

Leveraging Bitmap Indexing for Subgraph Searching

Título Leveraging Bitmap Indexing for Subgraph Searching

Autores David Luaces, José R.R. Viqueira, Tomás F. Pena, José M. Cotos

Tipo Comunicación para congreso

Fonte  22nd International Conference on Extending Database Technology (EDBT), Lisbon (Portugal), OpenProceedings.org, pp. 1-12 , 2019.

Rank  Ranked A in CORE

ISBN 978-3-89318-081-3

Abstract Deciding whether a query graph is a subgraph of some other in a very large database of small graphs is a problem of major interest in many application domains. As an example, it arises in the searching for specific molecular substructures in currently available molecular databases, whose sizes may reach levels close to one hundred million. State of the art methods to solve this problem follow a filter-then-verify (FTV) paradigm, where an indexing technique is first used in a filtering stage to obtain result candidates and a subgraph isomorphism algorithm is next applied to the candidates in a verification stage to obtain the final result. Among all the available techniques of the state of the art, two of them have demonstrated better performance when applied to large datasets, namely, the GraphGrepSX (GGSX) and CT-Index (CTI). In this paper, three new indexing techniques, one based on GGSX and two based on CTI, are proposed. In particular, Bitmap GGSX (BM-GGSX) leverages the use of bitmaps in the trie structure used by GGSX to achieve performance gains of around 90% in the filtering stage. Column-Wise CT-Index (CW-CTI) exploits a column-wise representation of the fingerprints (bitmaps) used by CT-Index to reduce the filtering times around 80% for small queries (8 edges). Finally, K-Means CT-Index (KM-CTI), constructs a binary tree of bitmaps from the CT-Index fingerprints to reach filtering time reductions of around 70% for medium queries (20 edges) and 75% for large queries (40 edges).

DESCARGAS

 Referencia BibTex

 Descargar versión do editor

PROXECTOS DE INVESTIGACIÓN

Nextchrom: Cromatografía virtual avanzada e xestión integral de datos analíticos e moleculares

PROGRAMAS CIENTÍFICOS

Procesamento aproximado