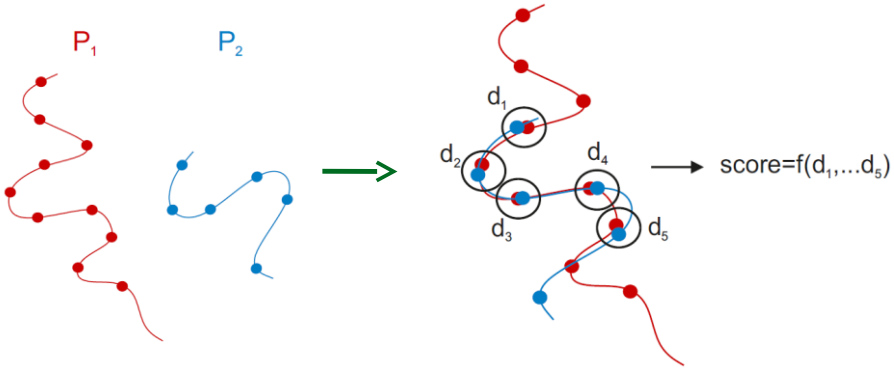
SIRET research group

Department of Software Engineering
Faculty of Mathematics and Physics
Charles University in Prague

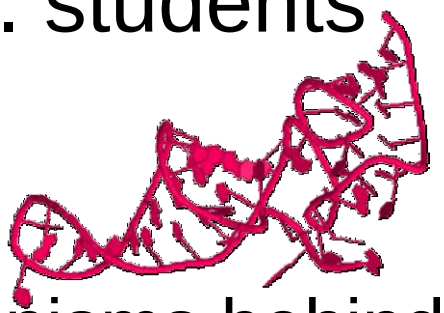
http://siret.ms.mff.cuni.cz

SIRET = **Si**milarity **Ret**rieval

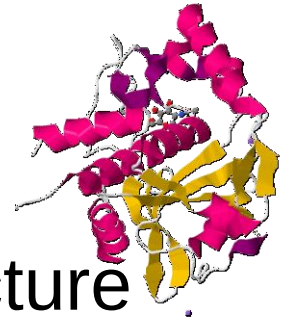


Bioinformatics & Cheminformatics

- David Hoksza and 3 Ph.D. students
  - hoksza@ksi.mff.cuni.cz
- RNA Function Elucidation



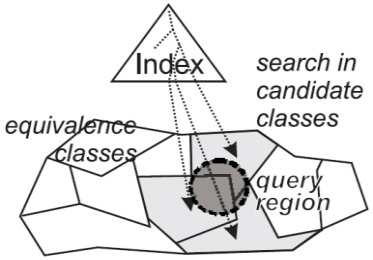
  - understanding of the mechanisms behind RNA function
- Protein Structure Similarity



  - funcionality is based on their structure
- Tandem Mass Spectrometry, Protein Sequences Identification

General methods of indexing similarity

- Tomáš Skopal, Jakub Lokoč and 3 Ph.D. students
  - skopal@ksi.mff.cuni.cz lokoc@ksi.mff.cuni.cz
- Indexing multimedia databases for content-based retrieval
  - efficient and effective similarity search
  - multimedia exploration
- Similarity analytics
  - support for general modeling of similarity



Real-time Exploration of Multimedia Collections

Juraj Moško

Supervised by Tomáš Skopal

Department of Software Engineering

Faculty of Mathematics and Physics

Charles University in Prague

mosko@ksi.mff.cuni.cz

Multimedia retrieval

- Content based similarity search
- Range and nearest neighbor queries
- Traditional scenario – user-defined query by example

