Chained Forests for Fast Subsumption Matching

Overview

- Content is defined using profiles. Queries select a subset of profiles.
- Assumption: There are subsumption relations in the input sets.
- Idea: build forests of each set (profiles/queries) and maintain mappings between these sets that determines subsumption between sets.
- Essentially chain two forests with the mappings. The mechanism generalizes to a chain of forests.
- Research question: How to minimize the number of subsumption test operations when inserting/deleting a profile or a query?

Introduction

- Content provisioning and delivery are becoming increasingly popular and important.
- We present a novel scheme for the maintenance and matching of partial orders.
- Partial order derives from the subsumption relation inherent the collection of objects being matched.
- The proposed chaining technique has applications in information routing, collection tracking, and peer-to-peer information exchange.
- Track result set of a continuous query with insertions and deletions.

Algorithm and Optimizations

- Add profile: All elements that subsume the input profile are found by traversing the query forest towards subsuming elements.
- Add query: All subsumed elements are found by traversing the profile forest. All matching elements and their children added to the result set.
- Optimizations: Find candidate set for testing inclusion into the result set. Determine upper and lower bounds for the result set and inspect their intersection.
- Experimental results suggest significant improvement with the optimization.

Results

- Generalizes to multiple sets, each corresponding to a forest in a chain of forests.
- Separates mappings from the forests.
- Improved insertion and lookup cost. Improved or degraded deletion cost and structure size depending on the workload. Sparse bitstrings may be used to alleviate space concerns.
- Nature of optimizations suggest that copes well with self-similar workload.
- Future work investigates how to compact mappings.