CSC 560	Modern DBMS Architectures	Alexander Dekhtyar

Assignment 3: Individual Report Proposal and Bibliography

Due date: Thursday, October 9, in-class

Assignment

This is Stage 1 of your individual course assignment. As such, each students needs to complete it.

Individual Course Assignment Overview

Your individual course assignment is a survey paper on a **modern research topic in the field of databases**. To make this assignment more straightforward, be break it into several parts. The first part essentially asks you to pick a topic and do an initial literature search for it.

The overall requirements for the survey paper are as follows.

R1. Currency of topic. The topic you choose to survey must be subject to current, significant ongoing research. *Topic currency* is defined as *several papers have been published on this topic in the past five years. Topic significance* is defined as: *the work published on this topic can be found in top journals and conferences in the field of databases* (see below).

Notice, that not all papers you survey must be published in the past five years. You are encouraged to trace the history of your research topic and to find and discuss the papers that established it. But you also must discuss current research. (Similarly, you may pick some papers that were published outside of the list of venues specified below, but it is expected that some of the papers you survey are from that list.)

R2. Scope of survey. You must survey the work of *multiple research groups* on the topic. This essentially means two things. First, you must read and dicuss in your survey multiple research papers. Second, while multiple papers by the same groups of authors or research group are allowed (and sometimes — encouraged), you *must survey work done by at least two independent research groups.*

R3. Related work. You must survey work that is *truly related.* If you choose a broad enough topic, there may be a lot of work in the area of your interest that is completely unrelated — the problems being addressed are different, the approaches cannot be compared, and so on. *Preparing such a survey is a mistake!* Pick the work that truly has traceable connections and is comparable. This can be accomplished in a number of ways:

- *Trace the history.* Trace the *vertical* of one research direction from its roots to the modern-day work. Use the bibliography in the most recent paper to discover what work it extends, find this work, use its bibliography to find prior work, repeat for as long as deemed necessary.
- *Explore branches.* Survey the *horizontal* snapshot of a research topic by discussing alternative approaches to solving the same problem or similar problems. This may or may not require going one step back in history, but the stress of the survey will still be on the state of the art and the *alternative solutions* to reaching pretty much the same goal.
- *Survey problems.* Rather than surveying multiple solutions to the same problem, take a step back, and discuss the space of all (comparable) problems in the field of your choosing.

  Example. The difference between this approach and the previous one can be described on the following simple example. Suppose you are interested in relational DBMS. An *explore branches* survey would pick one topic - for example cost-based query optimization, and would describe multiple cost models for query optimization and attempt to compare them based on the evaluation results presented in the surveyed papers. The *survey problems* approach would, for example, decompose the problem of building an efficient RDBMS into a collection of subproblems: data management on disk, buffer manager construction, transaction management and scheduling, crash recovery, relational algebra implementation, query compilation and optimization, query parsing. It would then describe a consensus, most popular, or, simply, a clearly influential solution to each of the smaller problems. (Note, your surveys don't need to break big problems into so many smaller ones).

R4. Number of papers. You must survey at least the equivalent of five conference papers. Journal papers, depending on their length and significance of contribution (as well as the breadth of scope) may be counted for multiple conference papers. This does NOT mean that you all have to survey *exactly five papers.* Your paper count should match the scope of your survey, and should be arrived at naturally, not in an artificial way.

Survey Proposal

For this assignment, each of you shall prepare a short survey proposal that includes the following information:

- A brief description of the area of database research you are concentrating on in your survey, and the specific problems you plan to study.
- A brief informal description of why this problem appeals to you (please, be honest — if you are picking a specific problem because it is easy to find material for it, or because it is the topic of your M.S. research — just state so. It won't affect the grading.
- A preliminary bibliography of the work you plan to survey (and any other work that you know you will need to cite). You must include a **complete academic citation** for each paper you include. You may include a URL for the electronic version of the paper (this might make my life a bit easier when grading), but it is not necessary, and you won't lose any points over it.

The proposal should be *terse*.

Picking a topic. The second key reason why you were asked to read up on 20 years of self-assessment studies from the database community is because these documents describe a lot of emerging or just-emerged problems in the area of databases and data management. The hope is that by reading these papers you are able to chose a problem that is relevant, current, and appeals to you personally as an object of study.

Literature Search

The following venues are considered to be top-tier in the field of databases.

Journals:

- ACM Transactions on Database Systems (ACM TODS)
- IEEE Transactions on Data and Knowledge Engineering (IEEE TKDE)
- VLDB Journal (a.k.a., Journal of the Society for Very Large Databases)
- Data and Knowledge Engineering (DKE)
- ACM SIGMOD Record (technical papers only)
- Journal of the ACM (database related papers only)

Conferences:

- ACM SIGMOD Conference
- Principles of Database Systems (PODS)
- International Conference on Very Large Databases (VLDB)
- International Conference on Database Theory (ICDT)

- International Conference on Extending Database Technology (EDBT)
- International Conference on Data Engineering (ICDE)
- ACM Conference on Information and Knowledge Management (ACM CIKM)
- ACM SIGMOD Workshops
- International Conference of World Wide Web (WWW) (database papers only)

There is also a number of smaller, *niche* conferences that provide coverage of subareas of the field of databases and manage to publish high quality papers. There is also a number of other conferences in which good work often appears.

Links to the on-line proceedings (or at least the contents) of most of these venues will be provided to you on the course web page by Thursday, October 3.

In your search for material to cover, you may peruse the on-line proceedings of the conferences listed above, on-line journal archives (some allow you to download the papers from Cal Poly IP addresses), as well as by using traditional tools for literature search in Computer Science: DBLP, CiteSeer, Google Scholar, Microsoft's version of Google Scholar, ArnetMiner and others. Links to some of these resources will be provided to you on the course web page.,

Submission

I prefer that you do all your writing for the course in LaTeX, but will accept documents that were created in MS Word, OpenOffice or GoogleDocs, as long as the submitted documents are (a) formatted appropriately and (b) converted to PDF. Name your PDF file `<LastName>-proposal.pdf` (e.g., my proposal would be `Dekhtyar-proposal.pdf`.

For this assignment, use `handin` to submit your PDF document to me. The `handin` command is

```
handin dekhtyar 560-a03 <file>
```

Submit the proposal by the end of the day on **Tuesday, October 9**.

Good Luck!