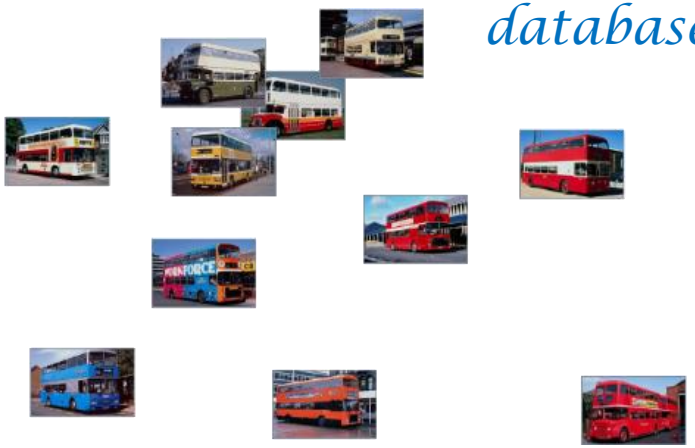
Query by Example

Query object

Query results

Multimedia Exploration

Visualized database



Multimedia retrieval

Query by Example

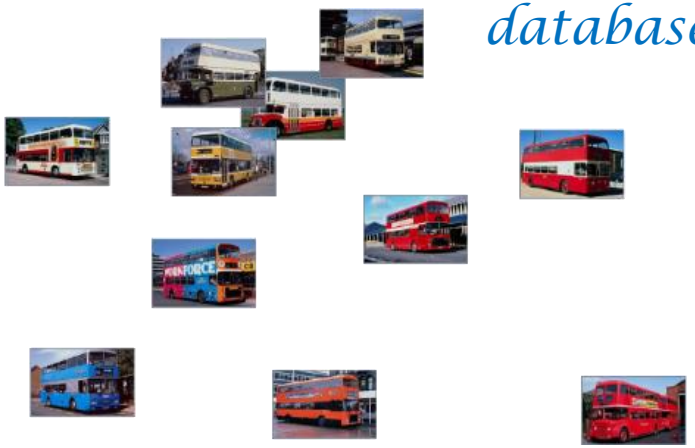
Query object

Query results

- users know what they are searching
- one simultaneous query
- no explicit need for immediate results

Multimedia Exploration

Visualized database

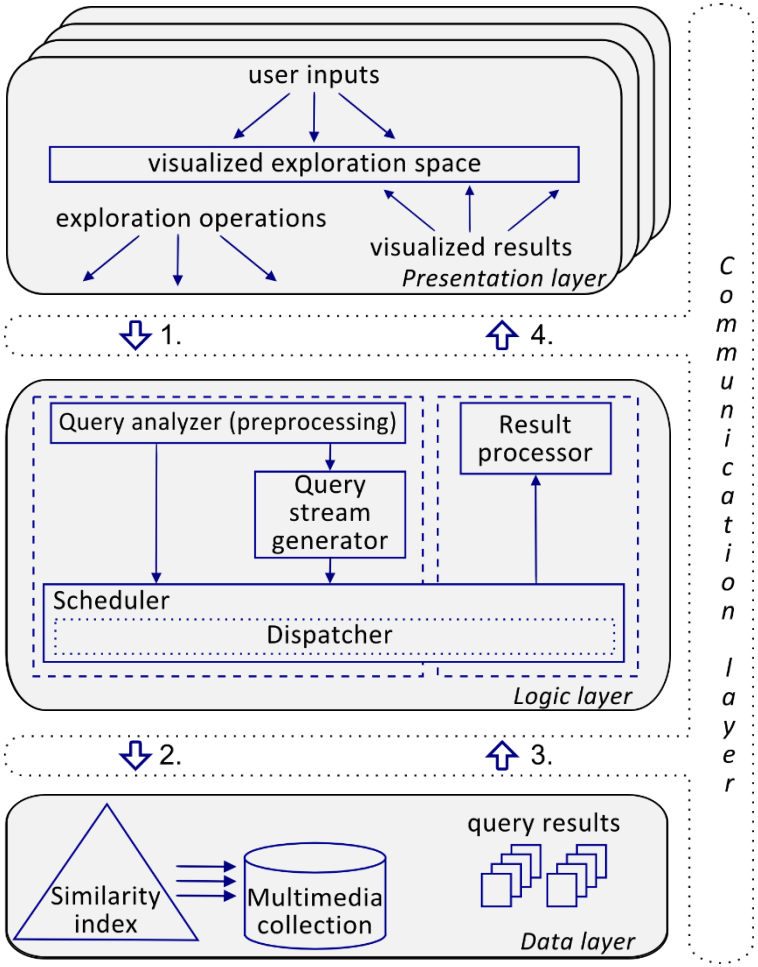


- users just explore
- multiple queries at once
- continuous interactive process, necessity for immediate responses

