

Summary of the 2nd meeting of the

DBLP Advisory Board

held on March 2nd, 2012,

in Mannheim, Germany



SCHLOSS DAGSTUHL

Leibniz-Zentrum für Informatik

In attendance:

- Andreas Butz (LMU Munich)
- Claudius Korzen (Univ. of Freiburg)
- Dietmar Saupe (Univ. of Konstanz)
- Hannah Bast (Univ. of Freiburg)
- Hans-Peter Lenhof (Saarland Univ.)
- Jürgen Teich (Univ. of Erlangen-Nuremberg)
- Marc Herbstritt (Schloss Dagstuhl LZI)
- Marcel R. Ackermann (Schloss Dagstuhl LZI, DBLP)
- Michael Ley (Univ. of Trier, DBLP)
- Michael Wagner (Schloss Dagstuhl LZI, DBLP)
- Mila Majster-Cederbaum (Univ. of Mannheim)
- Otto Spaniol (RWTH Aachen)
- Reinhard Wilhelm (Schloss Dagstuhl LZI, Saarland Univ.)
- Rüdiger Reischuk (Univ. of Lübeck)

Absent (excused):

- Oliver Günther (Univ. of Potsdam)
- Rüdiger Dillmann (Karlsruhe Inst. of Tech.)

Agenda

1. Opening remarks and DBLP progress report
2. Publications in computer science
3. Scope of DBLP
4. Bibliometrics
5. Perspectives and further development
6. Appointment of next meeting

Meeting begins: 2:00 p.m.

Item 1: Opening remarks and DBLP progress report

Hannah Bast welcomes the board members and opens the meeting. She introduces Michael Wagner, who joined the DBLP Team in January, and Claudius Korzen, who just started as a PhD student in Freiburg and who will collaborate with the DBLP Team.

Marcel R. Ackermann gives a brief overview on the recent progress of DBLP (c.f. the accompanying slides). The key developments are:

- Redesign of the DBLP web front end
- Development of an error report portal (with Florian Reitz, Univ. of Trier)
- Introduction of a workflow management infrastructure (JIRA)
- Testing of bibliometric polls at Dagstuhl Seminars
- Preparations for a bibliometric colloquium/workshop
- DBLP server in Schloss Dagstuhl
- Possible collaboration with Zentralblatt MATH
- New milestone: more than 1.9 million publications in DBLP

Item 2: Publications in computer science

Mila Majster-Cederbaum reports on the results of the Publications subcommittee. For the inclusion of new publication streams to DBLP, the installation of a formal review process is proposed. A submission form should query editors or steering committee members for the details of the stream. Using this information, the Advisory Board can come to a decision. The decision of the board does not have to be binary. Rather, the board can decide on the priority of inclusion.

With respect to journal series, the following standards/details are considered:

- Journal status: discernible thematic focus, longevity of the series, reputation of editorial board and authors, citation analysis, support of a professional society
- Peer reviewing process: number of reviewers per article, time spent on each review, acceptance rate
- Adequate style: type setting, proper usage of reference lists in papers, presence author's contact information, an English abstract for non-English articles

It is discussed whether entirely non-English publications should be included in DBLP. Consensus is reached that if a publication stream is important on a national level (e.g., GI publications), then it may be included even if it is not in English. However, in general, publications in DBLP should be relevant and available on an international level.

With respect to conference and workshop series, the same standards should apply as with journal series. In addition, the following standards/details are considered:

- Long term availability of paper and/or electronic resources
- Reputation of steering/program committee
- Number of paper assignments per reviewer/committee member
- (Re-)publishing policy

The case of monographs is briefly discussed. The inclusion of textbooks could be decided based on the renown of the publisher. PhD theses should also be included to DBLP.

Michael Ley points out that we cannot expect that all valuable publication streams will step forward and submit a proposal to DBLP. There is a general agreement that DBLP should continue to include new

streams without a formal review process if the team and the board are convinced of the quality of the publication stream. When in doubt, the editor or steering committee of the stream should be asked to submit an application.

Actions:

- Marcel R. Ackermann (in coordination with Mila Majster-Cederbaum and Reinhard Wilhelm) prepares a draft application form for new publication streams.
- The application process shall be supported by JIRA. Marcel R. Ackermann prepares a prototype.
- The DBLP team in Trier is asked to make explicit the criteria that usually lead to acceptance or rejection. The findings shall be incorporated as guidelines into the formal review process.

