Career-Enhancing Services at SIGMOD Online

Alexandros Labrinidis
Department of Computer Science
University of Maryland
`labrinid@cs.umd.edu`

Alberto O. Mendelzon
Department of Computer Science
University of Toronto
`mendelzon@cs.toronto.edu`

`www.dbjobs.org`

Abstract

This article serves three purposes. First of all, to introduce *dbjobs*, the database of database jobs, and also describe its functionality and architecture. Secondly, to present statistics for the dbgrads system, after 18 months of continuous operation. Finally, to describe exciting future projects for SIGMOD Online.

1 Introduction to dbjobs

As the name implies, dbjobs is the searchable, on-line database of database jobs. It is moderated, specializing in database-related job vacancies worldwide.

The idea for a job posting database originated in the DB Junior mailing list[1] in the summer of 2000. Since the SIGMOD database of graduating students, *dbgrads*, was already operational, dbjobs was created as a joint effort between ACM SIGMOD and DB Junior. Under the current implementation, the two databases share a common user base and cater both active and passive job seekers with database expertise from all over the world.

Although generic job sites boast hundreds of thousands of job postings, in comparing dbjobs to them, one needs to remember only two numbers: 100% and zero.

- **100%** is the percentage of jobs in the dbjobs database that are *database-related*. All postings must be approved by a moderator and thus only database-related vacancies will be listed. This approach is better than the keyword-based searching used at generic job sites, because we give the term *database-related* broader semantics that include for example data mining, knowledge bases, data warehousing, geographic information systems, etc.

- **zero** is the cost to post a job vacancy in the dbjobs database and is also the cost to search the database. DB Jobs is a free service to the database community from ACM SIGMOD and DB Junior, and is currently being hosted at the University of Maryland, College Park. Commercial job sites usually allow free searching, but charge for posting job advertisements. This typically prohibits universities, research labs and small companies from advertising their job openings.

The rest of the paper is organized as follows. In the next section we describe the functionality of dbjobs in more detail. In Section 3 we summarize the architecture behind the dbjobs system. Section 4 has an overview of new services at SIGMOD Online. Finally, we conclude in Section 5.

2 Using the dbjobs System

Free Registration In order to avoid misuse of the dbjobs system, all users need to register first. Registration is short and simple: only the user's name, email address and a password are required. The email address will be used as the account name

---

[1]DB Junior is the international organization of junior database researchers. DB Junior was founded in May 1999 during the EDBT Summer School. For more information please visit `www.dbjunior.org`.

and has to be verified. For this reason, an email is sent to the new user's address and he/she has to respond back or visit a pre-specified URL in order to activate the account.

Job Classification Job vacancies posted in dbjobs are always database-related. However, to help job-seekers narrow down their query results, we have implemented a unique classification system based on three characteristics:

1. The type of the employer: academia, government, industry, research lab, start-up, or other.

2. The employment type: permanent, temporary, contract, post-doc, or internship.

3. The type of the position, i.e. whether its duties involve development, management, research, and/or teaching. All that apply for the advertised position must be selected.

Searching the Jobs Database The dbjobs system aims at the global job market, so except for the job classification, users can search the jobs database by country, and also by state for searches within the US. Additional search parameters include the minimum academic degree required for the job and whether any experience is required. Finally, the job start date can be used to limit the search results.

Additional Job Information Except for the searchable attributes for each job vacancy, there is additional information that is mostly textual and thus non-searchable. It includes the name of the employer, the job title, a textual job description, contact information, and URL pointers to the employer's home page and/or to additional information for the advertised job.

Submitting a Job Vacancy Any registered user can submit a job vacancy. However, in order to guarantee the accuracy of the posted advertisements, the email address of the user posting the job vacancy must match the domain of the employer's URL. In other words, one will need to have an ibm.com email address to post a job for IBM. Note,

however, that there are no restrictions for the email address listed as the contact information, which can be different from the poster's registered email address. After a job advertisement is submitted, it needs to be approved by the moderator, before being published in the jobs database. After publication, a notification message is emailed to the user who submitted the job posting.

