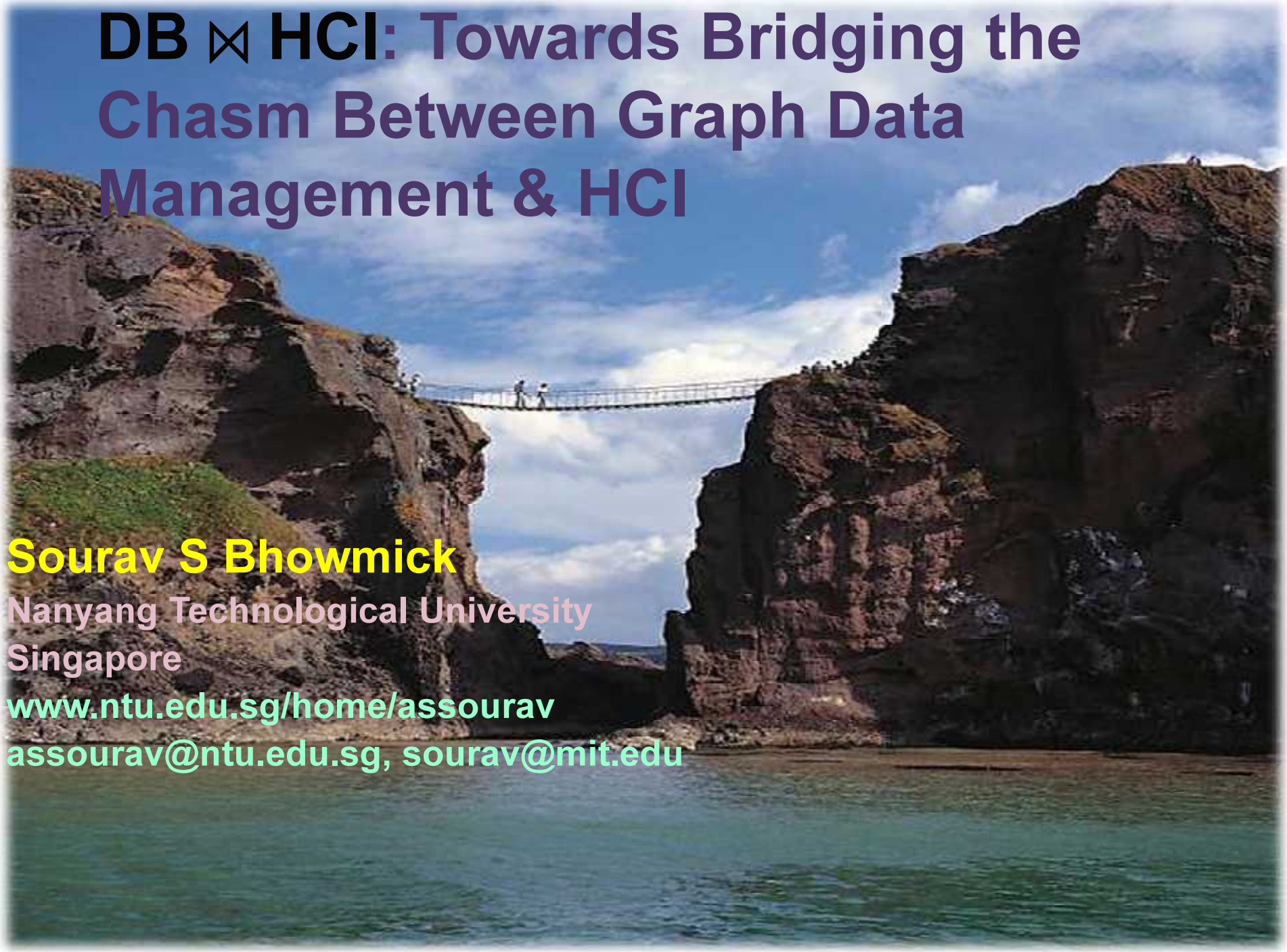
DB ⋈ HCI: Towards Bridging the Chasm Between Graph Data Management & HCI



Sourav S Bhowmick
Nanyang Technological University
Singapore
www.ntu.edu.sg/home/assourav
assourav@ntu.edu.sg, sourav@mit.edu

The World has Changed

NANYANG TECHNOLOGICAL UNIVERSITY

Then

- Data is generated in companies
- Resides in companies
- Used by companies
- DB-literate users

Now

- Data is generated by everyone
- Resides everywhere
- Used by everyone
- Non-DB literate users

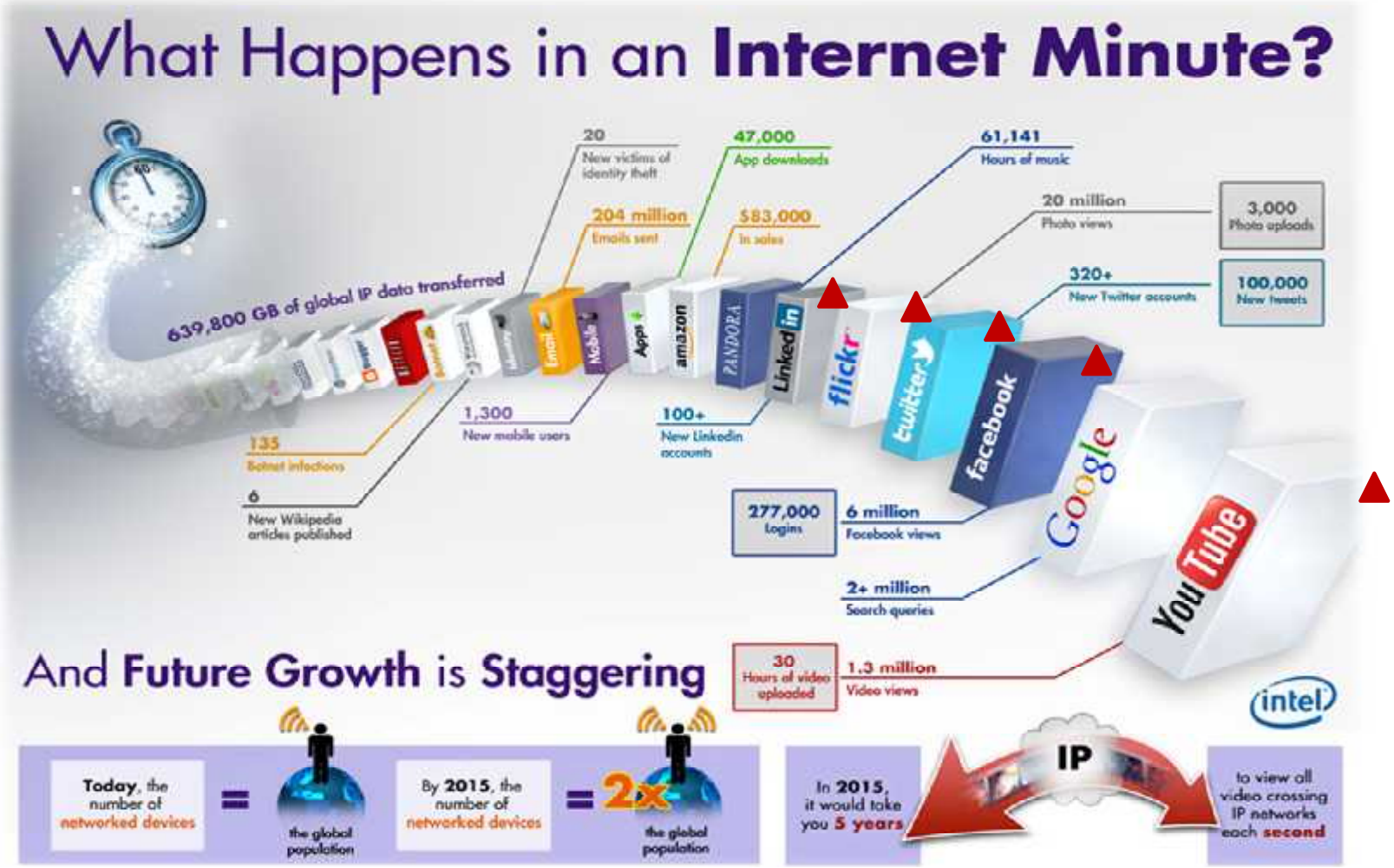
The World Has Changed

NANYANG TECHNOLOGICAL UNIVERSITY

"The old computing was about what computers could do; the new computing is about what people can do"

Ben Shneiderman (2002)

Users are willing to share information online



Source: http://www.intel.sg/content/www/us/en/communications/internet-minute-infographic.html

The Road Ahead

Complexity has Increased

Data streams

Parallel & Distributed DB

Cloud data management

DB+ IR

mobile data management

Object-Oriented Model

Object 1: Maintenance Report / Object 1 Instance

Date	01-12-01
Activity Code	24
Route No.	I-95
Daily Production	2.5
Equipment Hours	6.0
Labor Hours	6.0

Object 2: Maintenance Activity

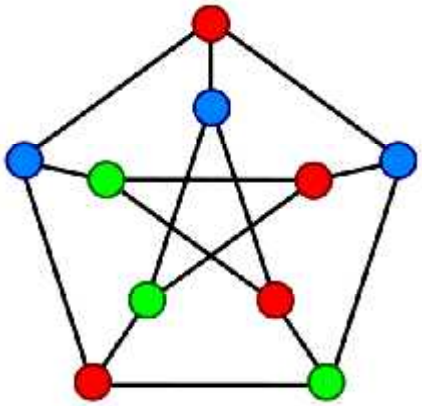
Activity Code
Activity Name
Production Unit
Average Daily Production Rate

OO DB

Key	Product ID	Price ($)	Prob.
a_1	a	120	0.7
a_2	a	80	0.3
b_1	b	110	0.6
b_2	b	90	0.4
c_1	c	140	0.5
c_2	c	110	0.3
c_3	c	100	0.2
d_1	d	10	1

Probabilistic DB

semistructured
data management



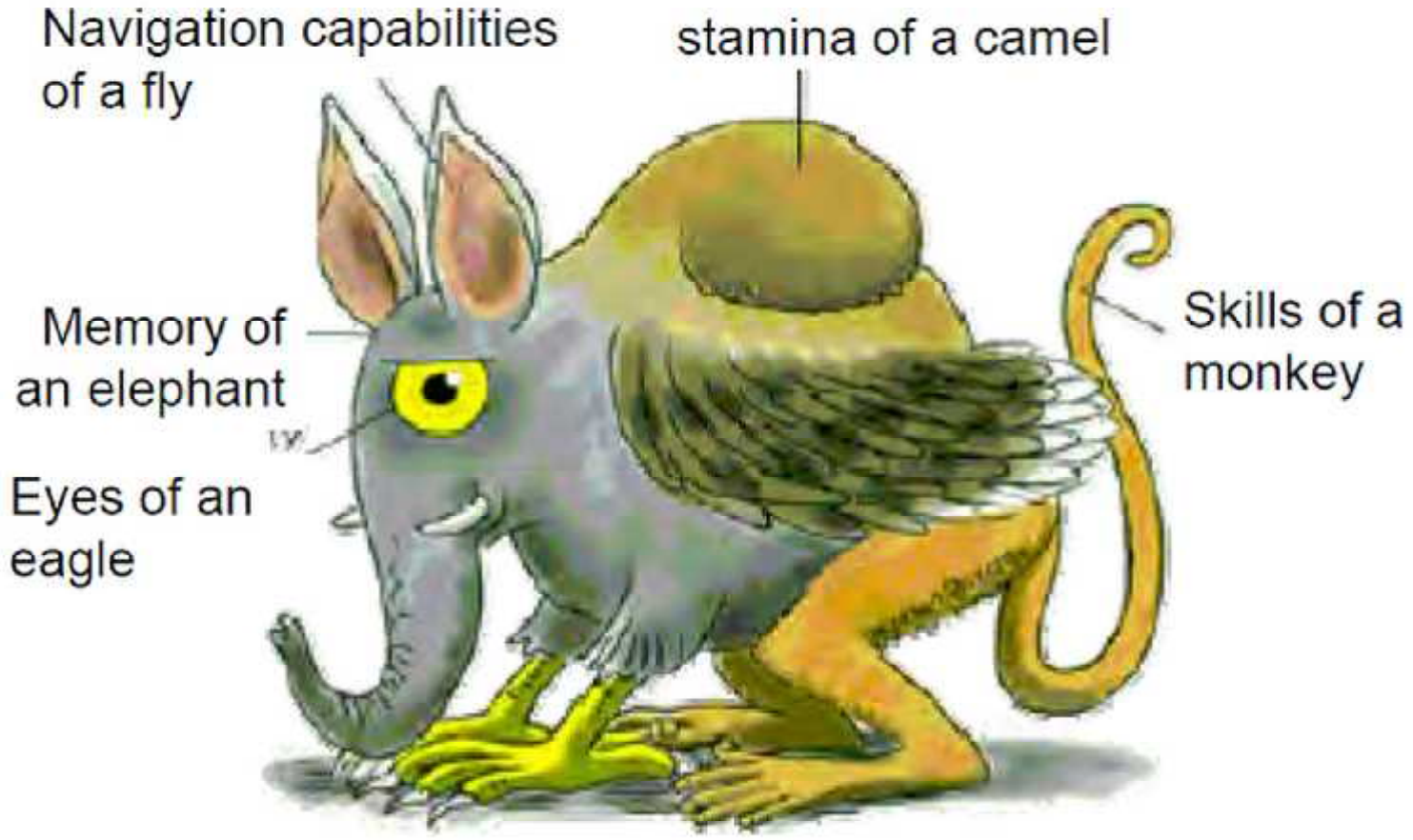
graph
data management

DB Community's Love Affair

SPEED, SPEED, SPEED
SCALE, SCALE, SCALE
FUNCTIONALITIES!!

