Answering Queries with Database Restrictions (Research Summary)

Chen Li

Stanford University, CA 94305
chenli@cs.stanford.edu,
Home page: http://www-db.stanford.edu/~chenli

I am a Ph.D. student in the Computer Science Department at Stanford University. My main research interests include query reformulation in the database domain. In particular, I work on the problems of how to do query planning and reformulation when relations have limited access patterns [7, 9]. That is, the binding patterns of the relations require values to be specified for certain attributes in order to retrieve data from a relation. Most of my work is based on the TSIMMIS project (http://www-db.stanford.edu/tsimmis/), a data-integration project [2, 10, 11] at Stanford.

The following are some query-reformulation problems that I have solved (with other database researchers at Stanford):

1. How to optimize queries on relations with binding restrictions [6, 13]. Since relations require values of certain attributes to return data, we cannot answer a query in the traditional way of answering queries. We prove that under the cost model that counts the number of source accesses, the problem of finding the optimal plan is NP-complete. We also give some heuristics for finding good plans and prove their bound.
2. How to compute the capabilities of mediators on relations with limited capabilities [12]. We consider other complicated relation capabilities besides binding restrictions.
3. How to do query reformulation to compute as many answers to a query as possible in the presence of binding restrictions [5]. We show that a query can be answered by borrowing bindings from sources not mentioned in the query. We also develop an algorithm for finding all the relations that need to accessed to answer a query.
4. How to test whether by query reformulation, it is possible to compute the complete to a query on relations with binding restrictions [3]. The complete answer to a query is the answer to the query that we could compute if we could retrieve all the tuples from the relations. Since we cannot retrieve all the tuples from a relation due to its binding restrictions, we need to do reasoning about whether the answer computed by a plan is really the complete answer.
5. How to test query containment in the presence of binding restrictions [4]. We show that the containment is decidable using the results of monadic programs [1], although containment of datalog programs in general is not decidable [8].

Currently I am working on the problem of how a mediator accesses sources to keep its cached data as fresh as possible, while minimizing the number of source accesses. It is an instance of a reformulation problem.

References

1. S. S. Cosmadakis, H. Gaifman, P. C. Kanellakis, and M. Y. Vardi. Decidable optimization problems for database logic programs. *ACM Symposium on Theory of Computing (STOC)*, pages 477–490, 1988.
2. M. R. Genesereth, A. M. Keller, and O. M. Duschka. Infomaster: An information integration system. In *Proc. of ACM SIGMOD*, pages 539–542, 1997.
3. C. Li. Computing complete answers to queries in the presence of limited access patterns (extended version). *Technical report, Computer Science Dept., Stanford Univ.*, 1999.
4. C. Li. Testing query containment in the presence of binding restrictions. *Technical report, Computer Science Dept., Stanford Univ.*, 1999.
5. C. Li and E. Chang. Query planning with limited source capabilities. *International Conference on Data Engineering (ICDE)*, 2000.
6. C. Li, R. Yerneni, V. Vassalos, H. Garcia-Molina, Y. Papakonstantinou, J. D. Ullman, and M. Valiveti. Capability based mediation in TSIMMIS. In *Proc. of ACM SIGMOD*, pages 564–566, 1998.
7. A. Rajaraman, Y. Sagiv, and J. D. Ullman. Answering queries using templates with binding patterns. In *Proc. of ACM Symposium on Principles of Database Systems (PODS)*, pages 105–112, 1995.
8. O. Shmueli. Equivalence of datalog queries is undecidable. *Journal of Logic Programming*, 15(3):231–241, 1993.
9. J. D. Ullman. *Principles of Database and Knowledge-base Systems, Volumes II: The New Technologies*. Computer Science Press, New York, 1989.
10. J. D. Ullman. Information integration using logical views. *International Conference on Database Theory (ICDT)*, pages 19–40, 1997.
11. G. Wiederhold. Mediators in the architecture of future information systems. *IEEE Computer*, 25(3):38–49, 1992.
12. R. Yerneni, C. Li, H. Garcia-Molina, and J. D. Ullman. Computing capabilities of mediators. In *Proc. of ACM SIGMOD*, pages 443–454, 1999.
13. R. Yerneni, C. Li, J. D. Ullman, and H. Garcia-Molina. Optimizing large join queries in mediation systems. *International Conference on Database Theory (ICDT)*, pages 348–364, 1999.