Personalization Every registered user must login before accessing the dbjobs system. Except for preventing misuse, limiting access to registered users will allow us to support personalization in the future. Personalization plans include storing previous queries for each user and employing "triggers" for new entries that match a pre-specified profile. Using cookies, we allow automatic logins for the same user from the same machine for up to ten days.

Advertising Currently, dbjobs does not accept paid advertisements. However, there are two non-paid forms of advertising on the dbjobs home page. The *job vacancy showcase* is a randomly selected job advertisement that is displayed on the dbjobs home page for about a week. Furthermore, all employers that have posted a job vacancy in dbjobs are listed in the *featured employers list*, which also appears in the dbjobs home page.

3 dbjobs System Architecture

The driving force in the design and implementation of the dbjobs system was its integration with dbgrads, the SIGMOD database of graduating database students (`www.dbgrads.org`). Specifically, we wanted to re-use as much of the dbgrads code as possible and also enable the dbjobs system to share the registered user database with the dbgrads system. Fortunately, modifying the dbgrads code was not a problem, since both systems were implemented by the same person[2].

We kept the original software setup from dbgrads in the combined dbjobs/dbgrads system;

---

[2]Alexandros Labrinidis did the implemention of dbgrads and dbjobs. For a list of the many people who provided valuable feedback please visit `www.dbjobs.org/credits.html`

we used the Apache web server with mod_perl, MySQL for the back-end DBMS, and CGI.pm/Perl for HTML generation and form parsing. Labrinidis and Roussopoulos [LR00] give a detailed description of the original dbgrads system.

We identified common tasks from both systems and packaged them as separate modules. Examples of common tasks are performing user authorization, sending email, and printing HTML headers and footers. For this to work properly, we needed to have different initializations for certain attributes, which we formalized by implementing our own Perl *template* library. The current dbjobs/dbgrads implementation uses templates heavily, which greatly simplifies the code and also facilitates the creation of "clone" services. For example, *PE grads*, the database of graduating students in performance evaluation (`www.pegrads.org`), is a direct clone of dbgrads.

Both dbjobs and dbgrads, as well as their "clones," share a unified database of registered users, which greatly simplifies user management for these services. Finally, in addition to making it easier for people to join, the shared database will enable future personalized services.

4   Online Services at SIGMOD

SIGMOD Online, at `www.acm.org/sigmod`, is the web site for the ACM Special Interest Group on Management of Data. It contains, among other resources, information about SIGMOD activities, officers, and awards, a free database software catalog, a calendar of database-related events, advance copies of the proceedings of the SIGMOD and PODS conferences, online versions of (parts of) the SIGMOD Anthology and Digital Symposium Collection, the web pages for the SIGMOD Record, and a mirror of Michael Ley's DBLP bibliography. In this article we highlight SIGMOD Online career-related services of the past, present, and future.

4.1   Past: dbgrads

The dbgrads system is the searchable, online database of graduating database students from all over the world. It is a free service to the database community by ACM SIGMOD, and is currently being hosted at the University of Maryland. Students can list themselves in `dbgrads`, specifying their qualifications and their job preferences, including when they will be available and what type of job they are looking for. Employers can query the database in order to locate talented candidates for their job vacancies, which can include summer internships.

The dbgrads system went online in December 1999 and has operated continuously for the past 18 months. 1195 people have registered for the service as of July 10, 2001. Table 1 has the (user-provided) location distribution for all registered dbgrads users. Clearly, dbgrads has a world-wide audience. More than half of the registered users live outside North America, with the biggest concentrations in Europe (27%) and Asia (15%).

continent	users	
not specified	10	1%
Africa	12	1%
North America	571	48%
South America	38	3%
Asia	179	15%
Europe	324	27%
Oceania	61	5%

Table 1: Location distribution for dbgrads users

Out of the 1195 registered dbgrads users, 451 (38%) also submitted their entry to the graduating student database. Out of the 451 dbgrads students, 91 (20%) will be graduating (or graduated) with a Bachelors degree, 198 (44%) with a Masters degree, and 151 (34%) with a PhD.