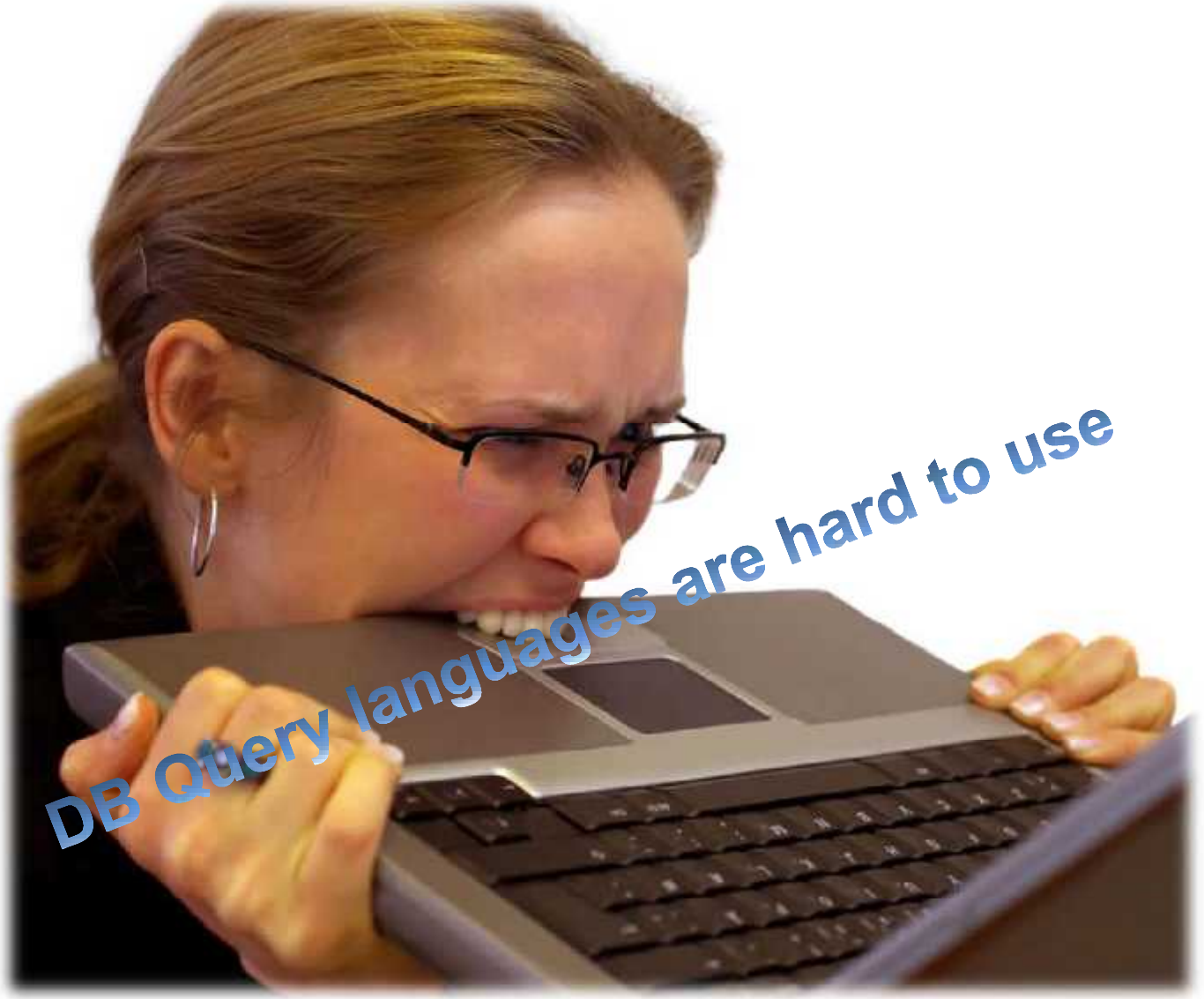
Assumptions Made By (Most) DM Systems

NANYANG TECHNOLOGICAL UNIVERSITY



The Real User

NANYANG TECHNOLOGICAL UNIVERSITY



Query Formulation using SPARQL

NANYANG TECHNOLOGICAL UNIVERSITY

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rss: <http://purl.org/rss/1.0/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

