INFO 290T

Human-Centered Data Management

Gestural Query Specification

Thoughts on Paper?

- Interface?
- Evaluation?
- Writing?

Key Ideas

- Touch-based interfaces to manipulate data
- Pros: people not used to databases can manipulate data

Questions

- Was the user study "fairly" conducted? What else would you have done?

Questions

- Was the user study "fairly" conducted? What else would you have done?
  - GestureQuery may be easy to specify small q but may be hard to specify more complex ones
  - Discoverability only on join: not clear if more complex queries are discoverable

Questions

- GestureQuery
  - Provides feedback as queries are being composed. Can there be issues?

Questions

- GestureQuery
  - Provides feedback as queries are being composed. Can there be issues?
    - Will only work for small tables. What about predicate pushdown after a cross-product?
    - Going against the declarative nature of databases if query results are composed iteratively.

Alternatives: Excel

- When would GestureQuery be better than Excel?

- When would they be worse?

Alternatives: Excel

- When would GestureQuery be better than Excel?
  - More tactile and therefore more intuitive
  - Relational operations not supported by Excel
    - Joins not supported
    - Primarily formulae rather than relational expressions
- When would they be worse?
  - Plotting charts
  - Looking at all your data at once

Alternatives: Visual Analytics Tools

- When would GestureQuery be better than Tableau?

- When would it be worse?

Alternatives: Visual Analytics Tools

- When would GestureQuery be better than Tableau? When would it be worse?
  - When selection, creation of new tables, joins, is key rather than aggregate queries

Alternatives: Query By Example

Moshe Zloof, IBM, '70s

Sailors(*sid:* `integer`, *sname:* `string`, *rating:* `integer`, *age:* `real`)

Boats(*bid:* `integer`, *bname:* `string`, *color:* `string`)

Reserves(*sid:* `integer`, *bid:* `integer`, *day:* `dates`)

Sailors	*sid*	*sname*	*rating*	*age*
P.			10	

Print all sailor tuples with rating 10

Sailors	*sid*	*sname*	*rating*	*age*
		P.	P.AO(2)	P.AO(1)

Print names, ratings, ages of all sailors ordered by a, r

Sailors	*sid*	*sname*	*rating*	*age*
	_Id	P._S		

Reserves	*sid*	*bid*	*day*
	_Id		

Print all sailors with a reservation

Sailors	*sid*	*sname*	*rating*	*age*
	_Id			> 25

Reserves	*sid*	*bid*	*day*
	_Id	_B	'8/24/96'

Boats	*bid*	*bname*	*color*
	_B	Interlake	P.

Print colors of interlake boats reserved by sailors on 8/24/94, with age > 25

Alternatives: Query By Example

Sailors	*sid*	*sname*	*rating*	*age*	
			G.P.	_A	P.AVG._A

Sailors	*sid*	*sname*	*rating*	*age*
		P.		< 30
		P.		> 20

Sailors	*sid*	*sname*	*rating*	*age*
	_Id	P.		< 30
	_Id			> 20

- Print average age grouped by rating
- Print sailors <30 or < 20
- Print sailors <30 and > 20

Alternatives: Query By Example

Advantages, Disadvantages?

Alternatives: Query By Example

Advantages, Disadvantages?

Advantages: more powerful, requires less visual manipulation, few keystrokes

Disadvantages: less “fun”? Less “intuitive”?

Alternatives: Keyword Search in DB

Key Idea of a **Data Graph**: Captures **relationships and their strengths**, among data and metadata items

Nodes

- Classes, tables, attributes, field values
- May be **weighted** – representing authoritativeness, quality, correctness, etc.

Edges

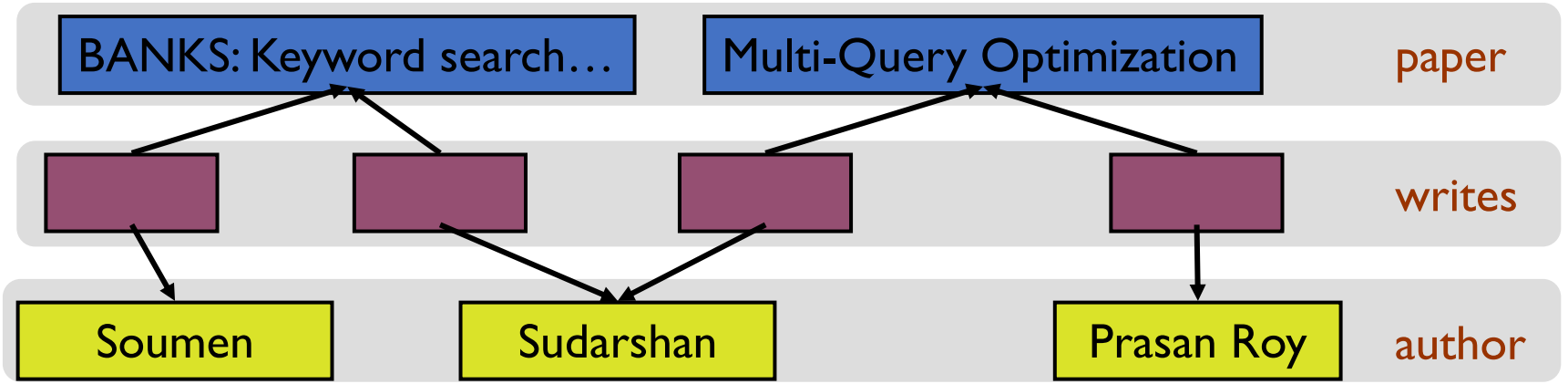
- is-a and has-a relationships, foreign keys, hyperlinks, record links, possible joins, …
- May be **weighted** – representing strength of the connection, probability of match, etc.