Students who added their entry to the dbgrads database had the option of specifying a preference for a type of employer: industry, research lab, or academia. If they had more than one preference (e.g. research lab and academia), they were instructed to declare "no preference." Table 2 has the distribution for these preferences, grouped by year of graduation. Although these numbers are aggregated over students from all over the world, we suspect we can see the results of the recent high-tech slowdown: only 18% of students prefer working in

the industry in 2001, compared to 31% last year.

job type	2000		2001	
not specified	147	51%	90	58%
industry	91	31%	28	18%
research lab	32	11%	23	15%
academia	19	7%	13	9%

Table 2: Preference for type of employer

Graduating students were also able to declare preference for the type of position they were interested in, and were similarly instructed to declare no preference if they were interested in more than one type. Out of 451 students, 169 (37%) declared no preference, 231 (52%) were interested in a permanent position, 36 (8%) in an internship, and 15 (3%) in a post-doc position.

keyword	occurrences
database/s	216
data	160
web/internet/www	96
mining (data)	88
system/s	77
information	51
warehouse/-ing/olap	40
java	33
development	33
distributed	30
XML	27
query	26

Table 3: Frequency of keywords in interests list

Finally, in Table 3 we present the frequencies of keywords in the interests field for dbgrads students. 401 out of 451 students (89%) supplied a list of interests. The word "database" or "databases" occurred 216 times (48%). Other popular words were data, web/internet/www and data mining.

4.2 Present: dbjobs

As described above, dbjobs is the moderated, searchable, online database of database-related jobs. It is a joint effort between ACM SIGMOD and DB Junior. After rigorous testing, the dbjobs system (`www.dbjobs.org`) went live in May 2001.

So far, in addition to the 1195 registered dbgrads users, 87 new users registered through dbjobs, bringing the total to 1282.

	active	passive
employers	dbgrads	dbjobs
job-seekers	dbjobs	dbgrads

Table 4: Applicability of dbgrads/dbjobs

Combined, dbgrads and dbjobs fulfill the needs of both employers and job-seekers, whether they are active or passive in their search for talented applicants or job vacancies (Table 4).

4.3 Future: napster-type paper search

Here is a sketch of a future SIGMOD Online service that would be of use not only to those of you starting your careers, but to the community as a whole. Simply put, it's a Napster [nap] for papers. We include it in this article because some of you may be interested in contributing, and this is at the moment an idea in search of champions to carry it forward.

A bit of background: a couple of years ago, when one of us took over as SIGMOD Information Director, we had the idea of building a repository of (pointers to) online database papers. However, two parallel developments made the task seem less urgent: the establishment of CoRR, the Computing Research Repository [H+98], and the expansion of the SIGMOD Anthology and DiSC. CoRR has not grown as hoped; we can only speculate as to the reasons, but we suspect that by the time CoRR was set up, the existence of a fairly convenient mechanism for paper sharing–the Web–weakened the need for a central repository. The Anthology and DiSC have been extremely successful, but they contain, for the most part, already published material.

There is still the need for easy dissemination of things like full versions of conference papers, early descriptions of work in progress, graduate theses, etc.; exactly the kind of publications that used to go by the name of technical reports and were distributed by physical mail in the distant past that most of you are too young to remember. Of course such a mechanism could also be used for already

published papers, subject to whatever intellectual property constraints apply. It is probably obvious to every reader of this bulletin that having papers freely available online is A Good Thing; objective evidence for this is given by Lawrence [Law01], who reports strong correlation between free online availability and number of citations.