SELECT ?title ?known_name ?link

FROM http://planetrdf.com/index.rdf
FROM NAMED <phil-foaf.rdf>
WHERE {
        GRAPH <phil-foaf.rdf> {
        ?me foaf:name "Phil McCarthy".
        ?me foaf:knows ?known_person .
        ?known_person foaf:name ?known_name .
        } .
        ?item dc:creator ?known_name .
        ?item rss:title ?title .
        ?item rss:link ?link .
        ?item dc:date ?date.
    }
ORDER BY DESC[?date] LIMIT 10
```

Reality Check!

NANYANG TECHNOLOGICAL UNIVERSITY

Reality

> Thirty years of research on query languages can be summarized by: we have moved from SQL to XQuery. At best we have moved from one declarative language to a second declarative language with roughly the same level of expressiveness. It has been well documented that end users will not learn SQL; rather SQL is notation for professional programmers.

The Lowell Database Research Self-Assessment,
Communication of the ACM (May 2005)

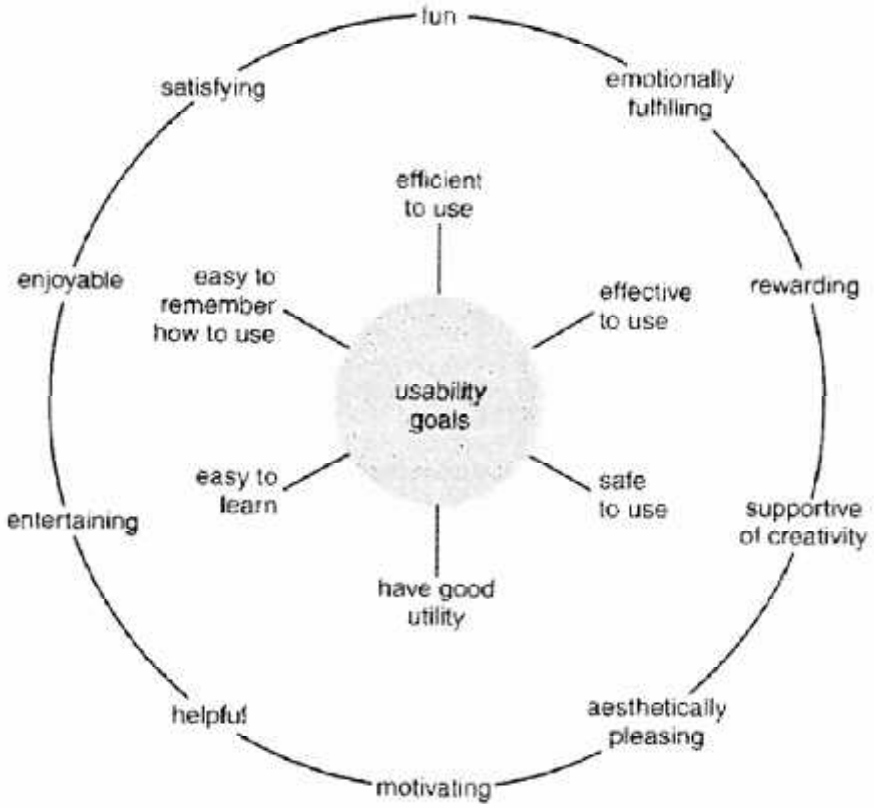
Usability [Preece et al.]

What is it?

How well users can use the system's functionality

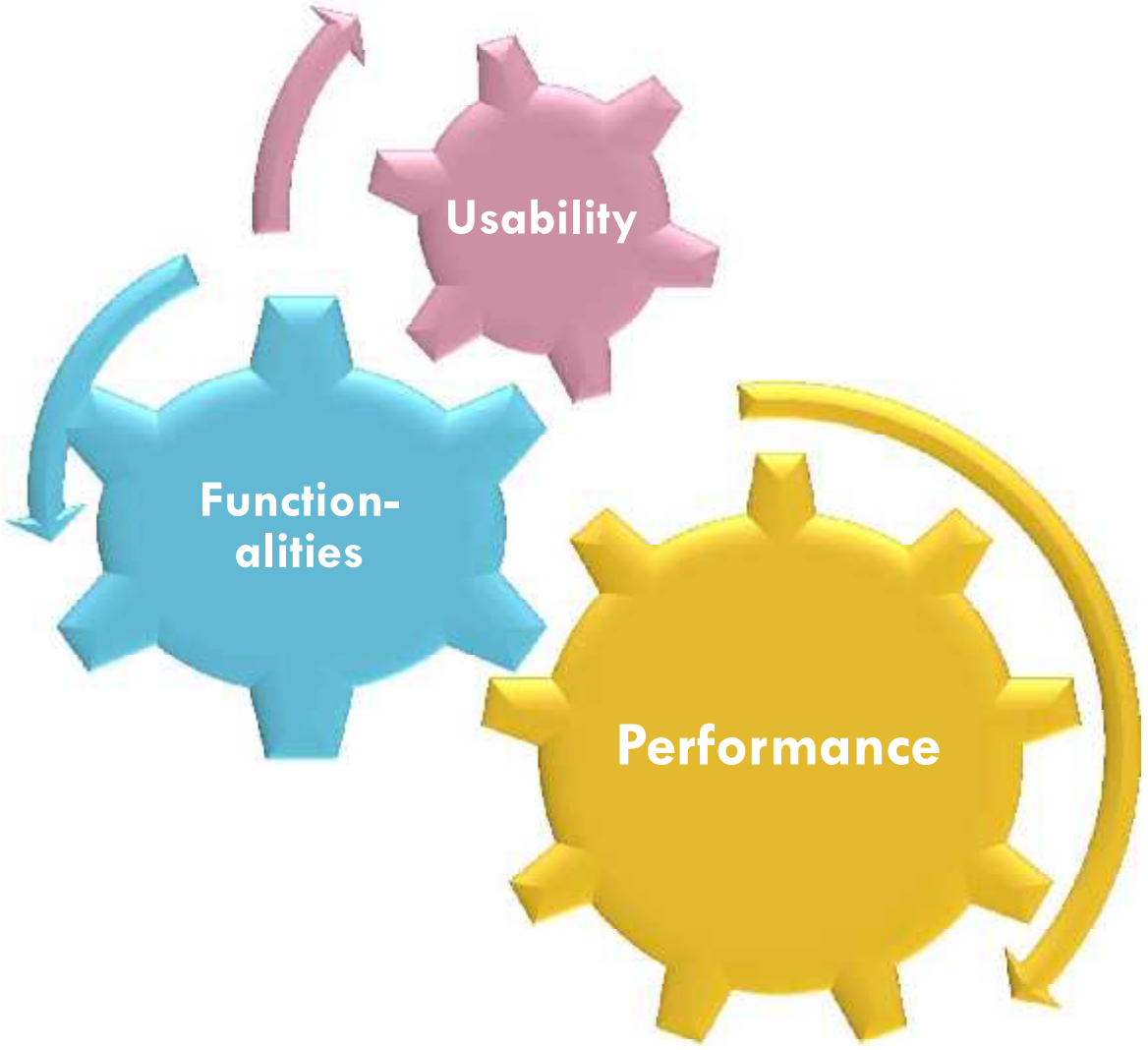
Dimensionality

- Learnability: is it easy to learn?
- Efficiency: once learned, is it fast to use?
- Memorability: is it easy to remember what you learned?
- Errors: are errors few and recoverable?
- Satisfaction: is it enjoyable to use?



DB Research Since 2006

NANYANG TECHNOLOGICAL UNIVERSITY



NANYANG TECHNOLOGICAL UNIVERSITY

Next..

What is Human Computer Interaction (HCI)?

NANYANG TECHNOLOGICAL UNIVERSITY

"concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them"

ACM SIGCHI (1992)

Usability and HCI

NANYANG TECHNOLOGICAL UNIVERSITY

Studying and improving usability is part of Human-Computer Interaction (HCI).

Usability and good UI design are closely related

“The interface is the system.”

NANYANG TECHNOLOGICAL UNIVERSITY

View

- Interface provides/conveys the ***only*** view of the “underlying” system

Perception

- Usable software sells better

Superficiality

- Users blame themselves for UI failings
- People who make buying decisions are not always end-users

