



DIESEL: Blending Formulation and Processing of Hand-drawn Data Series Queries

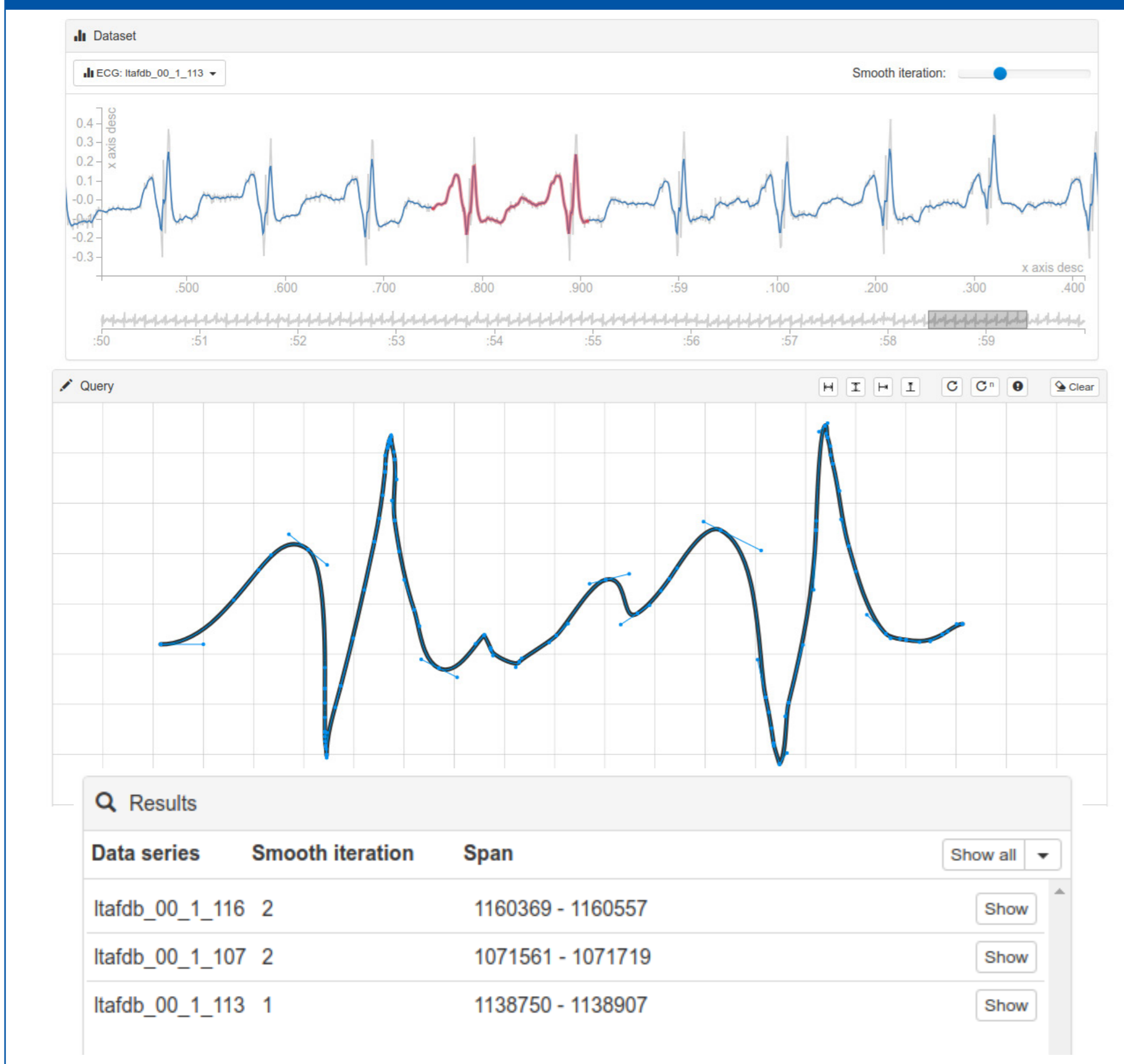
ABSTRACT

Efficient data series query reduces processing time needed to search for matching subsequences in a large database. These approaches often focus on queries that are predefined, and are often of fixed, predetermined length. Recent works on hand-drawn queries however, are not able to large datasets. We present DIESEL, a novel approach to query subsequences using hand-drawn sketches efficiently. Queries will be conducted in a scale-free manner and match based on pattern rather than absolute value. DIESEL first indexes the data based on sections, followed by retrieval of subsequence during the query, according to our matching metrics developed for hand-drawn sketches.