Requirements   So, we are proposing a service that supports registering and searching for research papers that may be stored anywhere on the Web. What should such a service be like?

First, it should be easy to get one's papers registered. We can envision a graded system where the more information an author supplies, the better the service provided. The minimum level of information is a pointer to a bunch of papers. The author can supply this in the form of a URL, or, as in Napster, by placing the papers in a certain directory, or in some other equally simple way. A higher level of information is a set of XML descriptions of each paper (using, e.g. the standard bibtex XML encoding described by Hendrikse [Hen01]).

Second, the data should be fresh. Broken or obsolete links will quickly discourage users. Of course, some of the responsibility for freshness rests with the data providers–the authors.

Third, search should be fast and effective. The more information an author has supplied, the smarter the search should be.

Technical Challenges   There are some interesting questions posed by the design and implementation of such a system.

- Is there a centralized index like in traditional Web search engines or Napster? Distributed index? Replicated index? No index, a la Gnutella [gnu]? Gribble *et al.* [GHI+01] give an overview of problems related to peer-to-peer databases.

- Given the minimum author-supplied information, a URL, what exactly is indexed? Are pages pointed to by this URL crawled? How far?

- If an index is built, how is its freshness ensured? Do authors provide updates, or are they automatically detected? (Or both?)

- Who gets to "publish" their papers? Is there any sort of access control?

- How do we ensure there are no copyright violations? (Or that, if there are, they are the authors' responsibility, not SIGMOD's).

- What is the mechanism for participating as an author? Filling out a Web form? Sending an email? Downloading a program? Being "nominated" by someone else?

- What is the data model? The query language? How is query processing done?

5   Conclusions

We have surveyed the past, present and future of some SIGMOD Online services. We would like to emphasize that SIGMOD is a volunteer-driven organization and that people at the start of their careers tend to make great volunteers, so please get involved by sending mail to either one of us. Many of the services that we take for granted now, such as the DB World mailing list, the DBLP, the SIGMOD Anthology, and the DiSC collection, exist because of the hard work of many volunteers.

References

[GHI+01]   Steven Gribble, Alon Halevy, Zachary Ives, Maya Rodrig, and Dan Suciu. "What Can Peer-to-Peer Do for Databases, and Vice Versa?". In *Proceedings of the Fourth International Workshop on the Web and Databases (WebDB'2001)*, Santa Barbara, California, USA, May 2001.

[gnu]   Gnutella resources. Available online at `gnutella.wego.com`/.

[H+98]   Joseph Halpern et al. "CoRR: Computing Research Repository". Available online at `www.acm.org/repository`/, November 1998.

[Hen01]   Zeger Hendrikse. "An XML equivalent of BibTeX: bibteXML". Available at `xml.coverpages.org/bibteXML.html`, June 2001.

[Law01]   Steve Lawrence. "Online or Invisible?". *Nature*, 411(6837):521, 2001.

[LR00]   Alexandros Labrinidis and Nick Roussopoulos. "Generating dynamic content at database-backed web servers: cgi-bin vs mod_perl". *SIGMOD Record*, 29(1), March 2000.

[nap]   Napster homepage. Available online at `www.napster.com/`.

Alexandros Labrinidis   is a PhD candidate at the Department of Computer Science of the University of Maryland, College Park. He is currently the Associate Editor for the Career Forum column and the webmaster for SIGMOD Record. His research interests include scalable database-backed web servers, Quality of Service, Quality of Data, and data warehousing.

Alberto O. Mendelzon   is a professor at the Department of Computer Science of the University of Toronto. He received his PhD degree in Electrical Engineering and Computer Science from Princeton University. He is currently the ACM SIGMOD Information Director and the webmaster for SIGMOD Online, as well as Associate Editor for ACM TODS and the VLDB Journal. His research interests include OLAP and data warehousing, semistructured data, and web databases.