Should and Must Do It Right

NANYANG TECHNOLOGICAL UNIVERSITY

Ben Shneiderman

"*Always should* have "good" interfaces. Computing time (power) is getting cheaper but users' time isn't.."

"If the user can't use it, it doesn't work"

Susan Dray,
Distinguished Engineer of ACM

The World of Cool Interfaces!

NANYANG TECHNOLOGICAL UNIVERSITY

Role of visual interfaces

> A picture is worth a thousand words. An interface is worth a thousand pictures
>
> Ben Shneiderman, 2003

Easy-to-use "dummy" visual interfaces are key to the spread of data management tool to wider community.

Lessons from HCI: Schneiderman's 8 Golden Rules

- Strive for consistency.
- Give shortcuts to the user.
- Offer informative feedback.
- Make each interaction with the user yield a result.
- Offer simple error handling.
- Permit easy undo of actions.
- Let the user be in control.
- Reduce short-term memory load on the user.

How do these "rules" impact data management?

Summary

"A USER INTERFACE IS LIKE A JOKE. IF YOU HAVE TO EXPLAIN IT, IT'S NOT THAT GOOD"

Anonymous, LinkedIn

NANYANG TECHNOLOGICAL UNIVERSITY

The Chasm

Our Affinity to Visual Languages

NANYANG TECHNOLOGICAL UNIVERSITY

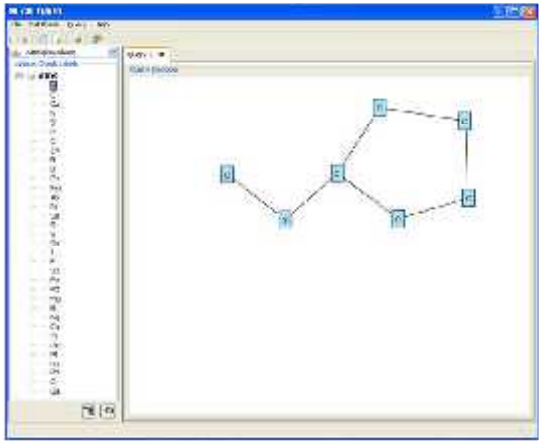
Cave painting

Cuneiform

Hieroglyphics

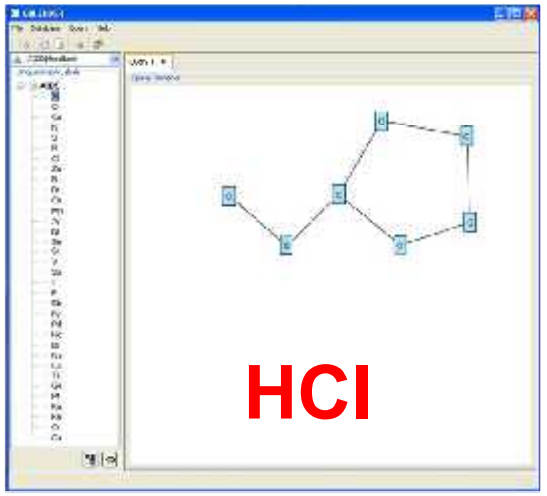
Coats of arms

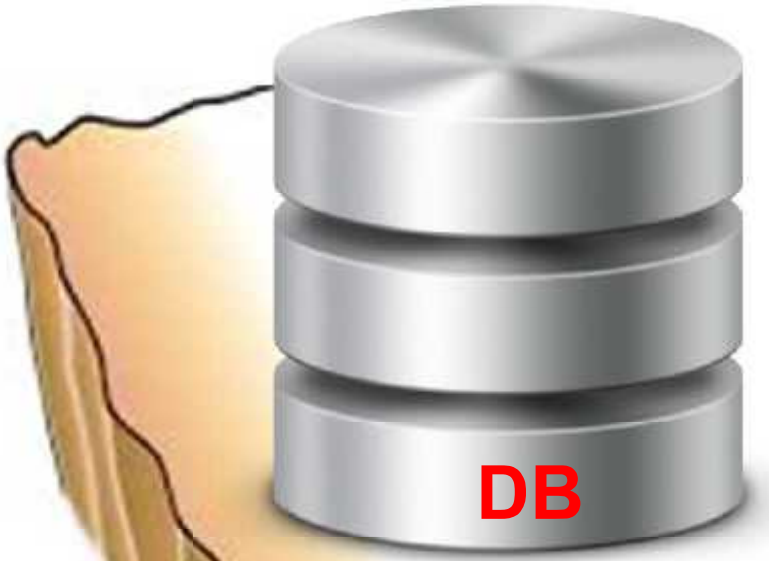
Emoji



Visual query language

Separate Ways for 40 years





NANYANG TECHNOLOGICAL UNIVERSITY

Data Management Research

Querying Graph Databases

Query Formulation

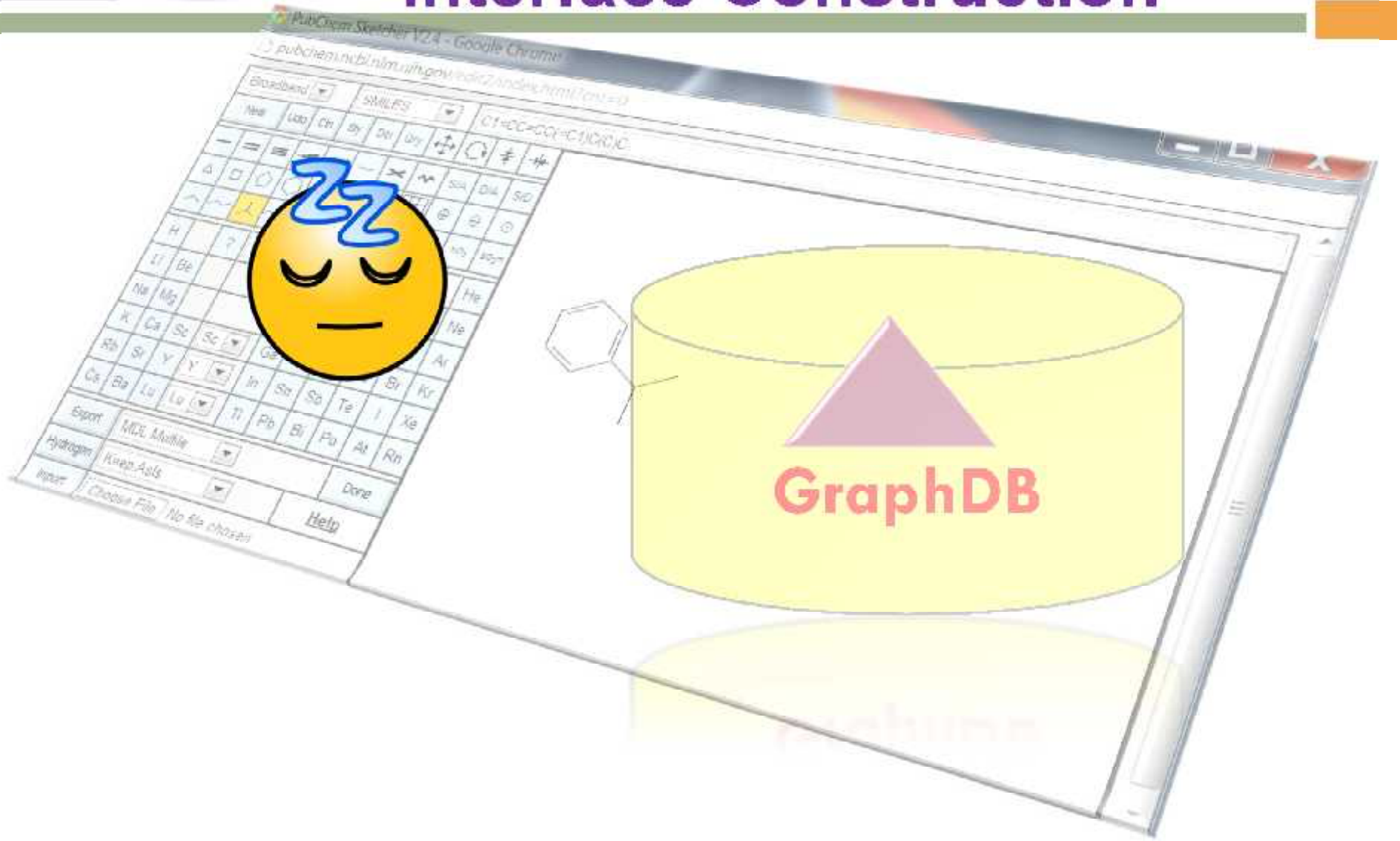
- Formal query language
- GraphQL, SPARQL, PQL

Query Processing

- Efficient algorithms and optimization techniques to process queries “quickly”
- FG-Index, Tree-PI, SSI, C-Tree, TALE

Classical Visual Query Interface Construction

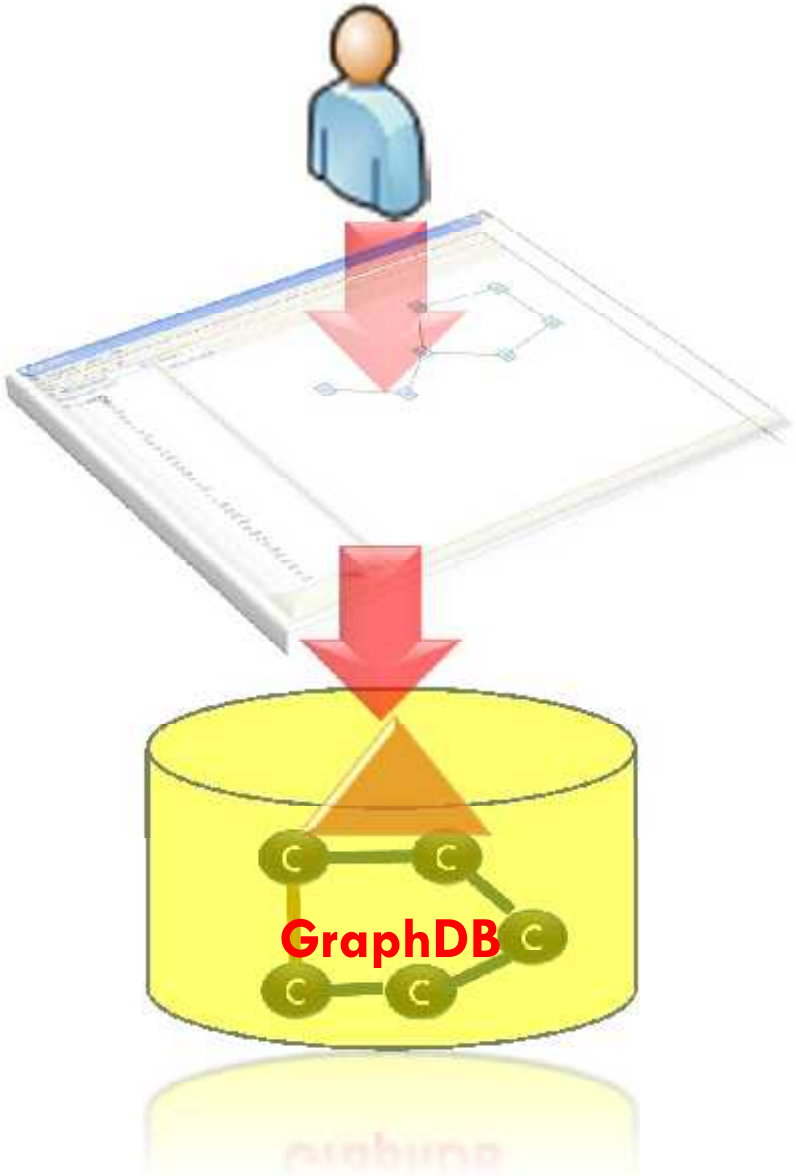
NANYANG TECHNOLOGICAL UNIVERSITY

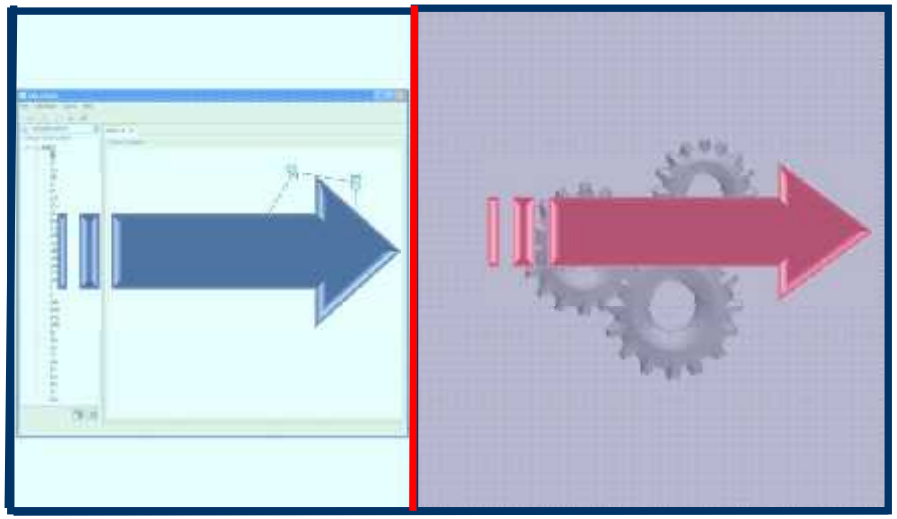


Classical Visual Query Interface Construction

NANYANG TECHNOLOGICAL UNIVERSITY

Classical Visual Querying Paradigm





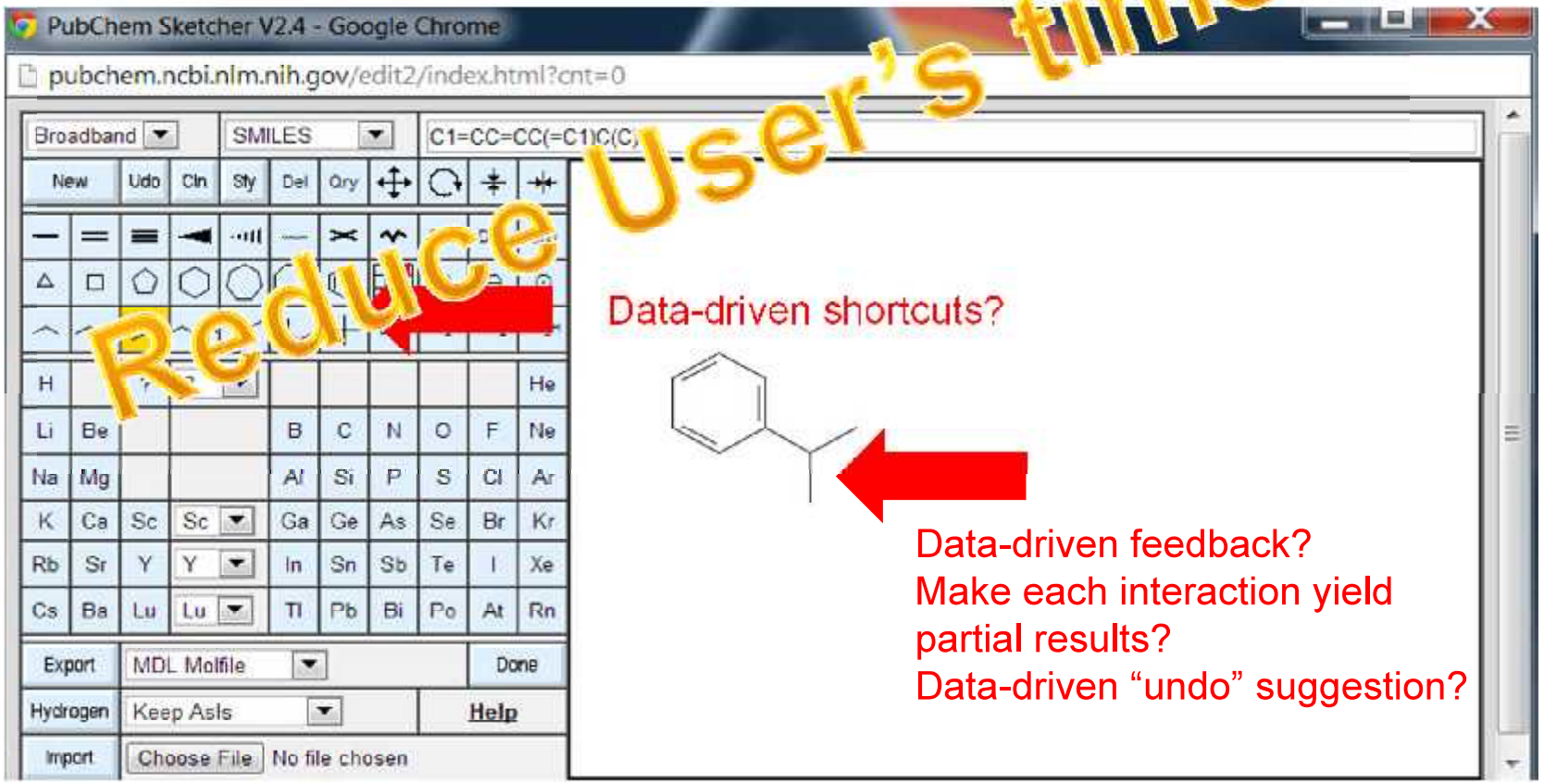