BACKGROUND

- Data series refers to a sequence of data points that are ordered often by time or other measures such as position or mass.
- State-of-the-art efficient query approaches utilizes euclidean distance to match subsequences based on exact length and absolute values which is hard to reproduce by sketches.
- Given a data series sketch from the user, the objective of DIESEL is to query for subsequences in the database which match the sketch pattern visually.
- In addition, hand-drawn queries incur GUI latency which we will utilize and blend query formulation with query processing to reduce runtime.

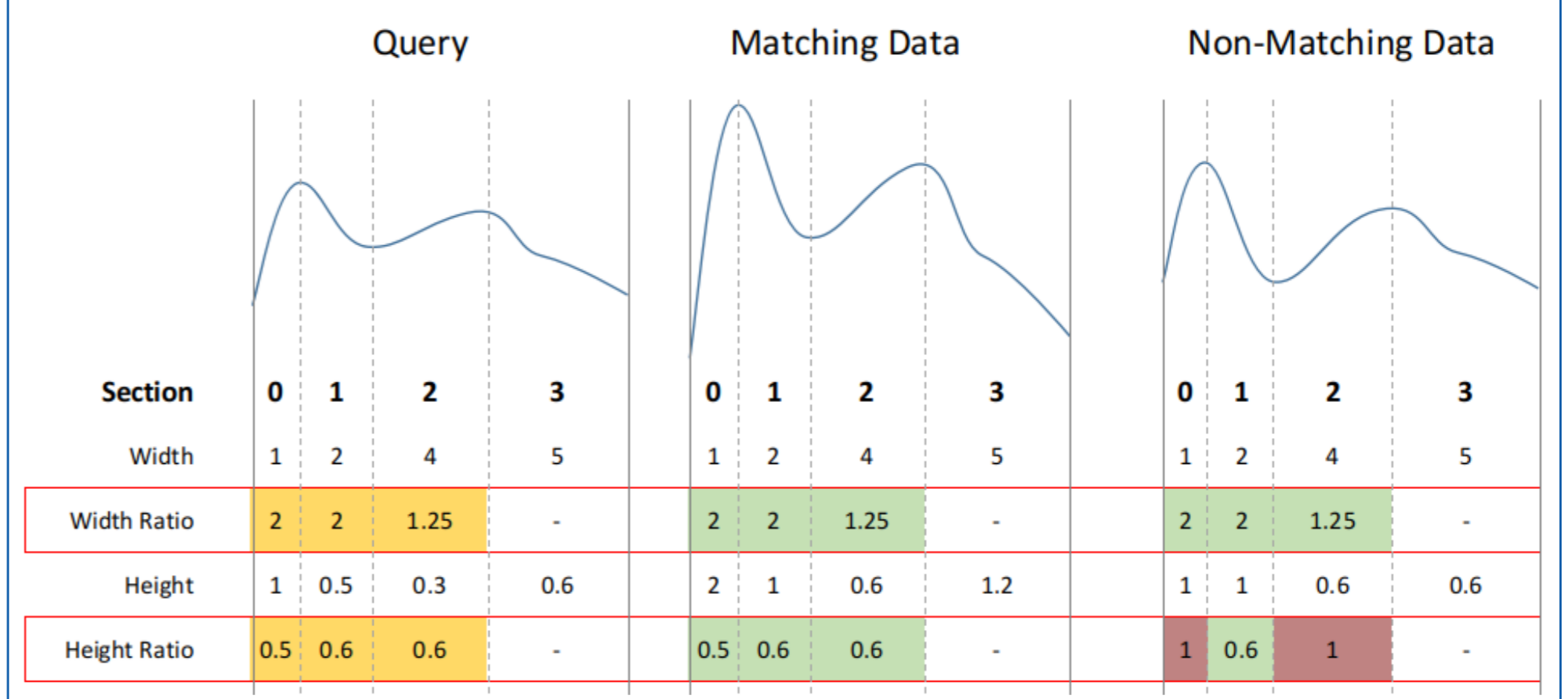
DIESEL QUERY INTERFACE



Student: Low Heok Hong | Supervisor: Associate Professor Sourav S Bhowmick