Multimedia exploration

- user-controlled (interactive) process of viewing and browsing

- RTExp System

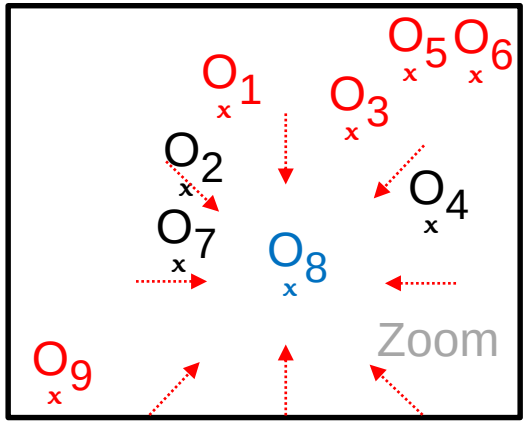


Real-time ...

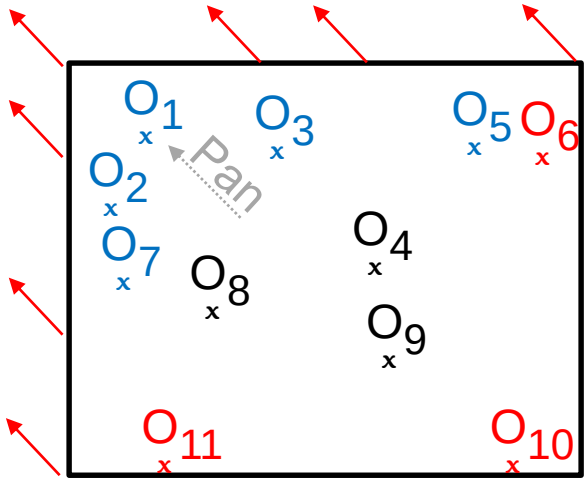
- Motivation
  - Smooth exploration process
  - Query real-time guarantee in arbitrary conditions
- Idea
  - Query evaluation is terminated whenever it is necessary
  - Partial results are returned