Query formulation Query processing

time

What Happens if we Bridge the Chasm?

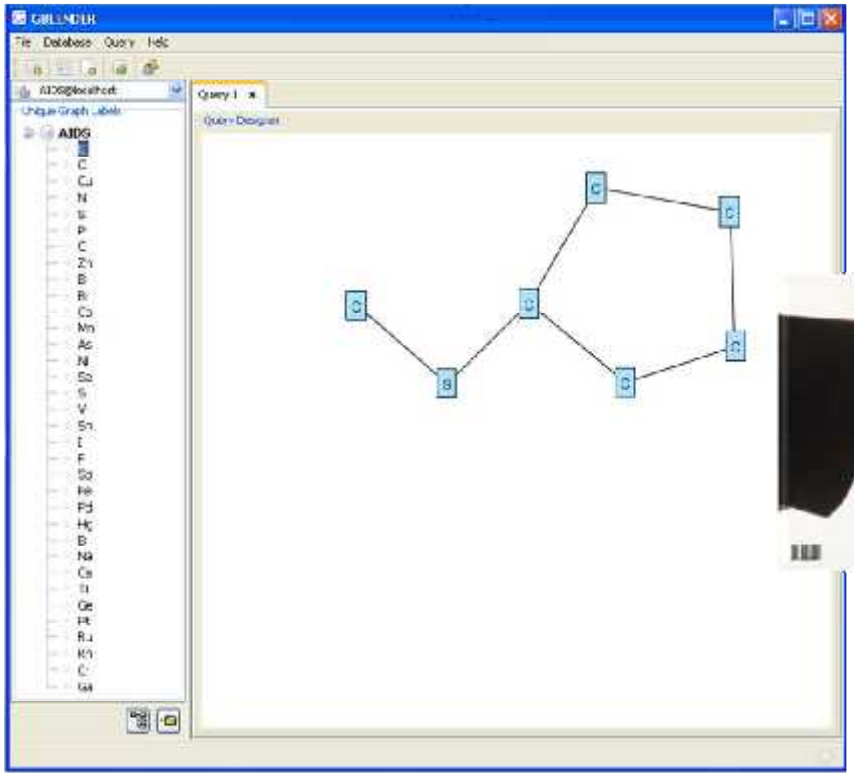
NANYANG TECHNOLOGICAL UNIVERSITY

- Strive for consistency.
- Give shortcuts to the user.
- Offer informative feedback.
- Make each interaction with the user yield a result.
- Offer simple error handling.
- Permit easy undo of actions.
- Let the user be in control.
- Reduce short-term memory load on the user.



HCI-aware Data Management

NANYANG TECHNOLOGICAL UNIVERSITY

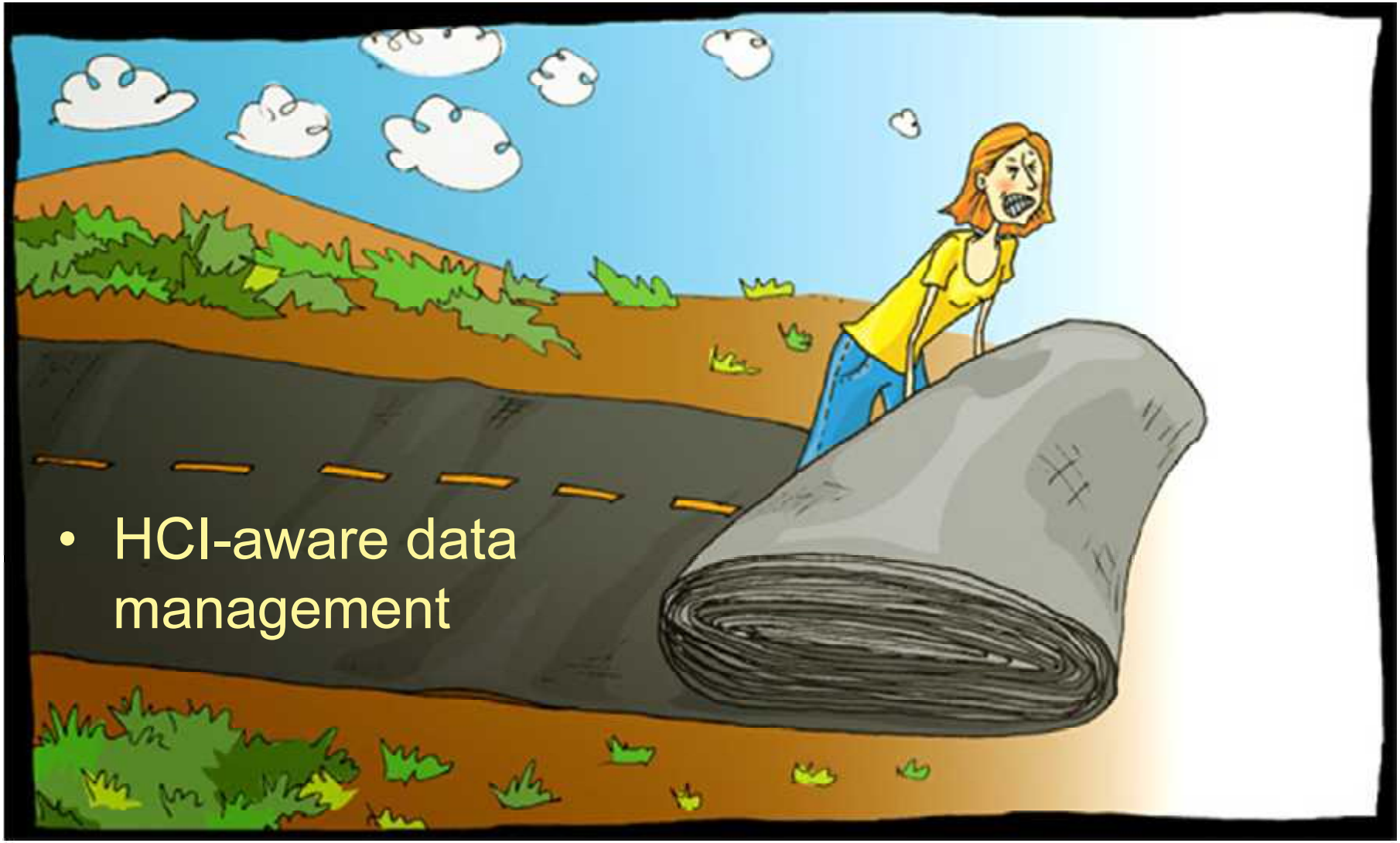


HCI

DB

NANYANG TECHNOLOGICAL UNIVERSITY

Next..



Broad Goals

NANYANG TECHNOLOGICAL UNIVERSITY

- Data-driven construction of visual query interface
- Interleave visual actions on the GUI and query processing
- Data-driven feedback and guidance to users centered around usability
- Minimize users' time

Functional Architecture

Data-driven Visual Query Interface Manager

Visual Action-aware Query Performance Simulator

Visual Action-aware Query Processing Engine

Visual Action-aware Indexer

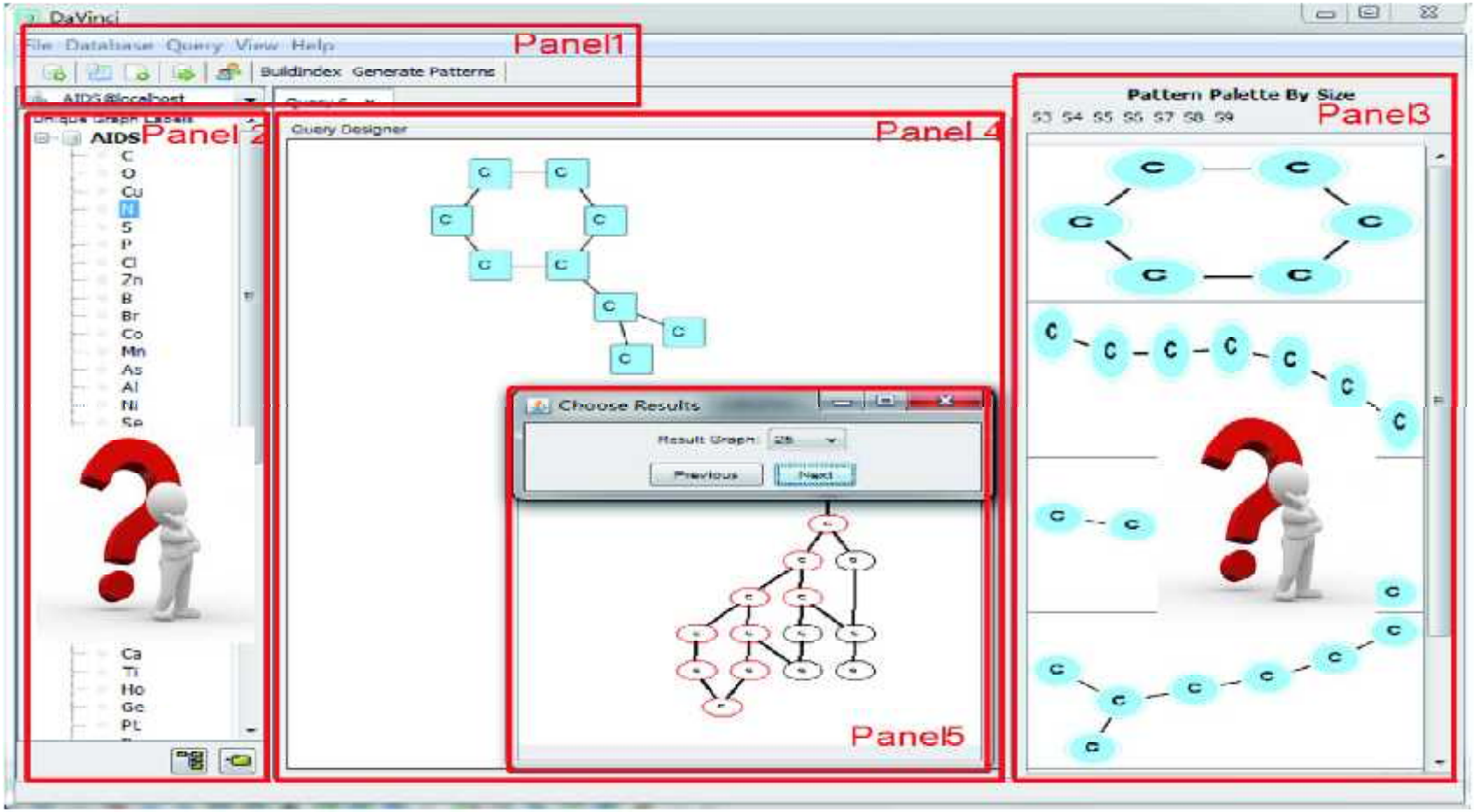
Graph DB

Data-driven Visual Interface Management

NANYANG TECHNOLOGICAL UNIVERSITY

Goal 1

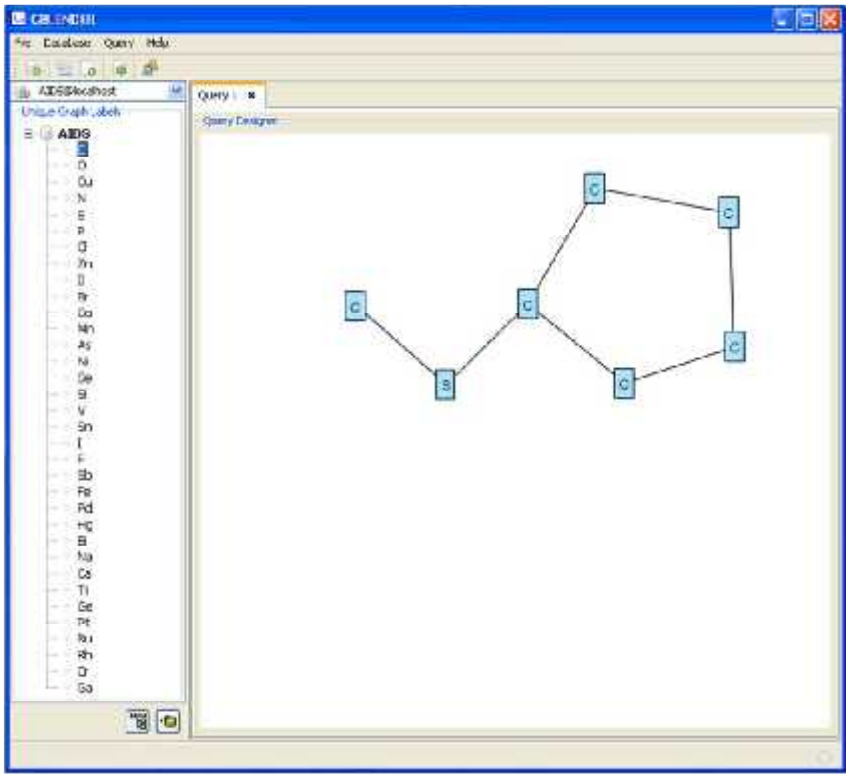
- Data-driven construction of visual query interface



Challenges

Panel 2

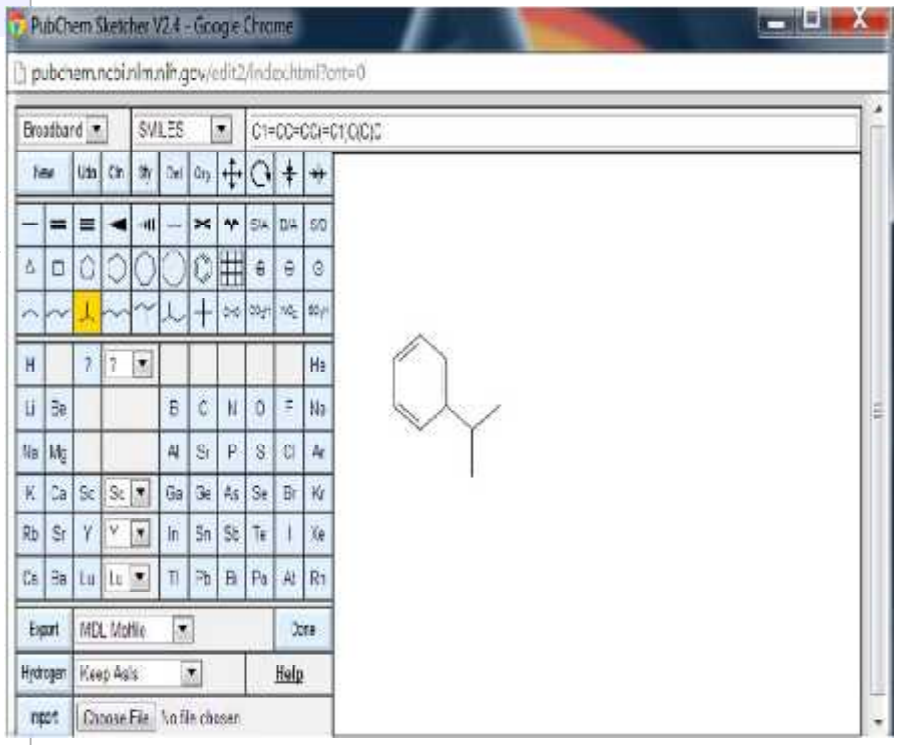
- Alphabetical, semantic, unordered
- How do we place the items for querying?
  - Search time to the same target position increases as the number of items in the panel increases [Byrne et al., CHI 99]
  - Semantic and alphabetical is faster [Halverson et al.,CHI 08]
  - Users are faster at selecting targets that are closer to the top [Cockburn et al, CHI 07]
  - Last item effect [Bailly et al, CHI 14]



Challenges

Palette Selection

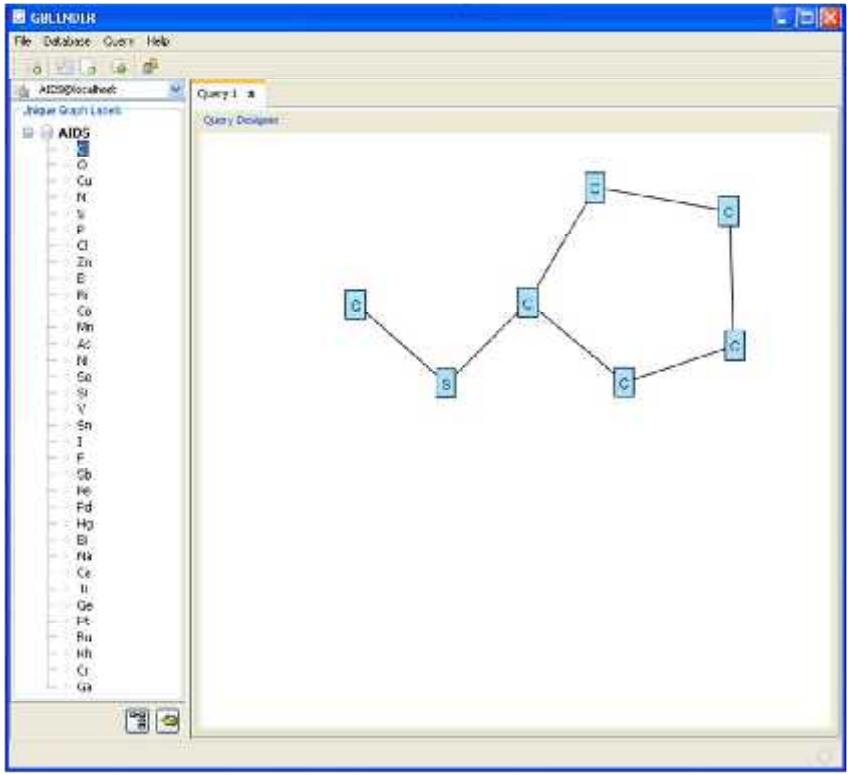
- Which subgraph patterns should be in the palette?
  - Formulate query easily and faster
  - Give shortcuts
- Issues
  - "Data-driven" selection
  - Size of the palette
  - Frequent and infrequent patterns
  - Maximally covers the DB
  - Minimal redundancy among patterns
  - Dynamically maintained