Item 3: Scope of DBLP

Hans-Peter Lenhof and Dietmar Saupe report on the results of the Scope subcommittee. Lists with missing publication streams from interdisciplinary fields such as bioinformatics, medical informatics, information systems, geoinformatics, etc are presented. Michael Ley points out that the problem of many interdisciplinary publication streams lies in the unavailability of high quality meta data.

It is discussed to include information from other bibliographic databases such as PubMed into DBLP. Another option can be to provide links from a DBLP author page to the author's publication list in other databases. Such an approach is currently in negotiations with Zentralblatt MATH.

To obtain information on missing interdisciplinary publication streams it is suggested to query the GI's special interest groups. Interdisciplinary streams should be assessed with respect to scientific impact as well as to their the vicinity to core computer science.

Actions:

- The DBLP team will check the prepared lists for possible additions to the database.
- The DBLP team will contact the GI's special interest groups
- Each member of the board is encouraged to propose additional information on interdisciplinary fields or streams.

Item 4: Bibliometrics

Rüdiger Reischuk presents a conceptual framework for bibliometric evaluations in computer science. As objective of bibliometric analyses, the evaluation of single authors or institutions is assumed. Rüdiger Reischuk argues that the evaluation of all relevant publications individually will require too much time and resources to be viable. As an approximation, it is suggested to use the estimated quality of a publication stream as a proxy for the quality of a paper.

In conclusion, Rüdiger Reischuk proposes a non-automated process which bases its evaluation on objectively verifiable standards and the judgment of representative expert groups. The results should not be a precise ranking but rather a relative categorization of quality (e.g., top – standard – sub-standard). The relative categorization should only apply within the particular sub-field of computer science the publication stream belongs to, not within computer science as a whole discipline. These evaluations should be repeated periodically.

Jürgen Teich presents the results of a poll he initiated within the GI FB TI (special interest group on computer engineering). 25 of the roughly 50 members of the special interest group chose to participate. The data shows a clear agreement on the top conferences and top journals in the field. This top group is followed by a "heavy tail" of publication streams that were not unanimously supported. Jürgen Teich also compared the poll results with the average number of citations that papers of a given publication stream receive. Citation information was provided by the Microsoft Academic Search database. It turns

out that both approaches show similar results.

Actions:

- Michael Wagner will investigate whether the GI's special interest groups are available as expert groups.
- Michael Wagner (in coordination with Rüdiger Reischuk and Jürgen Teich) prepares a concept of a large-scale poll with several expert groups.

Item 5: Perspectives and further development

Otto Spaniol reports on his inquiry among the researchers of RWTH Aachen. There were a number of small requests, but no hints on major problems with the service of DBLP.

Andreas Butz presents the prototype of an explorative visualization tool for the DBLP dataset. The tool allows to browse the data by several axes like year, co-authorship or keywords. Ideas from the prototype could be adopted to DBLP.

It is discussed whether DBLP should use crowd-sourcing concepts to improve its service. The DBLP team reports that a new system will be incorporated into the website to allow users to report errors and claim missing publications. Submitted information are checked by the DBLP team before its integration to the data set. Unmoderated user edits are not considered viable as long as there is no adequate concept to guarantee for the quality of submitted data and to avoid fraud or vandalism.

It is discussed that additional data such as an author's affiliation or thematic tagging of publication streams should be included to the DBLP dataset. The DBLP team and Hannah Bast's research group will work together to collect the data.

Marc Herbstritt points out that the scientific citation of data and source code (e.g., the DataCite initiative) is becoming increasingly significant. In the long run, DBLP might consider to cover these fields.

Actions:

- Hannah Bast, Claudius Korzen and Marcel R. Ackermann will work on collecting author affiliations and publication stream tags. First results should be presented on the next meeting.
- Marc Herbstritt will prepare a presentation on DataCite for a future meeting.

Item 6: Appointment of next meeting

The next meeting will take place on June 1st, 2012, in Mannheim, Germany. Otto Spaniol volunteered to arrange a meeting room at the computing center in Mannheim. If unavailable, Mila Majster-Cederbaum will try to arrange the chancellor's meeting room again. This is highly appreciated by all participants.

Meeting adjourns: 5:30 p.m.

Meeting notes: Marcel R. Ackermann, Trier, Germany, March 9th, 2012