since authors usually do provide an abstract: It starts with the word “Abstract” and ends either by a paragraph starting with “Key Words” or “Category”, or the number and name of the first section.

Unfortunately, authors submit papers with no keyword- and no category-field. A simple solution to this problem would be to let the author select the keywords and/or categories by hand from a given list (or via a simple text field). This would be easily possible because articles are categorized in terms of the ACM-CCS. For the current prototype it is planned to provide the author with these additional mandatory fields. After a short evaluation with some selected authors the prototype will be available to all authors. Many other publishing systems and conference supporting systems require the same information from authors uploading a paper. This additional workload is distributed to the authors, therefore the workload for the editorial team is reduced without losing quality in the whole workflow.

After the successful upload of a paper tools check the given abstract, keywords and categories with the automatically extracted entities. In Hyperwave (HIS, [Maurer, 1996]) some knowledge management tools like Autonomie’s automatic summary are already integrated and available to the developer of a web-application. Automatic clustering is also available in the used publishing system. Thereby it is also possible to automatically categorize the submitted article in an ACM-CCS-style collection hierarchy.

Automatically extracted parts of the document in combination with the given parts of the document (such as abstract, keywords and categories) create a helpful profile of the article used as base for the review process.

### 4. Search for the Right Referee

After some helpful information is extracted from the submitted papers the most appropriate referees must be found. At the moment each referee must decide on the basis of title, abstract, keywords and category of the paper which paper to review and which not to review. While this process guarantees that a “self-assigned referee” is most appropriate for this topic, it also means that every referee receives every submitted abstract by email. This fact increases the workload for the referees. A preselection of potential referees may improve the reviewing process. In the following we illustrate the current available information of referees. Thereafter we sketch how to improve the dissemination of submitted papers more effectively.
4.1. Currently Available Information

We have already stored some information about referees in our database system. Such an entry is shown in figure 2. Referees express their interests by simply registering a category according to the ACM-CCS and usually also mentioning some keywords of their current main research interests. Links to their personal homepage are provided on this information page. With the currently available information it would be possible to distribute the abstracts of new submissions to the corresponding most related referees. This procedure will reduce the workload of the referees in the beginning, but there is an enormous drawback to this approach: later on, referees do not update their information page and do not provide us with information about changes in the main research area! Therefore the situation will be very unsatisfactory in a few years when referees are asked to review papers related to old research interests. The quality of publications will decrease over time because there are massive changes in computer science related topics.

Figure 2. Information about a Referee

Obviously there are two solutions to this problem: first, referees are actively asked to update their database record once a year or whenever there is a change in the research area, second, the publishing system tries to explore necessary information automatically without any additional human interaction. Unfortunately people will not update database records (even on request!) therefore we decided to explore and implement the automatic exploration approach.

On the assumption that publications of the referees express their main research areas we have based our ideas on improving the proper dissemination of abstracts. After the publications are found, the system performs a similarity search of submitted articles and recent publications of the referees. We think that we are able to assure the quality of publication on a long term without any additional human interaction!

4.2. Search for Publications of Referees

In the following you find some thoughts of possible solutions to the problem of finding the publications of our referees. We have explored some of the freely available services.

To enable similarity search our goal is to find at least the abstract of papers written by a particular referee. Several available search-engines index a lot of material stored all over the world. The most powerful search-engine at the time of writing is the Google search engine with more than $2 \times 10^{9}$-indexed web pages. An application program interface (API) for Google using SOAP (Simple Object Access Protocol) is available for different programming languages. The problem of formulating the right search query still exists and there are some problems in extracting the right entities from the result. Therefore Google is unfortunately not suitable for our task in searching the publications of referees.

Many referees do supply us with the address of a personal homepage where a list of publications is very often provided. Often authors are allowed to make at least the abstract available at their home page after signing a copyright transfer statement with the publisher of the paper. Older papers (usually papers older than one year) are often available in fulltext on the homepage. Nevertheless, homepages differ in many respects (structure as well as format) and an automatic extraction of the publications of an author is not easily possible. There are still some referees without a homepage or without a list of publication page linked from the homepage. It is conspicuous that information extraction from homepages is difficult and unmanageable in a long perspective.

Unstructured data is difficult to put into the right context therefore we were looking for structured, freely available material about publications. For computer science related material a service located at University of Trier ([Ley, 2002]) is very appropriate. The data is stored in an XML formatted file and is accessible via a high sophisticated search interface. Hyperlinks to online available electronic resources are provided. It is easy to search using first and last name of an author and get back a list of known publications of this author including some meta data (year, type, title etc.). We think that this service is a good start point for further explorations of publications of an author.

Another well-known service for the search of publications is NEC Research Institute’s ResearchIndex ([NEC Research Institute, 2002]). Additionally to information offered by the system from University of Trier, Re-
searchIndex offers a citation index and caches also the publication itself. Since we are going to explore this system also in another project, we decided to use this system as a source of publication-gathering from our referees as well.