Visual Query Formulation

Characteristics

- Node/edge-at-a-time or subgraph-at-a-time approach
  - Size of the query fragment increase by 1 or k
- Query can be modified at any time
  - Size may not increase monotonically
- Different formulation sequences
- At any step, the partial query is either frequent or infrequent
  - The chance of a fragment to remain frequent diminishes as the size of the query grows

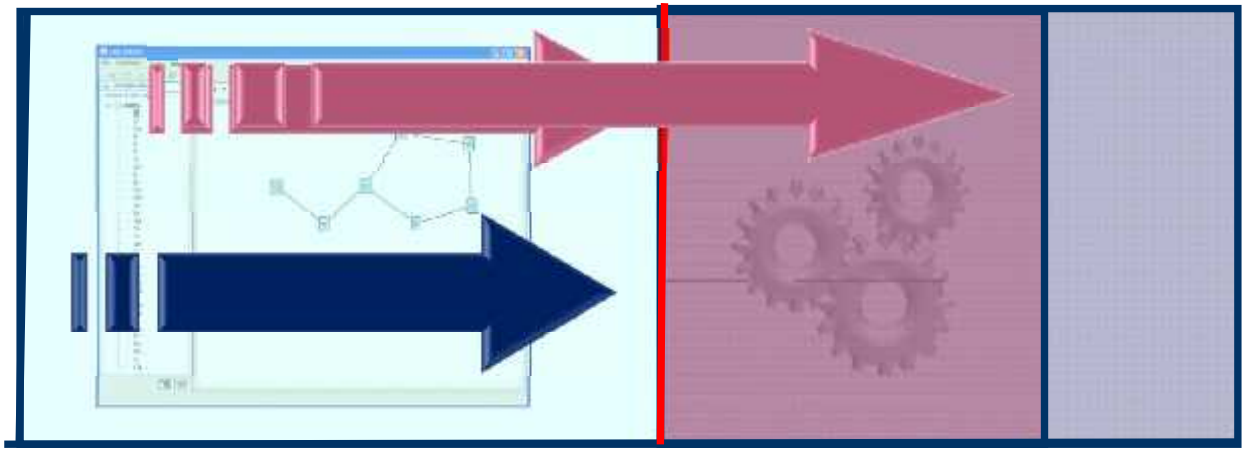


Visual Action-aware Query Processing

NANYANG TECHNOLOGICAL UNIVERSITY

A novel paradigm

- Why wait for the complete visual query to be constructed before initiating query evaluation? How can we blend these two steps?
- By initiating query processing "early", can we significantly reduce the user's waiting time?



Non-traditional Challenges

Action-aware Indexing

•Prune irrelevant results even when partial query graph is known

• Efficient traversal from g' (Step i) to g'' (Step i+1) where g'$\subset$ g'' and |g''|=|g'|+k

• Support of pruning graph-structured frequent and infrequent fragments

• Smaller-sized graphs should be efficiently indexed

• Minimize candidate verification of partial results

Intermediate results materialization

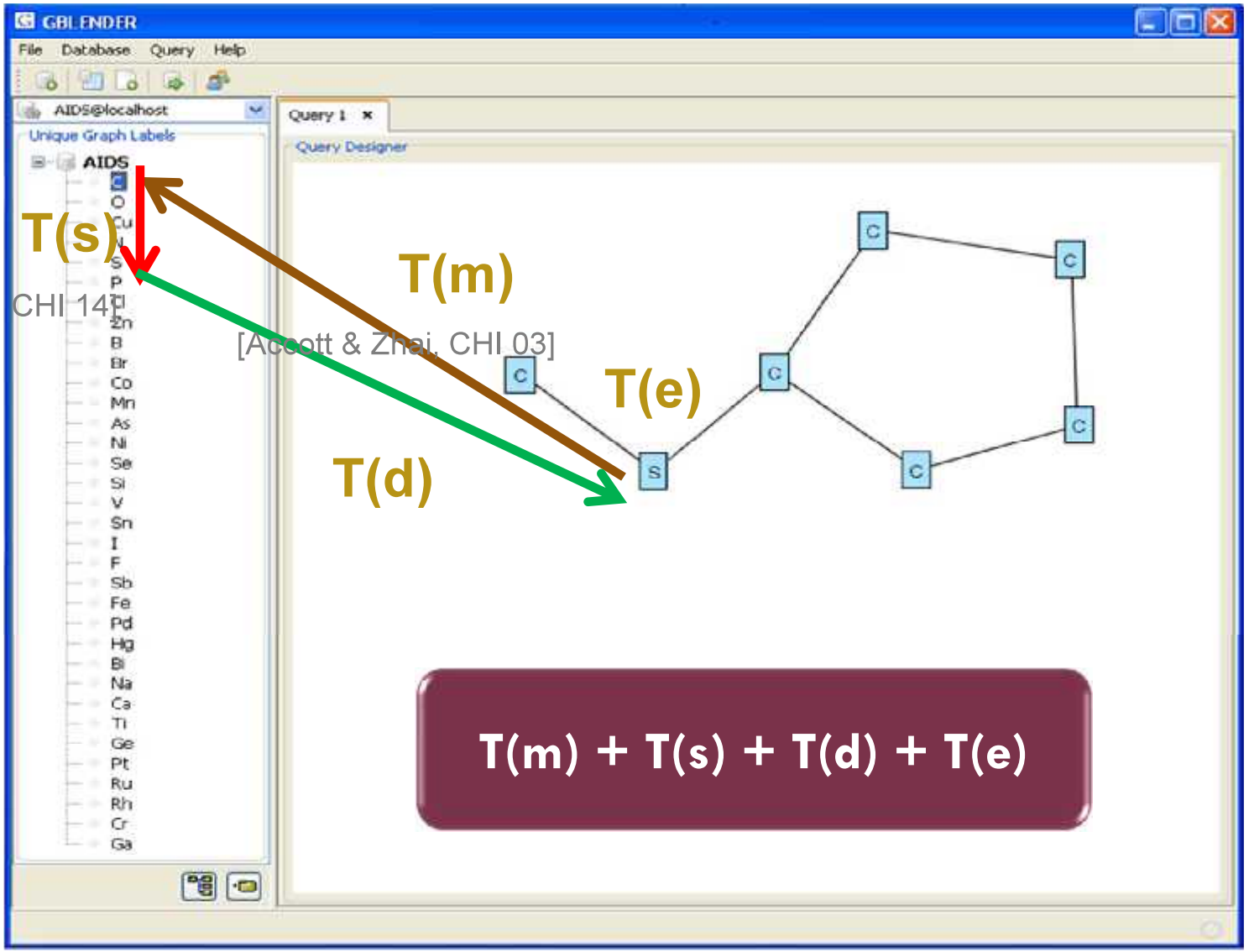
- Materialization of all partial candidate graphs matching the query fragment is necessary
- Unreasonable in traditional DB
- Challenging due to computational hardness of subgraph matching

Bounds on materialization time

NANYANG TECHNOLOGICAL UNIVERSITY

[Bailly et al, CHI 14]

[Accott & Zhai, CHI 03]



Non-traditional Challenges

Selectivity-free query processing

•Selectivity-based query processing is impractical
• Query fragments are drawn in any arbitrary sequence
• Complete query is not available during query evaluation due to the paradigm
  • "Push-down" highly selective fragment is not possible
  •Classical approach of physical query plan generation is ineffective

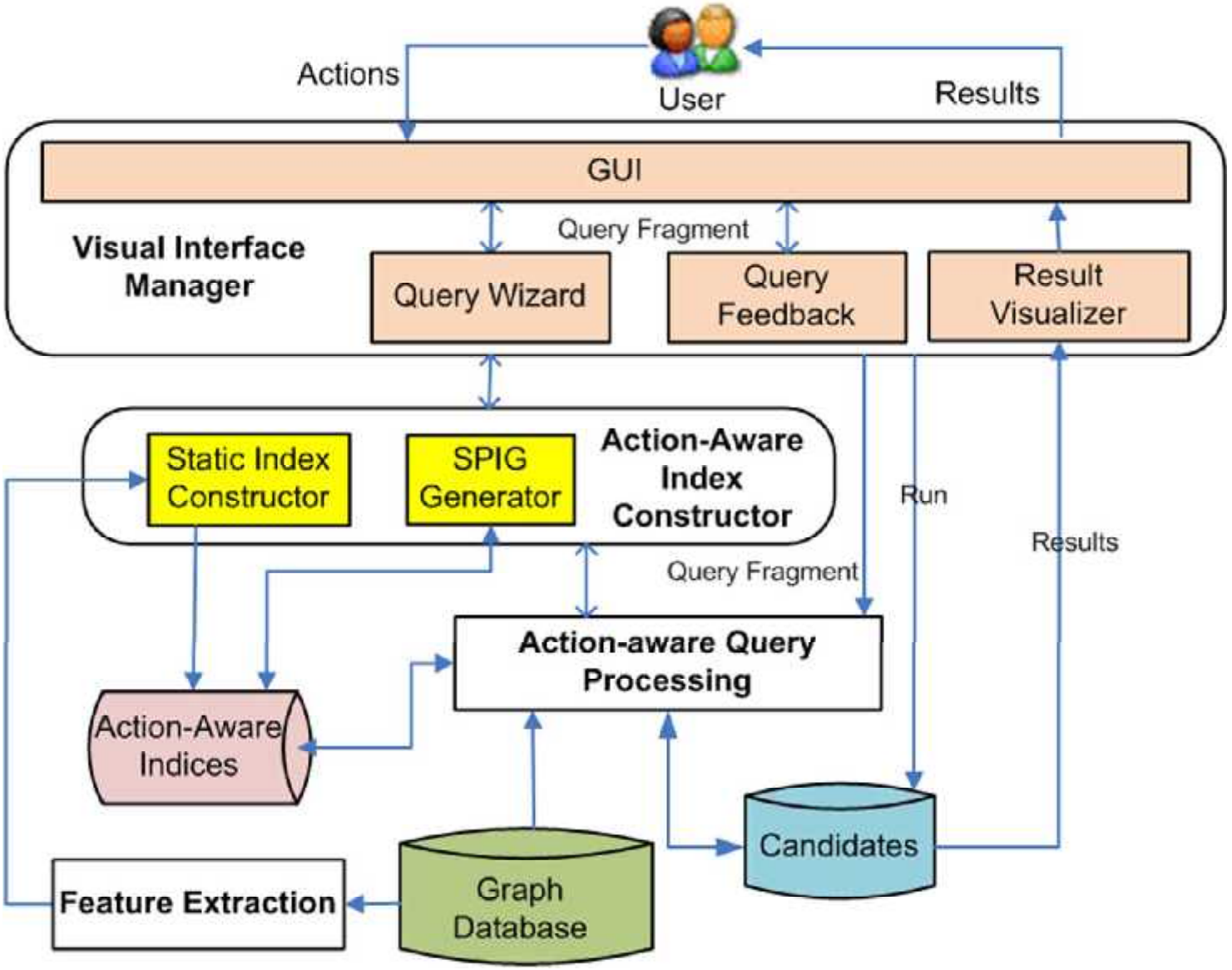
Focus on waiting time of users

- Quick formulation of queries
- System response time (SRT) matters more than backend processing cost!