... Exploration

- Exploration operations
  - Zoom

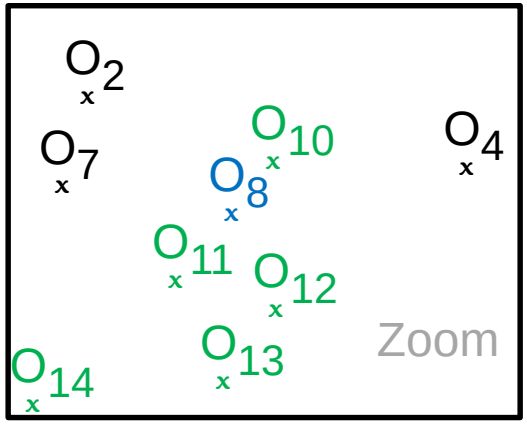


  - Pan

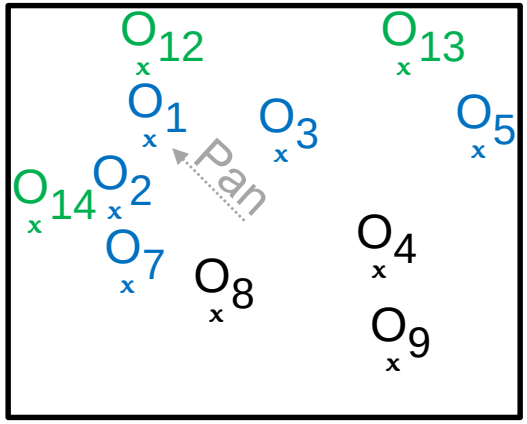


... Exploration

- Exploration operations
  - Zoom

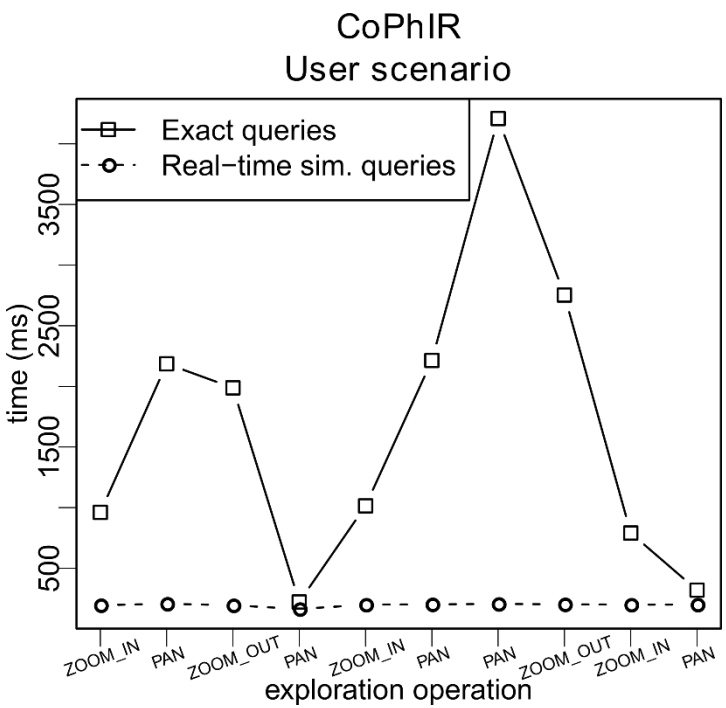


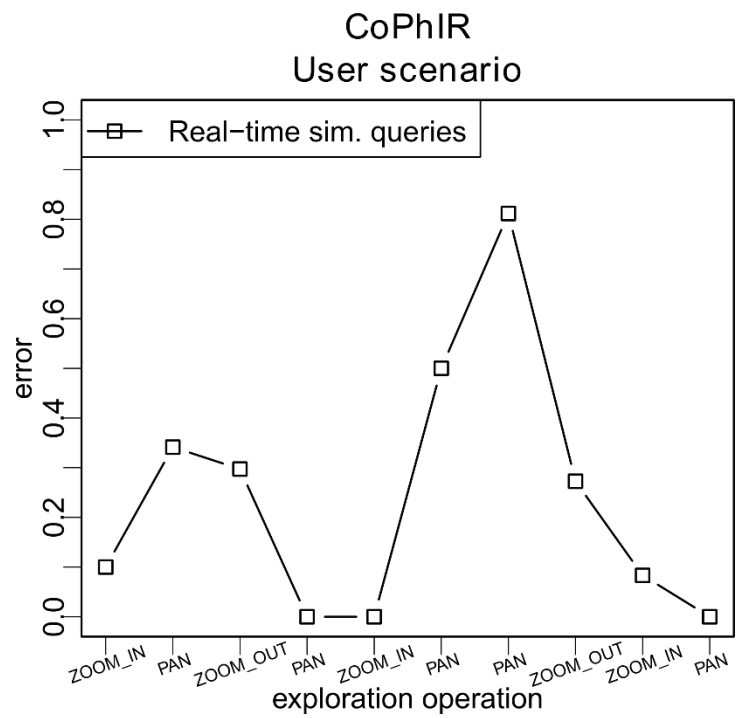
  - Pan



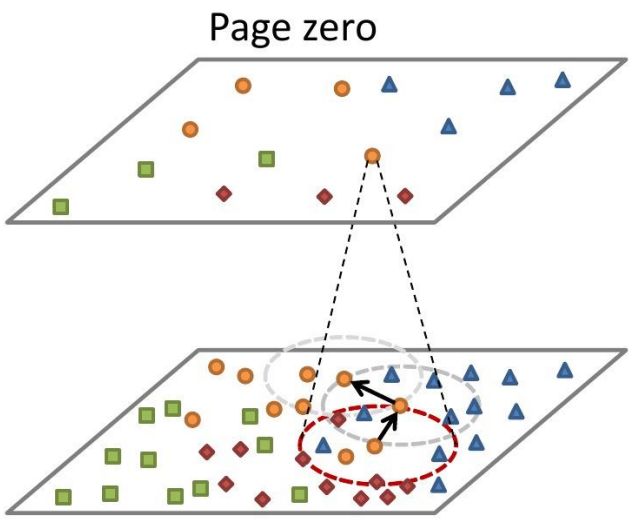
Experiments

- Simulating user behavior
- Limit 200 ms per exploration operation
- Visualized 50 objects from 1 million that reside in multimedia collection

