

Report on DEIS'10: Advanced School on Data Exchange, Information, and Streams (A GI-Dagstuhl Seminar)

Phokion G. Kolaitis*

Maurizio Lenzerini†

Nicole Schweikardt‡

1. INTRODUCTION

Most computer scientists are familiar with the Schloss Dagstuhl - Leibniz Center for Informatics or, simply, Dagstuhl as the place “where computer scientists meet”. Over the years, literally thousands of computer scientists have attended one or more Dagstuhl seminars in which participants spend a week interacting with colleagues in an informal setting by sharing new results and work in progress, exchanging ideas, or embarking on new collaborations. Alongside these year-round seminars, however, Dagstuhl also hosts a different kind of event that is expressly geared towards students and post-doctoral scholars. Specifically, Dagstuhl is also the home of the GI-Dagstuhl Seminars¹, which are sponsored jointly by the German Society for Informatics (GI) and the Schloss Dagstuhl - Leibniz Center for Informatics. The designated purpose of GI-Dagstuhl Seminars is to enable young researchers to learn about new developments in a particular area of research through active engagement in the seminar, which is typically organized by an international team of senior researchers. GI-Dagstuhl Seminars take place by far less frequently than regular Dagstuhl-Seminars; actually, only one or two such seminars has taken place each year during the past six years. Furthermore, GI-Dagstuhl Seminars are limited to at most 20-25 participants, including the organizers.

This paper reports on GI-Dagstuhl Seminar 10452, an Advanced School on Data Exchange, Integration, and Streams (DEIS'10), which took place from November 7 to November 12, 2010 and was organized by the three authors.

2. SCIENTIFIC THEME

DEIS'10 focused on data exchange, data integra-

*UC Santa Cruz & IBM Research - Almaden, USA

†Università di Roma La Sapienza, Italy

‡Goethe-Universität Frankfurt am Main, Germany

¹<http://www.dagstuhl.de/en/program/gi-dagstuhl-seminars/>

tion, and data streams. These are three different, yet inter-related, facets of information integration that have been investigated in depth by the research community in recent years.

Both data exchange and data integration deal with the execution of information integration, but they adopt distinctly different approaches. Data exchange is the problem of transforming data residing in different sources into data structured under a target schema; in particular, data exchange entails the materialization of data, after the data have been extracted from the sources and re-structured into a unified format. In contrast, data integration can be described as symbolic or virtual integration: users are provided with the capability to pose queries and obtain answers via the unified format interface, while the data remain in the sources and no materialization of the restructured data is required.

In the basic data stream model, the input data consists of one or several streams of data items that can be read only sequentially, one after the other. This scenario is relevant for a large number of applications where massive amounts of data need to be processed. Typically, algorithms have to work with one or few passes over the data and a memory buffer of size significantly smaller than the input size.

3. PROCESS AND TIMETABLE

In response to a call for proposals for GI-Dagstuhl Seminars, we submitted a proposal for DEIS'10 in November of 2009. In this proposal, we described the scientific theme of DEIS'10, listed the specialized topics, and also spelled out the procedure for selecting participants in this event. After our proposal was accepted, we disseminated the plan for DEIS'10 via postings to a number of forums, including DBWorld, and through a dedicated web page at <http://www.tks.cs.uni-frankfurt.de/events/deis10>

Potential applicants were asked to submit by July 15, 2010 an application consisting of a letter of in-

terest, a curriculum vitae, up to three representative papers or theses authored by the applicant, and a letter of recommendation from an academic supervisor or other senior colleague. We received 31 applications, out of which 22 applicants were selected to participate in DEIS'10; together with the organizers, this brought the total number of DEIS'10 participants to 25, which is the maximum that can be accommodated in a GI-Dagstuhl Seminar. The great majority of the applications received were of very high quality. In fact, we would have gladly accepted more applicants had there been more room. Of the 22 successful applicants, 18 were graduate students and 4 were postdoctoral scholars. In terms of geography, 18 were located in Europe, 3 in North America, and 1 in South America. We note that the participants were selected not on the basis of a paper they submitted, but, instead, on their ability and potential to participate in the event in a meaningful way.

The participants were notified of their selection on September 1, 2010. Each participant was asked to study the relevant literature in a specialized topic that was assigned to him or her by the organizers of DEIS'10, based on the interests and expertise of the participants. Each participant was assigned one of the three organizers as mentor. Mentors and mentees interacted via email during September and October 2010. In particular, participants were asked to send their mentors a progress report with an outline of their presentation on October 1, 2010, followed by a semi-final draft of the slides of their presentation on November 1, 2010.

