Holistic twig join algorithm facilitates fast, efficient data retrieval from XML databases

Retrieving data from tree-structured XML databases can be a tedious and computationally intensive process that produces extraneous intermediate partial matches that are not ultimately returned as a search result. This technology is an XML query process, TwigStack, which provides a compact search path for easy retrieval of search matches. TwigStack implements a novel holistic twig join algorithm for matching an XML query twig pattern. The technique uses a chain of linked stacks to compactly represent partial results, which are then composed to obtain matches for the twig pattern. TwigStack provides a streamlined query solution to return relevant search results quickly, allowing XML databases to be navigated more efficiently.

TwigStack provides optimal query pattern matching by avoiding the creation of large intermediate results

Prior XML query algorithms have typically decomposed the twig pattern into binary structural relationships and achieved twig matching by either (i) using structural join algorithms to match the binary relationships against the XML database, or (ii) stitching together these basic matches. A disadvantage of this approach is that the size of the intermediate results can be large even when the input and output sizes are smaller. TwigStack overcomes this limitation by first generating partial solutions to individual query paths, then merging them together to compute the answers to the twig pattern. This ensures that no intermediate solution is larger than the final answer to the query twig pattern.

Analysis of TwigStack using a range of real and synthetic experimental data showed that the algorithm was I/O and CPU optimal for a large class of query twig patterns.

Lead Inventor:
Nicolas Bruno, Ph.D.
Applications:

- Query searches in search engines such as Google, Google Scholar and Bing
- Query searches in databases such as library databases, medical records databases
- Query searches in relational database management systems such as Oracle, IBM, Microsoft
- Query searches within social networking sites to retrieve old posts
- Query searches for genetics databases

Advantages:

- Allows XML databases to be efficiently navigated
- Avoids the creation of large intermediate results
- Returns search matches quickly
- I/O and CPU optimal for a large class of query twig patterns

Patent Information:

- Patent Issued (US 8117189)
- Patent Issued (US 7483887)
- Patent Issued (US 7454415)

Tech Ventures Reference: IR Proxy71

Related Publications:


Inventors

Nicolas Bruno