With these two systems we are able to gather at least some of the publications from our referees and are able to extract either the fulltext or the abstract of the publication. We are going to write adapters to make this information available to our system. Additionally we are going to encourage our referees to supply the system with abstracts of their recently published papers.

At the moment we are developing tools for gathering information about our referees concerning publications and are also developing a web-application to enable upload of abstracts and additional information about publications of our referees. We are going to use web-data-extraction tools (a very nice overview of some of these tools is given in [Laender et al., 2002]) as well as simple text extraction tools to extract information out of web resources. Each publication will be represented in the Hyperwave Information Server (HIS) as an entity consisting of a collection holding all the necessary meta data (title, year, keywords etc.) and the abstract of the publication. This collection structure makes it possible to add future developments like fulltext of the publication or citations of the entity. We are aware of many problems (e.g. author’s name representation differ in format: one system supply the full information i.e. first-, middle- and lastname, other systems just supply the first letter of the firstname and the lastname...) already solved in the mentioned systems. The implementation of the prototype and the upcoming feedback and evaluation procedure will demonstrate if the described approach for gathering information about publications is applicable.

4.3. Similarity Search

Once the publications of the referees are stored in our database, searching for similar articles is just “a mouse click away” and very easy to handle in the used publishing server system. We use a highly sophisticated knowledge management systems (Hyperwave Information Server, HIS) where features such as “find similar articles” or “find a category for an article” are already integrated. Two systems are available at the moment: [Verity, 2002] and [Autonomy, 2002]. These systems support the programmer with slightly different parameters and implementations. Nevertheless, getting a similar document is very easy to implement.

Since similarity search may also be scoped (i.e. the search scope may be restricted to a specific part of the server) there is an enormous potential for this function. It is possible to limit the search scope to all articles published last year. If no appropriate referees are found then the search scope may be extended to a larger time-scope. If there is a lack of potential referees for a type of topic then the editorial board must be extended etc.

Search results may be sorted in different ways to guarantee quality in the review of articles. To give some examples: if the referee is the only author of the article and the article was written just a few months ago, then this referee is very suitable to review this new article. If, on the other hand, the referee is one of several co-authors and the article was published a long time ago, then the referee may not be the appropriate person to review the selected topic. As all reviewing activities of the referees are logged in the system, the activity of the referee may also be considered when selecting the referee for reviewing an article. Articles published in journals may be endowed with a higher priority than those published in conference proceedings. Further parameters including length of the article, additional ratings of other users of the article, relevance to the community represented by a citation index parameter etc. may be taken into account.

It has been shown that there are many parameters to adapt to find the right referee for a submitted paper. After an evaluation period of the system feedback from the referees will be evaluated to improve the performance of the system.

5. Further Applications

Many other applications are possible with the shown approach of information gathering and searching for similar documents. In every situation where experts about a topic must be found, this approach may be a good choice to improve the selection of experts.

Conference management systems (e.g. ConfMan, [Halvorsen et al., 1999], used at EurAsia 2002) support conference organizers in the reviewing process of submissions. Submitted papers are uploaded to the server system and meta data is added by the authors via web form. Additional information about a submission (like keywords) are added by the authors themselves. Members of the program committee sign-up for a review and return the results via email or web-form. These systems may be extended by the described ideas: automatic keyword extraction from the submitted papers will guarantee appropriate keywords, the most appropriate referee of a submission may be found by simply do a similarity search between the content of the submission and the abstracts or fulltexts of articles written by the referees. The number of referees per submission may be reduced and therefore time is saved without losing quality of reviews.

Refereeing of grant proposals may be improved by automatically selecting the appropriate referee for the proposal. Whenever an expert or expert group of a number of given persons must be found the described approach may
improve the current situation. This idea can be extended to introduce a support system for writing scientific papers ([Krottmaier, 2002]). Experts about a topic are found automatically and the system may generate an automatic reference section for publications in preparation. It is and will not be possible to explore and gather information about all experts. Nevertheless, it is planned that at least a selection of experts who published at least one paper in a well-known, reputable journal are considered to be included in the data base.

6. Conclusion and Future Work

Recent publications of authors express their current research interests in much greater details than information provided by the authors like unstructured formatted lists of keywords and/or ACM categories. Interests change over time and many experts do not update information given on their homepage.

It has been shown that there are several applications where experts must be found. An application is the assignment of a number of referees to submitted papers of a journal. Steady quality of publications and a minimum amount of time for the referees and editor(s)-in-chief are the main advantages of the described system. Several problems when finding publications of an author are already solved in other systems ([Lawrence et al., 1999] or [Ley, 2002]) therefore the system must not deal with these problems.

Future work is addressed to implement some of the ideas presented in this paper. Gathering of information about a referee will be implemented using the two mentioned systems. Contact to the operator of one system is already established. Other technologies of gathering information (like intelligent agent technology) are currently explored and compared to the described approaches. If publications are found, information must be extracted and the appropriate data must be inserted into the database system. Usability studies and feedback will guarantee a fast and easy to use application.

References


Acknowledgment

I like to thank Prof. Hermann Maurer for the fruitful discussions and Dana Kaiser for giving me insights to the work of the editorial team.