

Published: 02 December 2014

# Detecting common subexpressions for multiple query optimization over loosely-coupled heterogeneous data sources

[Mahesh B. Chaudhari](#) & [Suzanne W. Dietrich](#) 

*Distributed and Parallel Databases* **34**, 119–143 (2016)

**330** Accesses | **6** Citations | [Metrics](#)

## Abstract

---

The research presented in this paper supports the identification of common subexpressions as candidates for potential materialized views that form the basis of multiple query optimization in a loosely-coupled distributed system where query expressions access heterogeneous data sources, including relations and data-centric XML. This paper introduces a unifying mixed multigraph formalism to represent SQL, XQuery, and LINQ queries in a common query graph model and a heuristics-based algorithm to detect common subexpressions. The identified common subexpressions represent an opportunity for defining a materialized view to avoid repeating computation. The common subexpressions may access only relations, only XML, or a combination of relations and XML. The mixed multigraph model and the heuristic rules presented in this paper have distinguished advantages over the existing approaches that consider only relational or XML data sources individually. The mixed multigraph model can present SQL, XQuery, and LINQ queries in a single graph model and the heuristic rules are designed to consider the identical and subsumed conditions at the same time. A prototype implementation of the algorithm illustrates the applicability of the approach using various examples from the research literature as well as scenarios over a Criminal

Justice enterprise that include common subexpressions across relational and XML data sources.

This is a preview of subscription content, [access via your institution](#).

Access options	
<p>Buy article PDF</p> <p><b>US\$ 39.95</b></p> <p>Tax calculation will be finalised during checkout.</p> <p>Instant access to the full article PDF.</p> <p><a href="#">Rent this article via DeepDyve.</a></p>	<p>Buy journal subscription</p> <p><b>US\$ 79</b></p> <p>Tax calculation will be finalised during checkout.</p> <p>Immediate online access to all issues from 2019. Subscription will auto renew annually.</p> <p><a href="#">Learn more about Institutional subscriptions</a></p>

## References

1. 1.  
Antlr: Antlr v3 (2010). <http://www.antlr.org/>
2. 2.  
Chakravarthy, U.S., Minker, J.: Multiple query processing in deductive databases using query graphs. In: VLDB '86: Proceedings of the 12th International Conference on Very Large Data Bases, pp. 384–391. Morgan Kaufmann Publishers Inc., San Francisco (1986)
3. 3.  
Chaudhari, M., Dietrich, S.: Metadata services for distributed event stream processing agents. In: 19th International Conference on Software Engineering and Data Engineering, pp. 307–312. San Francisco (2010)
4. 4.

Chaudhari, M.B.: Materialized views over heterogeneous structured data sources in a distributed event stream processing environment. Ph.D. thesis, Arizona State University, Tempe, AZ, USA (2011)

5. 5.

Chen, F.C.F., Dunham, M.H.: Common subexpression processing in multiple-query processing. *IEEE Trans. Knowl. Data Eng.* **10**(3), 493–499 (1998)

6. 6.

Diao, Y., Florescu, D., Kossmann, D., Carey, M.J., Franklin, M.J.: Implementing memoization in a streaming xquery processor. In: Bellahsene, Z., Milo, T., Rys, M., Suciu, D., Unland, R. (eds.) *XSym. Lecture Notes in Computer Science*, vol. 3186, pp. 35–50. Springer, Berlin (2004)

7. 7.

Elsayed, I., Brezany, P., Tjoa, A.M.: Towards realization of dataspace. In: *DEXA '06: Proceedings of the 17th International Conference on Database and Expert Systems Applications*, pp. 266–272. IEEE Computer Society, Washington, DC, USA (2006).  
doi:[10.1109/DEXA.2006.140](https://doi.org/10.1109/DEXA.2006.140)

8. 8.

Franklin, M., Halevy, A., Maier, D.: From databases to dataspace: a new abstraction for information management. *SIGMOD Rec.* **34**(4), 27–33 (2005).  
doi:[10.1145/1107499.1107502](https://doi.org/10.1145/1107499.1107502)

9. 9.

Gupta, A., Mumick, I.S. (eds.): *Materialized Views: Techniques, Implementations, and Applications*. MIT Press, Cambridge (1999)

10. 10.