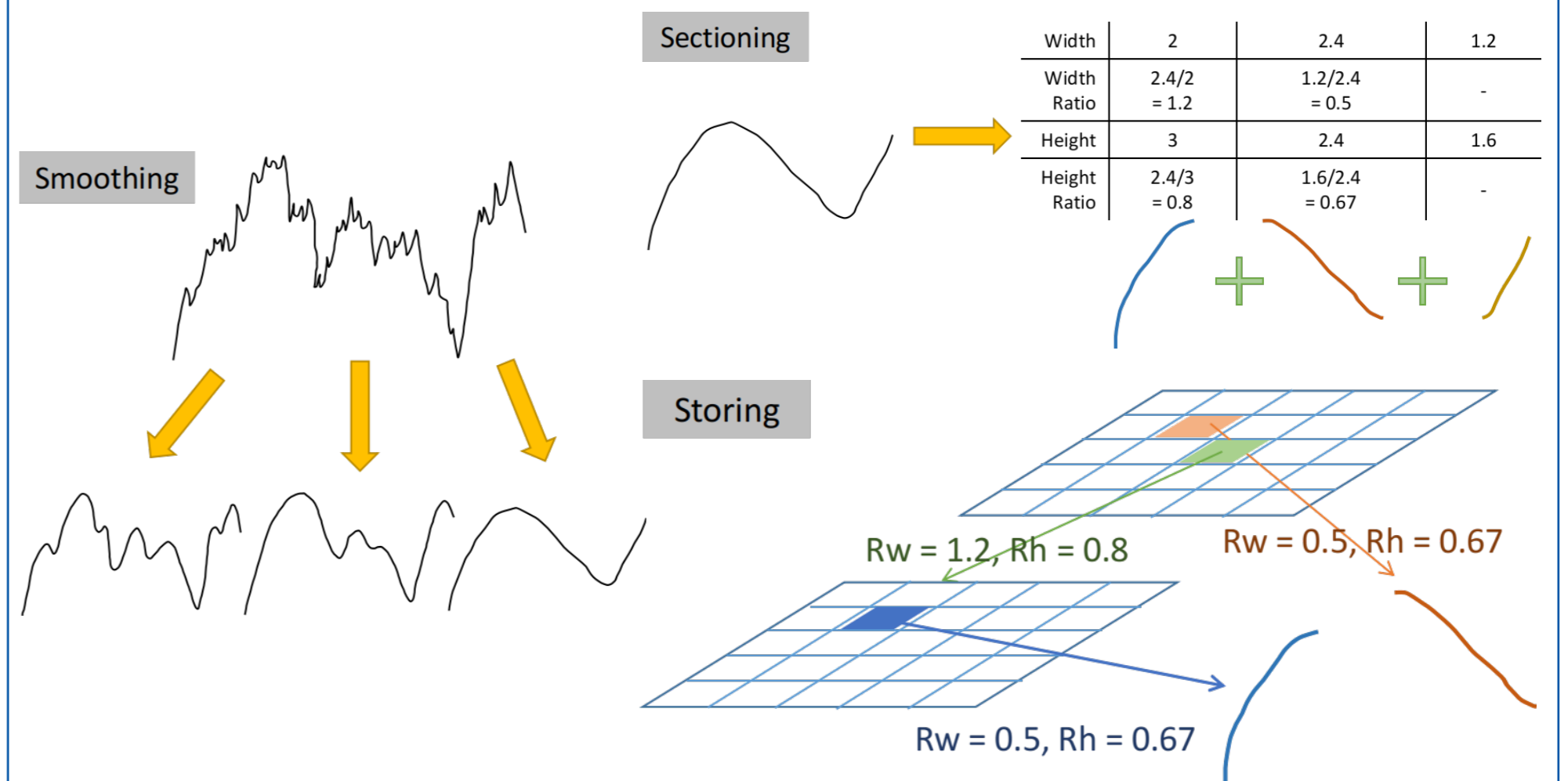
SCALE-FREE MATCHING

- A section is defined as a series of monotonically increasing or decreasing data sequence
- Two data series are similar if width and height ratios between all adjacent sections are within a certain bound



INDEXING

- Smoothing, Sectioning and Labelling, Storing in a tree based on width and height ratios of neighboring sections.



QUERYING

- Blends query sketching with processing of partial query
- Performs index tree traversal, retrieval and further pruning
- Online index containing query information and traversal status retained throughout the process
- Matches the head and tail ends after query formulation is complete

PERFORMANCE