- SRT should be robust to different query formulation sequences.

"Computing time (power) is getting cheaper but users' time isn't.."

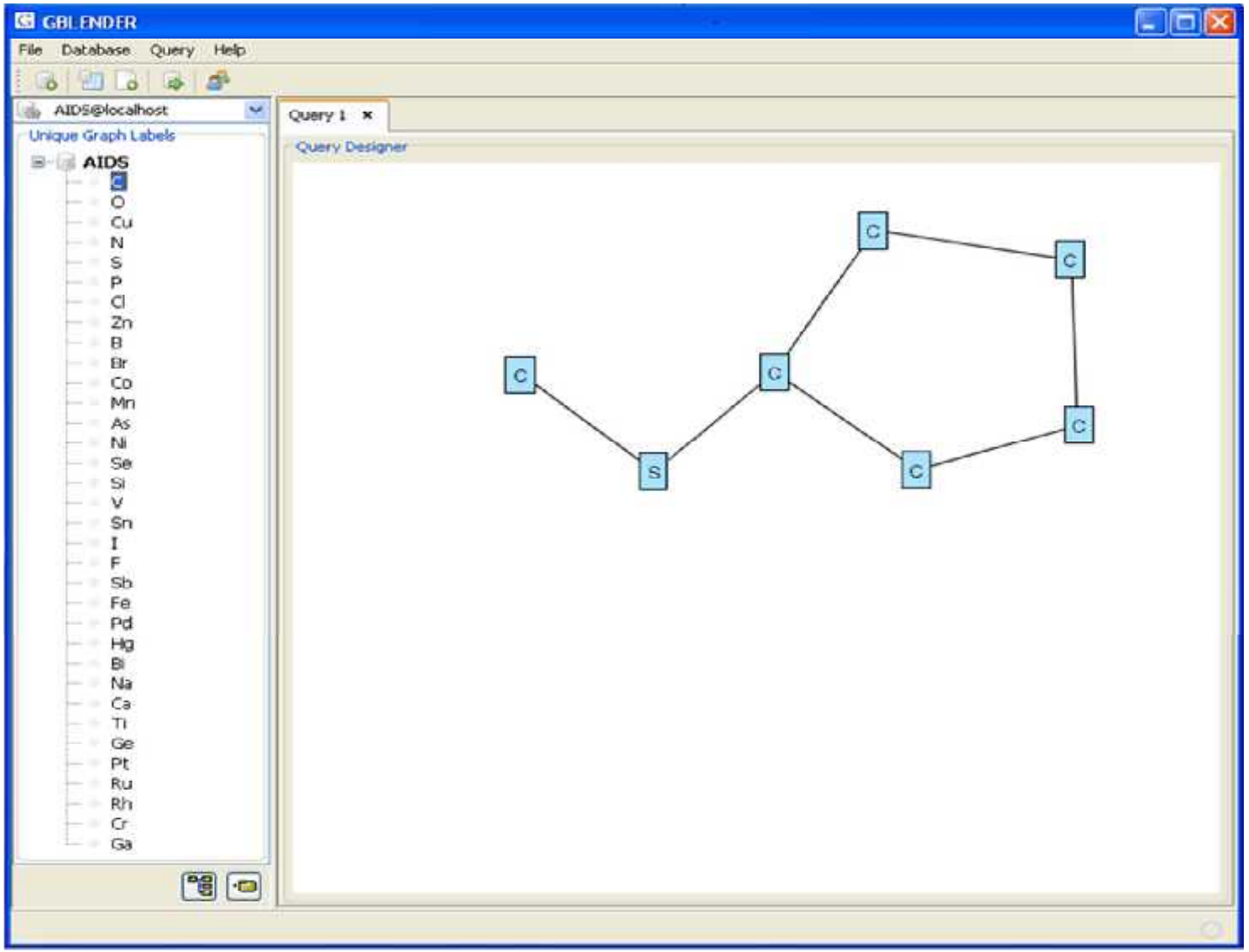
VOGUE [CIDR 13; SIGMOD 10, 11; ICDE 12]

NANYANG TECHNOLOGICAL UNIVERSITY



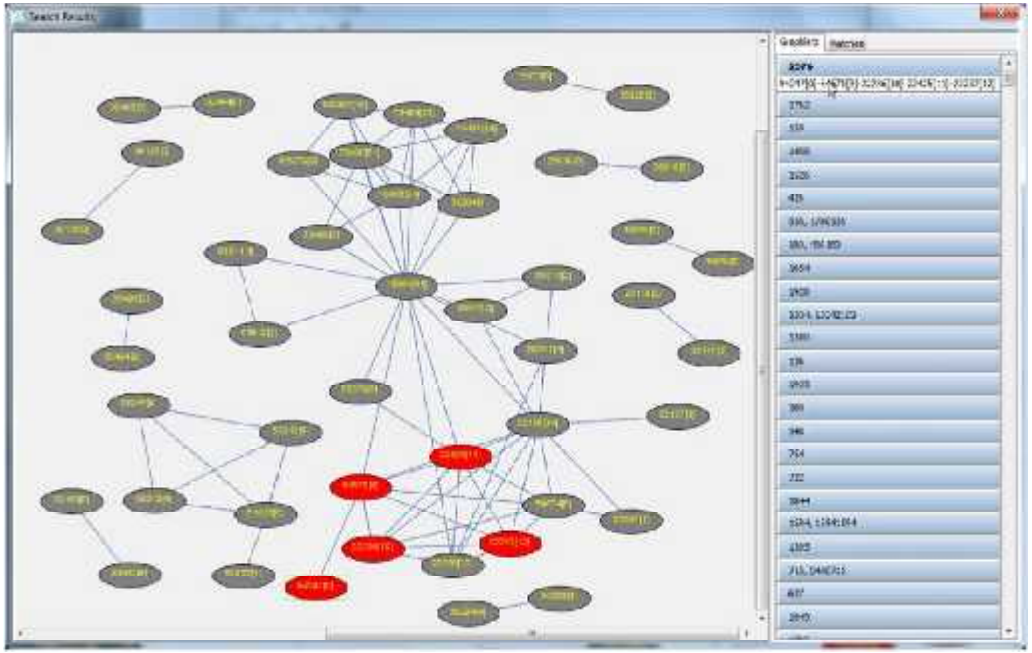
A Video of Visual Action-aware Query Processing

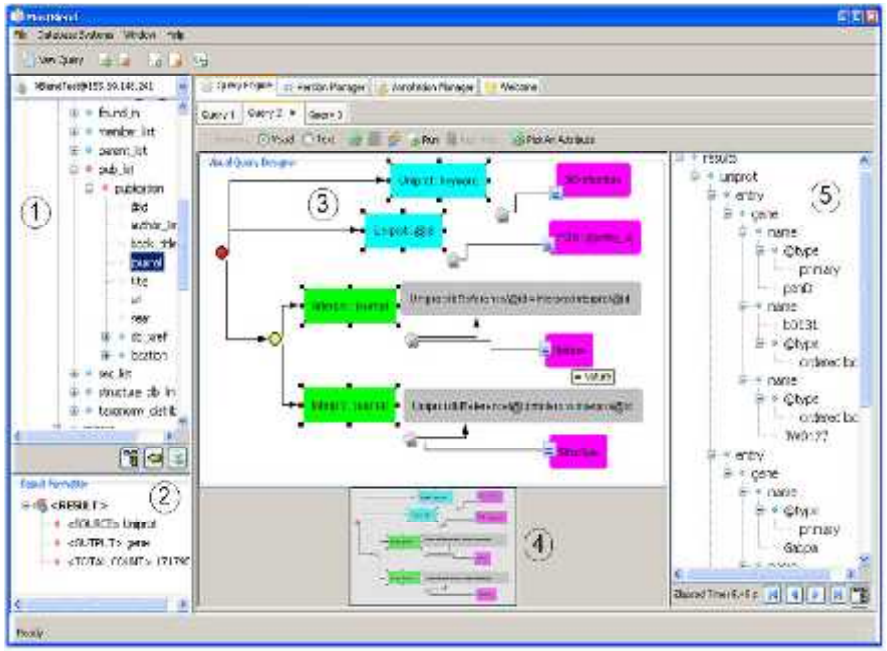
NANYANG TECHNOLOGICAL UNIVERSITY



Related Efforts

- ☑ QUBLE [SIGMOD 13, VLDBJ 14]
  - For large networks
- ☑ MUSTBLEND [DASFAA 13, ICDE 09, ICDE 06]
  - For XML data





Data-driven Visual Interface Management Revisited

Goal 2

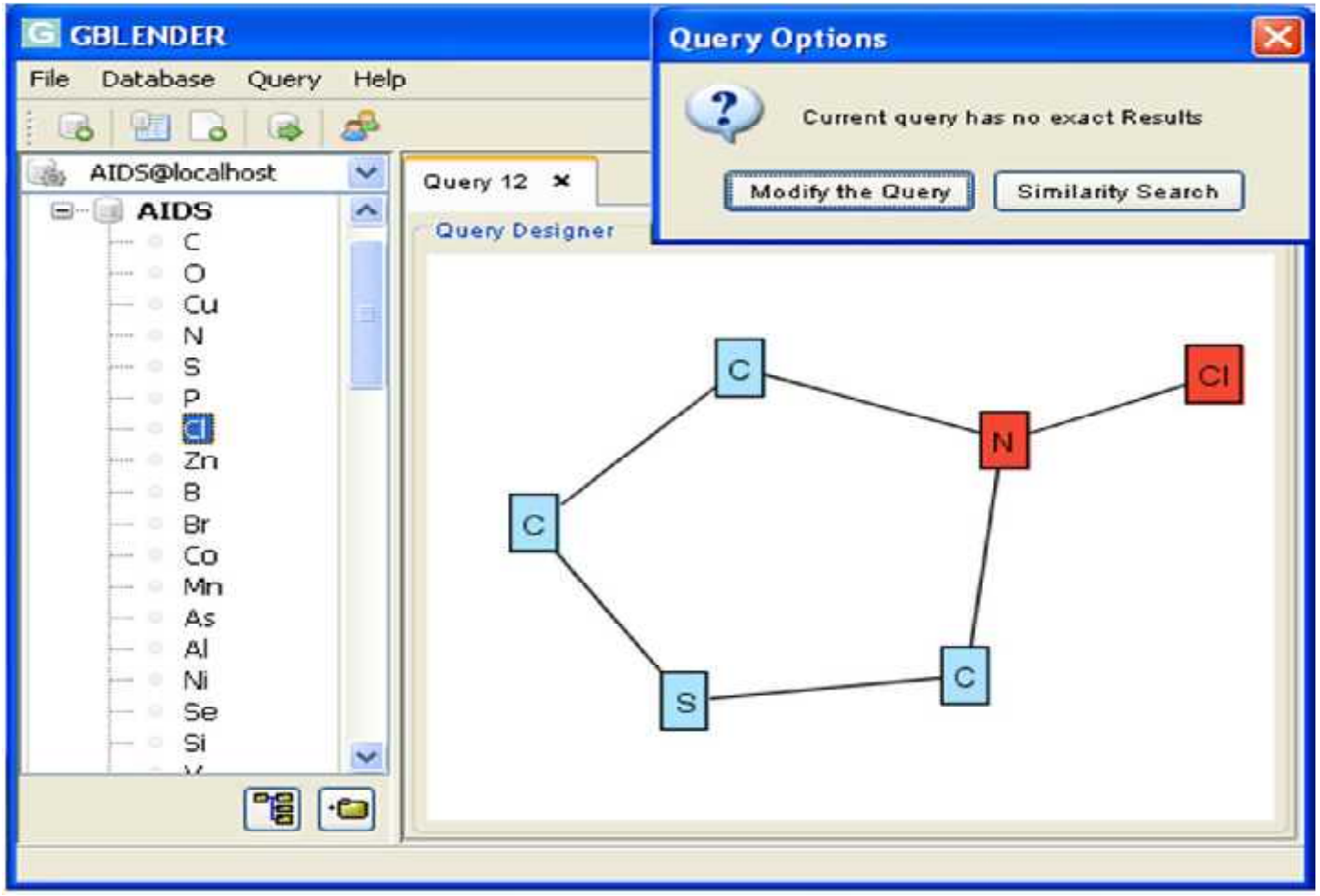
- Side-effects of intermediate results materialization
- Data-driven feedbacks and notifications

What kind of feedback?

- Aiding query construction by making appropriate suggestions
  - Given a query fragment, which top-k patterns are most likely to be added to the query?
- Empty results detection

Empty Results Detection

NANYANG TECHNOLOGICAL UNIVERSITY



Data-driven Visual Interface Management Revisited

Goal 3

- Interruption-sensitive notifications

Lessons from HCI & Cognitive Psychology

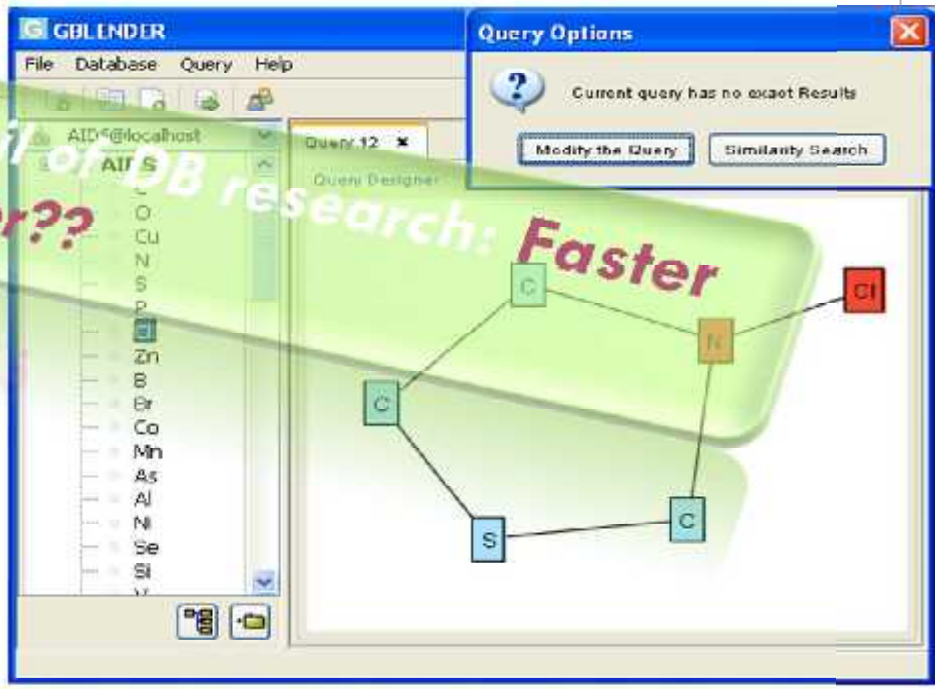
- Interrupting users engaged in tasks by delivering notifications inopportunely can negatively impact task completion time, lead to more errors, and increase user frustration [Bailey et al., JCSB 2006].
- Breakpoint-based notifications

Notification Management

Notification in the data management world

- Overly-aggressive!
- Immediate notification regardless of the cognitive impact on the user

Questioning the holy grail of DB research: Faster evaluation is always better??

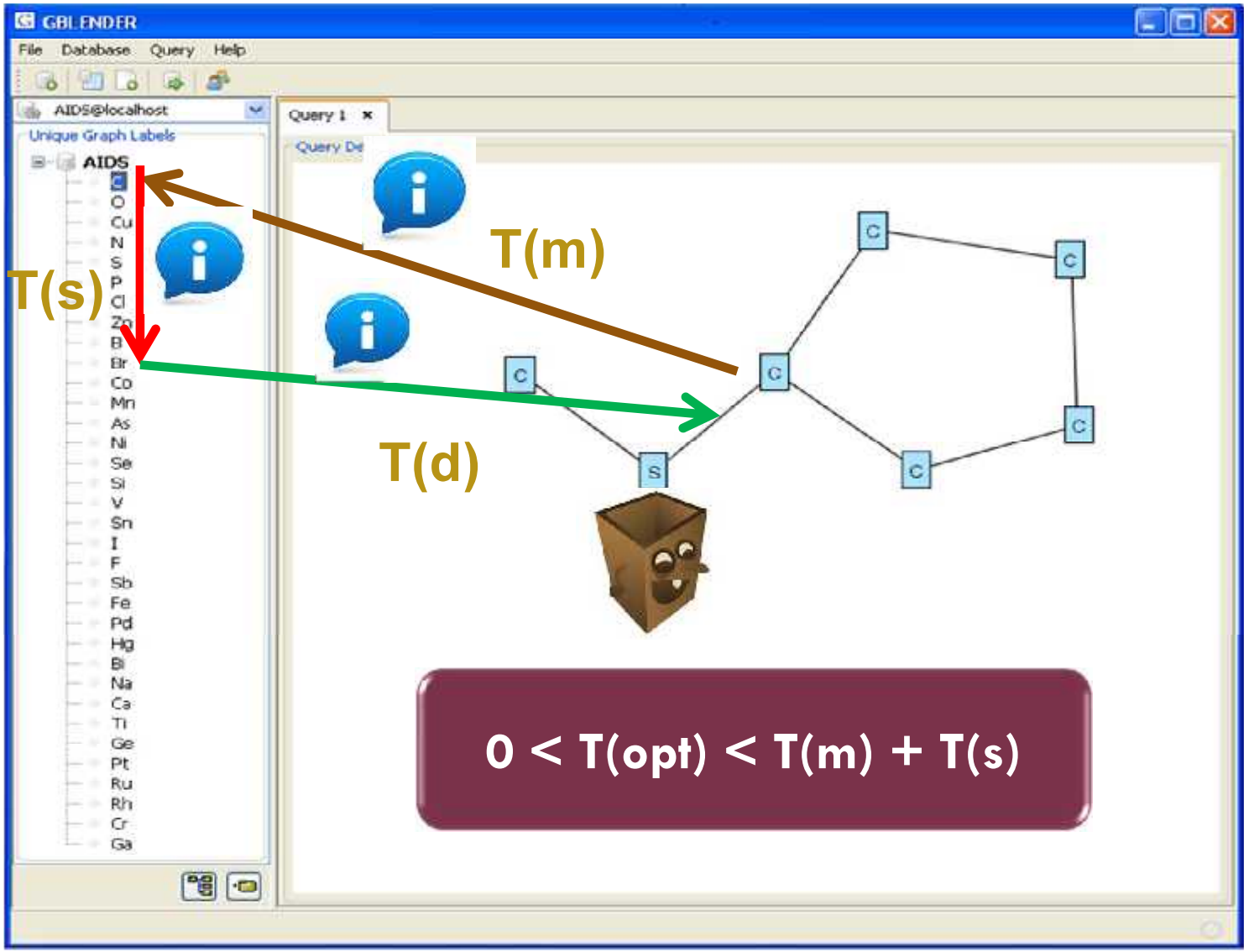


Solution

- Seamlessly integrate defer-to-breakpoint strategy [Iqbal et al., CHI 2007, 2008] with query formulation
- Defer notification to optimal breakpoints

Bounds on feedback generation time

NANYANG TECHNOLOGICAL UNIVERSITY



Visual Action-aware Query Performance Simulator

NANYANG TECHNOLOGICAL UNIVERSITY

Large-scale performance study

- Traditional approach