Alternatives: Keyword Search in DB

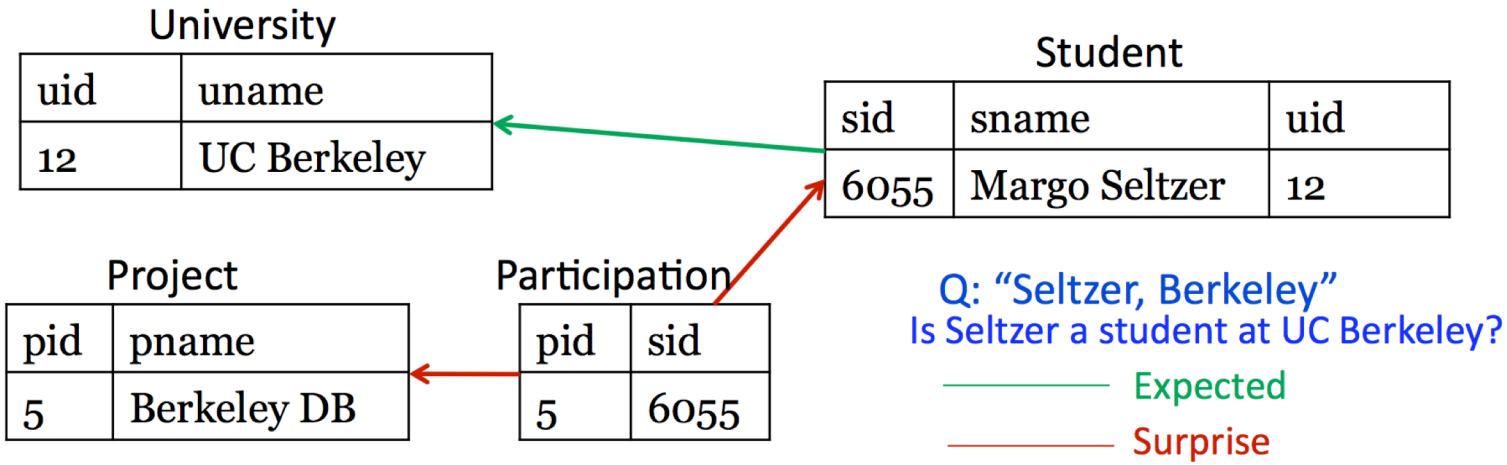
- Queries are expressed as sets of keywords
- We match keywords to nodes, then seek to find a way to "connect" the matches in a **tree**
- The lowest-cost tree connecting a set of nodes is called a **Steiner tree**
  - Formally, we want the **top-k Steiner trees**
  - **NP-Hard**

Examples (from original papers)

May be one way of reaching all keywords

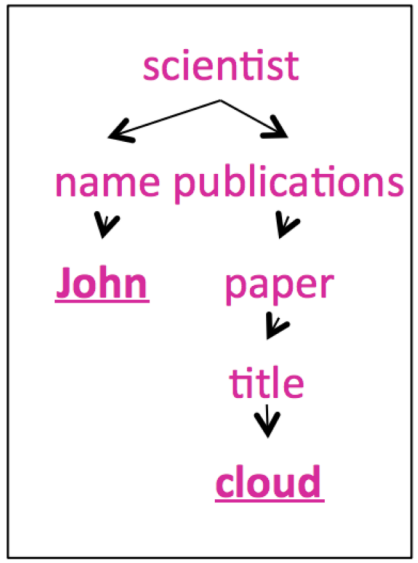


Or multiple ways

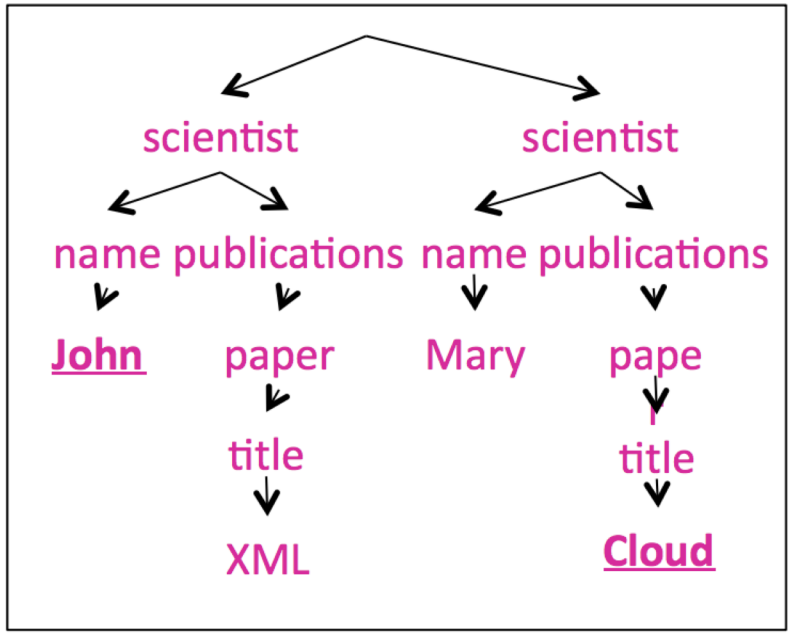


How to rank these?

Simple rule: Longer paths are worse



High Rank



Low Rank