Halevy, A., Franklin, M., Maier, D.: Principles of dataspace systems. In: PODS '06: Proceedings of the Twenty-fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 1–9. ACM, New York, NY, USA (2006).

doi:[10.1145/1142351.1142352](https://doi.org/10.1145/1142351.1142352)

11. 11.

Halevy, A.Y.: Answering queries using views: a survey. VLDB J. **10**, 270–294 (2001).

doi:[10.1007/s007780100054](https://doi.org/10.1007/s007780100054)

12. 12.

Jarke, M.: Common subexpression isolation in multiple query optimization. In: Query Processing in Database Systems, pp. 191–205. Springer (1985)

13. 13.

Kossmann, D.: The state of the art in distributed query processing. ACM Comput. Surv. **32**, 422–469 (2000).

doi:[10.1145/371578.371598](https://doi.org/10.1145/371578.371598)

14. 14.

Lenzerini, M.: Data integration: a theoretical perspective. In: Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS '02, pp. 233–246. ACM, New York, NY, USA (2002). doi:[10.1145/543613.543644](https://doi.org/10.1145/543613.543644)

15. 15.

Löwy, J.: Programming WCF Services. O'Reilly Media, Inc. (2007)

16. 16.

Microsoft Corporation: Linq.net language-integrated query (2009). <http://msdn.microsoft.com/en->

17. 17.

Park, J., Segev, A.: Using common subexpressions to optimize multiple queries. In: Proceedings of the Fourth International Conference on Data Engineering, pp. 311–319. IEEE Computer Society, Washington, DC, USA (1988)

18. 18.

QuickGraph: Quickgraph.  
<http://quickgraph.codeplex.com/> (2011)

19. 19.

Roy, P., Seshadri, S., Sudarshan, S., Bhoje, S.: Efficient and extensible algorithms for multi query optimization. SIGMOD Rec. **29**, 249–260 (2000).  
doi:[10.1145/335191.335419](https://doi.org/10.1145/335191.335419)

20. 20.

Sellis, T.K.: Multiple-query optimization. ACM Trans. Database Syst. **13**(1), 23–52 (1988).  
doi:[10.1145/42201.42203](https://doi.org/10.1145/42201.42203)

21. 21.

Sybase: Complex event processing.  
<http://www.sybase.com/products/financialservicesolutions/complex-event-processing> (2010)

22. 22.

Urban, S., Dietrich, S., Chen, Y.: An xml framework for integrating continuous queries, composite event detection, and database condition monitoring for multiple data streams. In: Chandy, M., Etzion, O., von Ammon, R. (eds.) Event Processing, no. 07191 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss

Dagstuhl, Germany, Dagstuhl, Germany (2007).

<http://drops.dagstuhl.de/opus/volltexte/2007/1142>

23. 23.

Vizing, V.: On incidentor coloring in a partially directed multigraph. *J. Appl. Ind. Math.* **3**, 297–300 (2009).

doi:[10.1134/S1990478909020161](https://doi.org/10.1134/S1990478909020161)

24. 24.

Xamarin: Mono (2013). <http://www.mono-project.com>

25. 25.

Zhou, J., Larson, P.A., Freytag, J.C., Lehner, W.: Efficient exploitation of similar subexpressions for query processing. In: *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, SIGMOD '07*, pp. 533–544. ACM, New York, NY, USA (2007). doi:[10.1145/1247480.1247540](https://doi.org/10.1145/1247480.1247540)

## Acknowledgments

---

This material is based upon work supported by the National Science Foundation under Grant No. 0915325. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## Author information

---

### Affiliations

1. School of Mathematical and Natural Sciences, Arizona State University, Phoenix, AZ, 85609-7100, USA

Mahesh B. Chaudhari & Suzanne W. Dietrich

## Corresponding author

Correspondence to [Suzanne W. Dietrich](#).

Rights and permissions

---

[Reprints and Permissions](#)

About this article

---

## Cite this article

Chaudhari, M.B., Dietrich, S.W. Detecting common subexpressions for multiple query optimization over loosely-coupled heterogeneous data sources. *Distrib Parallel Databases* **34**, 119–143 (2016).

<https://doi.org/10.1007/s10619-014-7166-6>

- Published 02 December 2014
- Issue Date June 2016
- DOI <https://doi.org/10.1007/s10619-014-7166-6>

## Keywords

- Distributed databases
- Common subexpressions
- LINQ
- SQL
- XQuery
- Event and stream processing
- Heuristic rules

Not logged in - 184.103.221.104

Not affiliated