During the first day of DEIS'10, each of the three organizers gave a 90-minute tutorial on one of the three main themes of the school. Specifically, there was a tutorial on "Schema Mappings and Data Exchange" by Phokion Kolaitis, a tutorial on "Data Integration" by Maurizio Lenzerini, and a tutorial on "Data Streams" by Nicole Schweikardt. The rest of the program consisted of the presentations by the participants. Each participant was given 45 minutes to present an overview of the specialized topic assigned to her or him; the presentations were followed by or were interspersed with questions by the audience, so that a total of one hour was allotted to each specialized topic.

While a small number of participants presented some of their own research work, most of the presentations were a synthesis of papers studied by the participants in the months before DEIS'10 took place. In total, well over 100 published papers were distilled and synthesized by the participants in their presentations. The slides of these presentations and

the relevant bibliographical references can be found at the web page of DEIS'10, which was given earlier in this section.

In addition to the tutorials and the presentations of specialized topics, an after-dinner problem session was held in the second day of DEIS'10. In this session, both the organizers and the participants presented selected open problems in each of the three main themes of DEIS'10. The last time slot of DEIS'10 was a wrap-up session during which feedback about the event was solicited and tentative plans for a follow-up event were discussed.

4. ASSESSMENT

From the viewpoint of the organizers, DEIS'10 was both an unqualified success and a truly satisfying experience. In the past, each of us had participated in typical advanced schools in which instructors give week-long short courses on a topic of their expertise. In that setting, attendees choose which courses to attend and then, for the most part, passively absorb the technical material presented by the instructors. The interaction between the instructors and the audience is usually limited to the questions that are asked from time to time during the course. As a result, it is unlikely that instructors get to know the attendees (or, at least, the majority of the attendees) and to have a quality interaction with them. Our experience with DEIS'10 was very different. The attendees of DEIS'10 were active and engaged participants who worked hard to first study a specialized topic and then present an overview of the topic assigned to them at the school. Furthermore, we got to know the participants well by evaluating their applications, communicating with them on a one-to-one basis via email before the event, and then meeting them in person at Dagstuhl and in a setting that fosters interaction and open communication.

From the viewpoint of the participants, DEIS'10 seems to have been a very positive experience for them. We base this assessment on both the feedback we received during the wrap-up session at the end of DEIS'10 and on the summary of the written survey that the participants completed by filling the form that is distributed to all participants of Dagstuhl seminars. In addition to high numerical ratings to questions concerning the scientific quality of the event, the inspiration of new ideas, and the identification of new research directions, participants expressed appreciation for the "interactive, informal atmosphere" and the "high quality of talks". In fact, one participant went as far as declaring that the worst aspect of this seminar is

that “it’s over”.

As regards criticisms and suggestions for improvement, there was a consensus that “the schedule was too dense”. As a remedy, some participants suggested having a mix of short and long presentations, instead of allocating the same amount of time to each participant. Others suggested cutting down the time of presentations uniformly for all participants. Of course, a different way to address this issue is to accept a smaller number of participants; however, this would be at the expense of turning down perfectly qualified and highly motivated applicants. Participants also suggested distributing open problems before the school and holding additional sessions during the school in which participants discuss open problems or present work in progress. Finally, at the scientific level, it was felt that some of the topics in the data streams area were disconnected from the topics in the other two areas of the school. Clearly, much remains to be done to strengthen the ties between research in data streams on the one side and research in data exchange and data integration on the other. Bringing people from all three areas together is the first step.

5. FOLLOW-UP

For some of the topics presented at DEIS’10, excellent survey articles already exist. Some other topics are still too nascent to justify survey articles. For several more mature topics for which no survey articles presently exist, we felt that the time is ripe to produce such survey articles. To this effect, we submitted a proposal to the Scientific Directorate of Dagstuhl to compile a volume consisting of state-of-the-art surveys of selected topics in data exchange, integration, and streams. Our proposal has now been accepted, and the planned volume will be published in the series *Dagstuhl Follow-Ups*, which is a new open-access publication venue for peer-reviewed articles based on Dagstuhl seminars (see <http://www.dagstuhl.de/dfu>). The survey articles in this volume will be authored by a number of DEIS’10 participants, who have already been invited and agreed to work on this project.

6. ACKNOWLEDGMENTS

We are grateful to the Schloss Dagstuhl - Leibniz Center for Informatics and to the German Society for Informatics (GI) for giving us the opportunity to organize this event. In particular, we thank Dr. Marc Herbstritt, Member of the Scientific Staff of the Schloss Dagstuhl - Leibniz Center for Informatics, for his help and encouragement at every step of this project. We also wish to acknowledge

the German Research Foundation (DFG) and the ACM Special Interest Group on Management Of Data (SIGMOD) for their generous financial support, which made it possible to provide travel support to participants of DEIS’10. Finally, we wish to thank all participants for their hard work, engagement, and enthusiasm that made DEIS’10 a successful event.