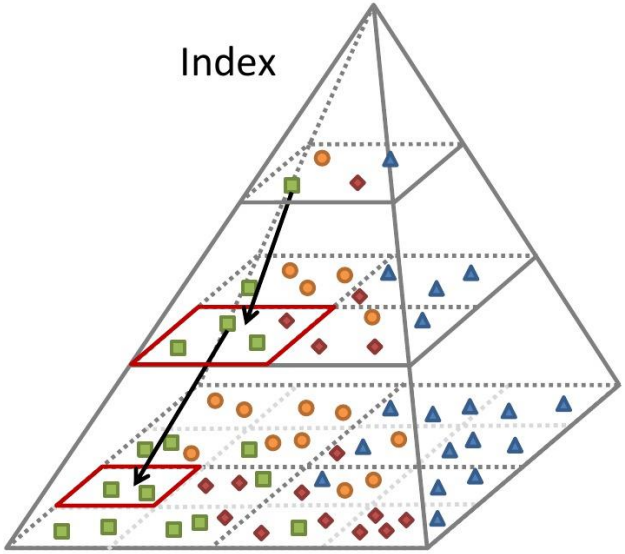




Index support



a)

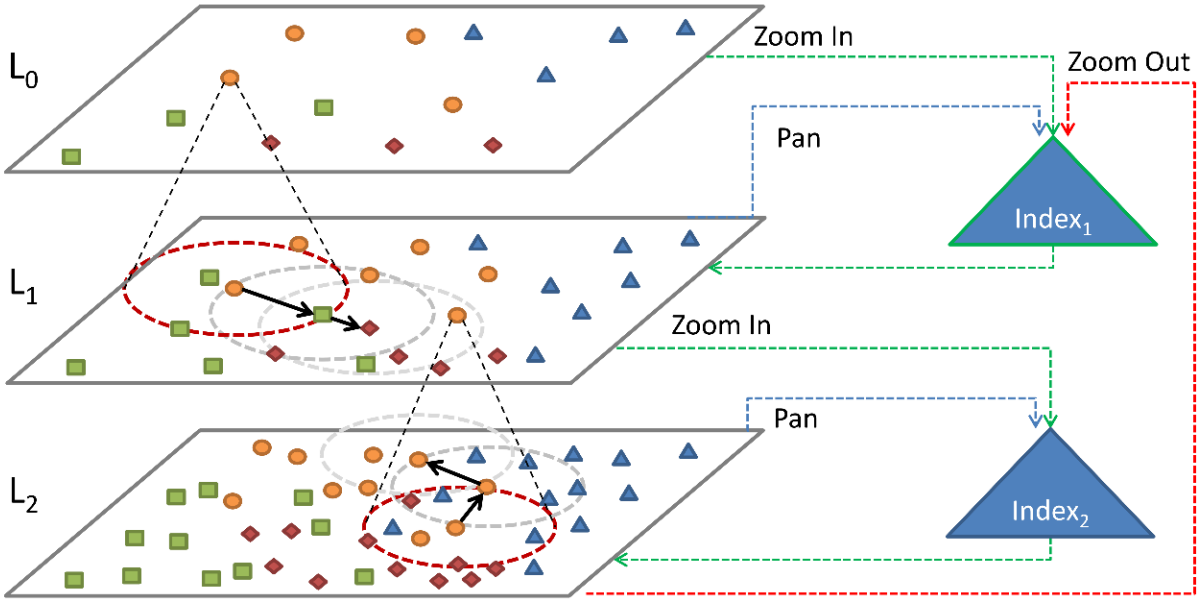


b)

- iterative querying
  - queries are always used to search the whole database
  - either too general or too specific view
- iterative browsing
  - too expensive to build

Index support

- Multilayer exploration structure
  - each object on layer L_i is also on layer L_{i+1}
  - designed for supporting exploration operations
  - Zoom In / Out – kNN query in lower / higher layer
  - Pan – kNN queries in the same layer, with exclusion of some objects



Current problems

- Index
  - Selection of objects for layers
  - Number of layers / number of objects in layers

- Operations
  - on arbitrary point in visualized space, not just on specific object (i. e. on the “white space” between images)
  - Pan
    - How to translate the size of panning?

Future

- Continuous query evaluation
  - Updating the visualized space with partial results while query streams are evaluating in the background
- Improvement of the system precision
  - Query analyzer

Future (2)

- Support for visualization
  - reduction of dimension
  - a distance between visualized objects can be the only known information
  - non-overlapping user interface

Thank you