  - Randomly extract subgraphs of different size and execute them
- Doesn't work in this paradigm!

Why?

- Queries need to be visually constructed by users
- GUI latency is critical for performance study

Challenge

- Users are expensive!
- How do we simulate visual query formulation?

NANYANG TECHNOLOGICAL UNIVERSITY

Next..

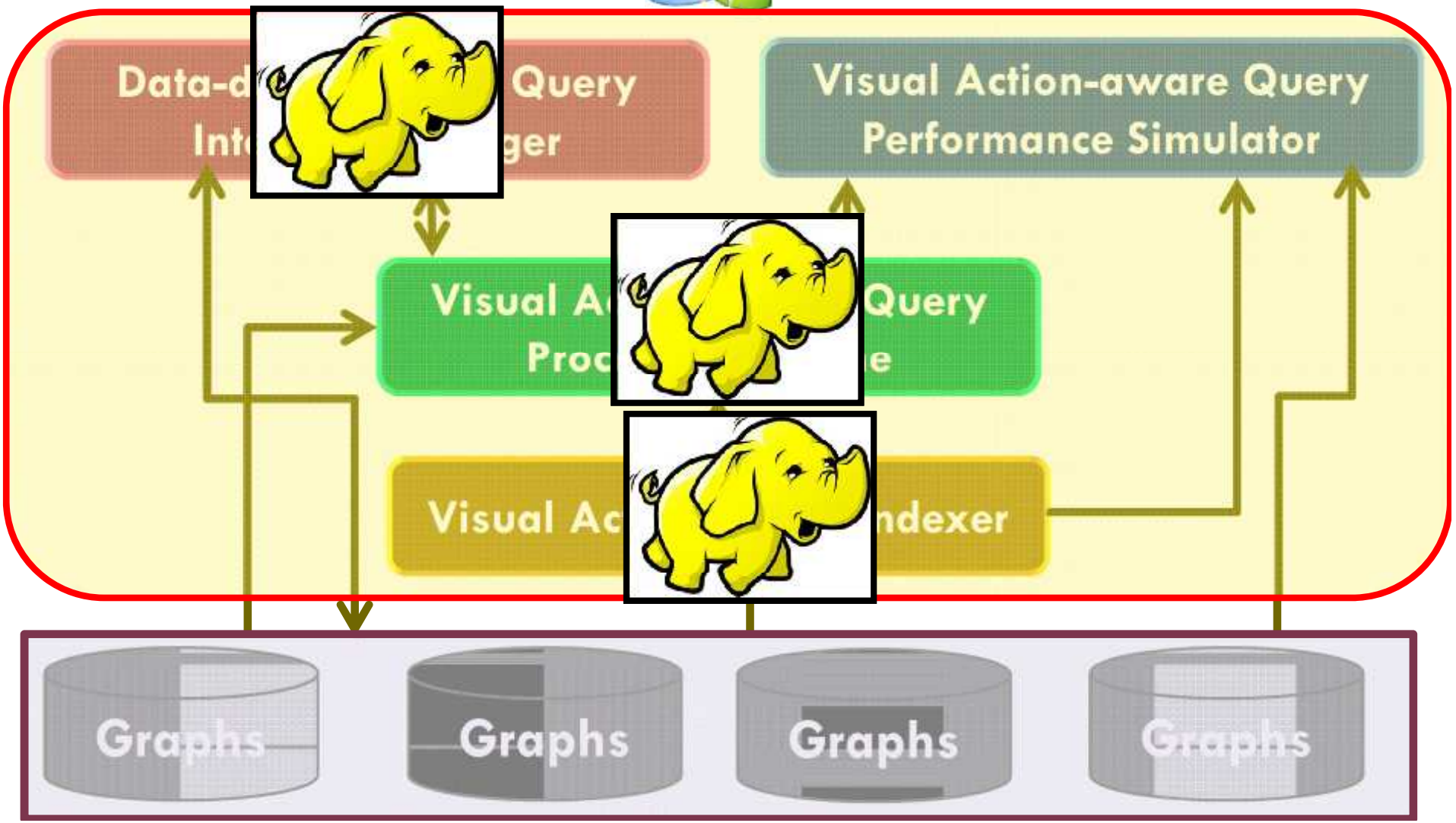
Blending more Complex Queries

NANYANG TECHNOLOGICAL UNIVERSITY

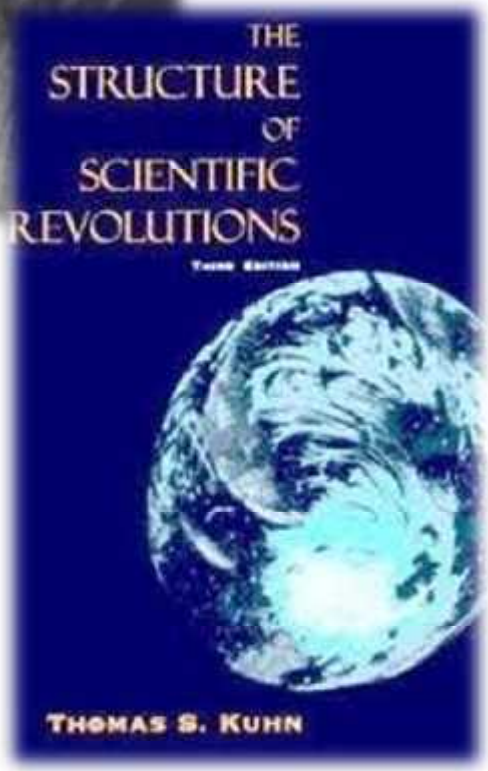
- Homeomorphic queries
- Distance-constraint queries
- Multi-attribute queries
- Graph simulation

HCI-aware Data Management in the Big Data Era

NANYANG TECHNOLOGICAL UNIVERSITY



Kuhn's Paradigm



1) Establish a paradigm

2) Mature science

3) Anomalies appear

Published in 1962

Kuhnian Example

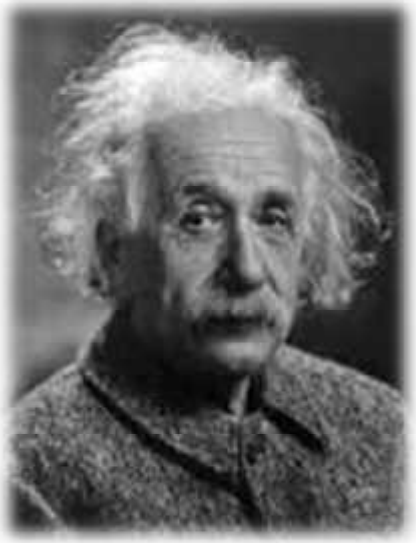
1700: Classical (Newtonian) mechanics

1800

Experiments with light show anomalous behavior

1900

1905: Relativity



Kuhnian Paradigm for Data Management

NANYANG TECHNOLOGICAL UNIVERSITY

1970s: Query Formulation → Query Processing

2000s: Visual query formulation ↔ query processing

- Blends data management and HCI
- Three key components
  - Data-driven visual interface management
  - Visual action-aware query processing
  - Visual action-aware performance simulation
- Can be extended to wide variety of data types

Broad goal

- **Pervasive desire to stimulate a cultural shift in our thinking by bringing together HCI and data management to "work" together**

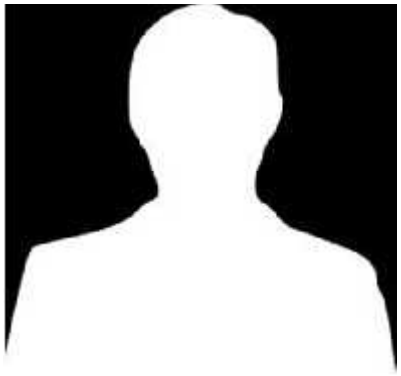
Acknowledgements

NANYANG TECHNOLOGICAL UNIVERSITY

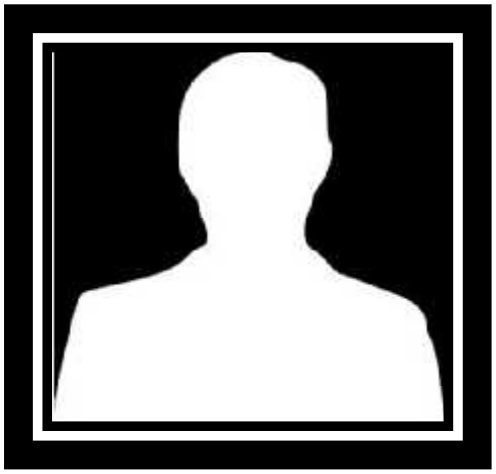
Byron Choi
Hong Kong Baptist Univ

Changjiu Jin

Shuigeng Zhou
Fudan Univ, China



Ba Quan Truong



Ho Hoang Hung


thank you
tenki
ขอขอบคุณคุณ
спасибо
הדות
dhanyavadagalu
дзякуй
hvala
dziękuje
bedankt
xièxie
gracias
mèsi
rahmet
danke
mochchakkeram
trugarez
dank je
arigatô
kop khun krap
ありがとう
gràcie
děkuji
ačiū
teşekkür ederim
obrigada
sukriya
obrigado
dakujem
sagolun
taiku
terima kasih
謝謝
merci
grazie
asante
mauruuru
go raibh maith agat
najs tuke
nanni
rahmat