INTERACTIVE EXPLORATION OF EVENT SEQUENCES IN TEMPORAL CATEGORICAL DATA

KRIST WONGSUPHASAWAT

DEPARTMENT OF COMPUTER SCIENCE &
HUMAN-COMPUTER INTERACTION LAB
UNIVERSITY OF MARYLAND

Dissertation Proposal
April 30, 2010

INTERACTIVE EXPLORATION OF EVENT SEQUENCES IN **TEMPORAL CATEGORICAL DATA**

KRIST WONGSUPHASAWAT

DEPARTMENT OF COMPUTER SCIENCE &
HUMAN-COMPUTER INTERACTION LAB
UNIVERSITY OF MARYLAND

Dissertation Proposal
April 30, 2010

TEMPORAL CATEGORICAL DATA

- A type of time series

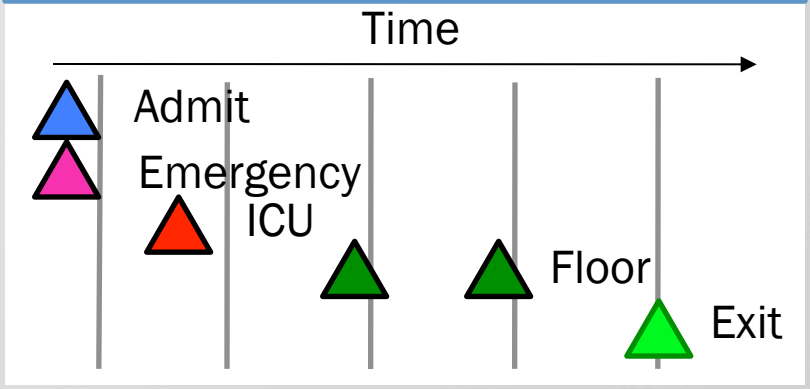
Numerical

Stock: Microsoft	
04/26/2010 10:00	31.03
04/26/2010 10:15	31.01
04/26/2010 10:30	31.02
04/26/2010 10:45	31.08
04/26/2010 11:00	31.16
04/26/2010 11:15	31.15

Categorical

Category Event

Patient ID: 45851737	
12/02/2008 14:26	Admit
12/02/2008 14:26	Emergency
12/02/2008 22:44	ICU
12/05/2008 05:07	Floor
12/08/2008 10:02	Floor
12/14/2008 06:19	Exit



TEMPORAL CATEGORICAL DATA

- Electronic Health Records: symptoms, treatment, lab test
- Traffic incident logs: arrival/departure time of each unit
- Student records: course, paper, proposal, defense, etc.
- Usability study logs
- Etc.

WORKING WITH PHYSICIANS AT

WASHINGTON HOSPITAL CENTER

INTERACTIVE EXPLORATION OF **EVENT SEQUENCES** IN TEMPORAL CATEGORICAL DATA

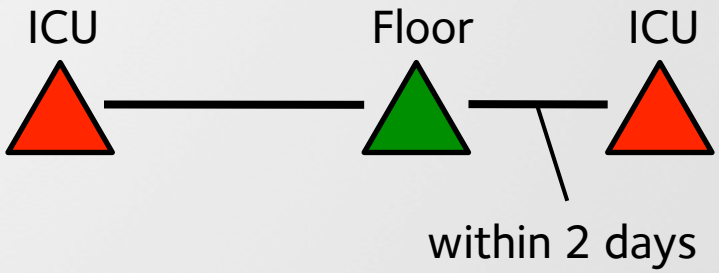
KRIST WONGSUPHASAWAT

DEPARTMENT OF COMPUTER SCIENCE &
HUMAN-COMPUTER INTERACTION LAB
UNIVERSITY OF MARYLAND

Dissertation Proposal
April 30, 2010

EVENT SEQUENCES

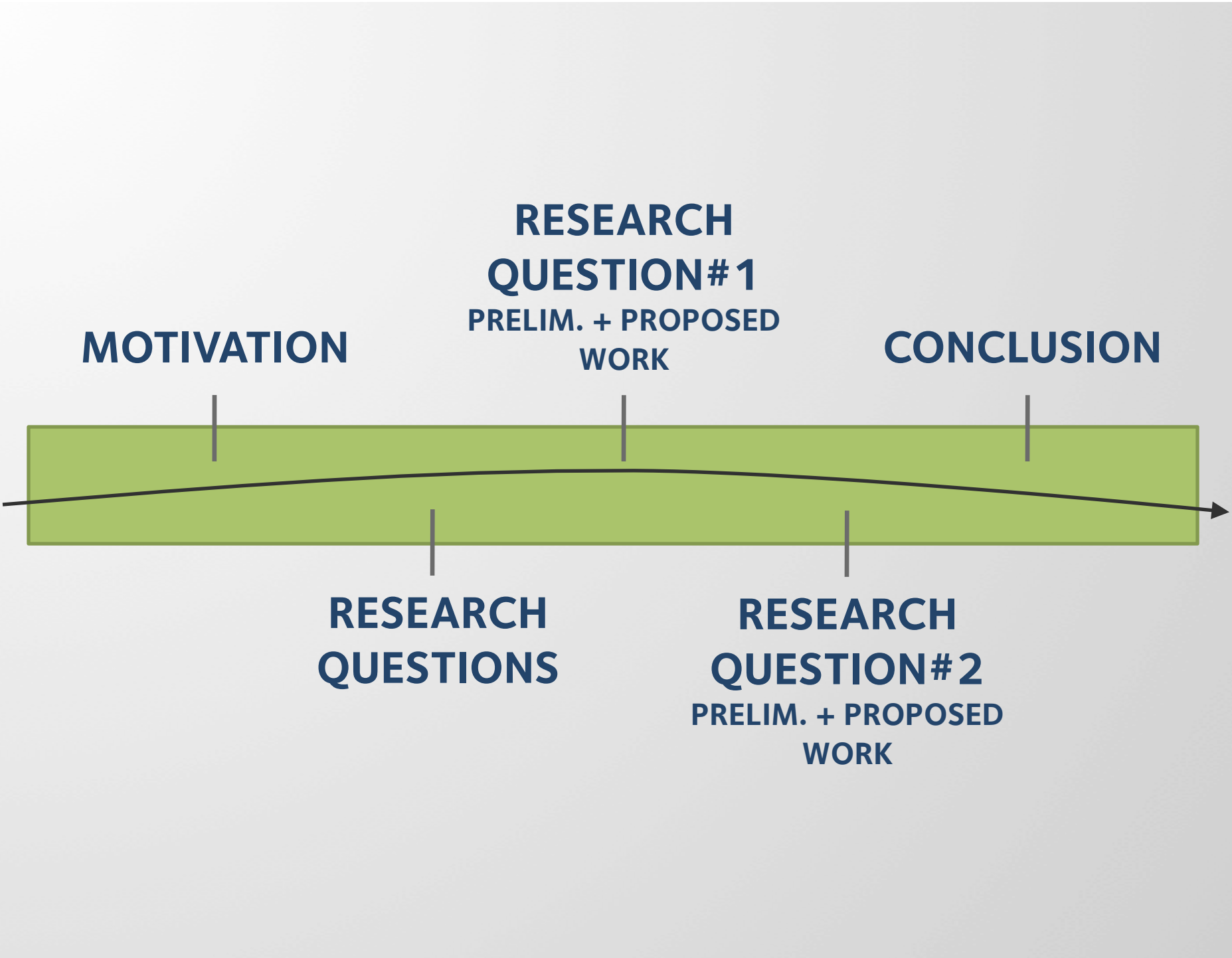
- Examples: “Bounce backs”



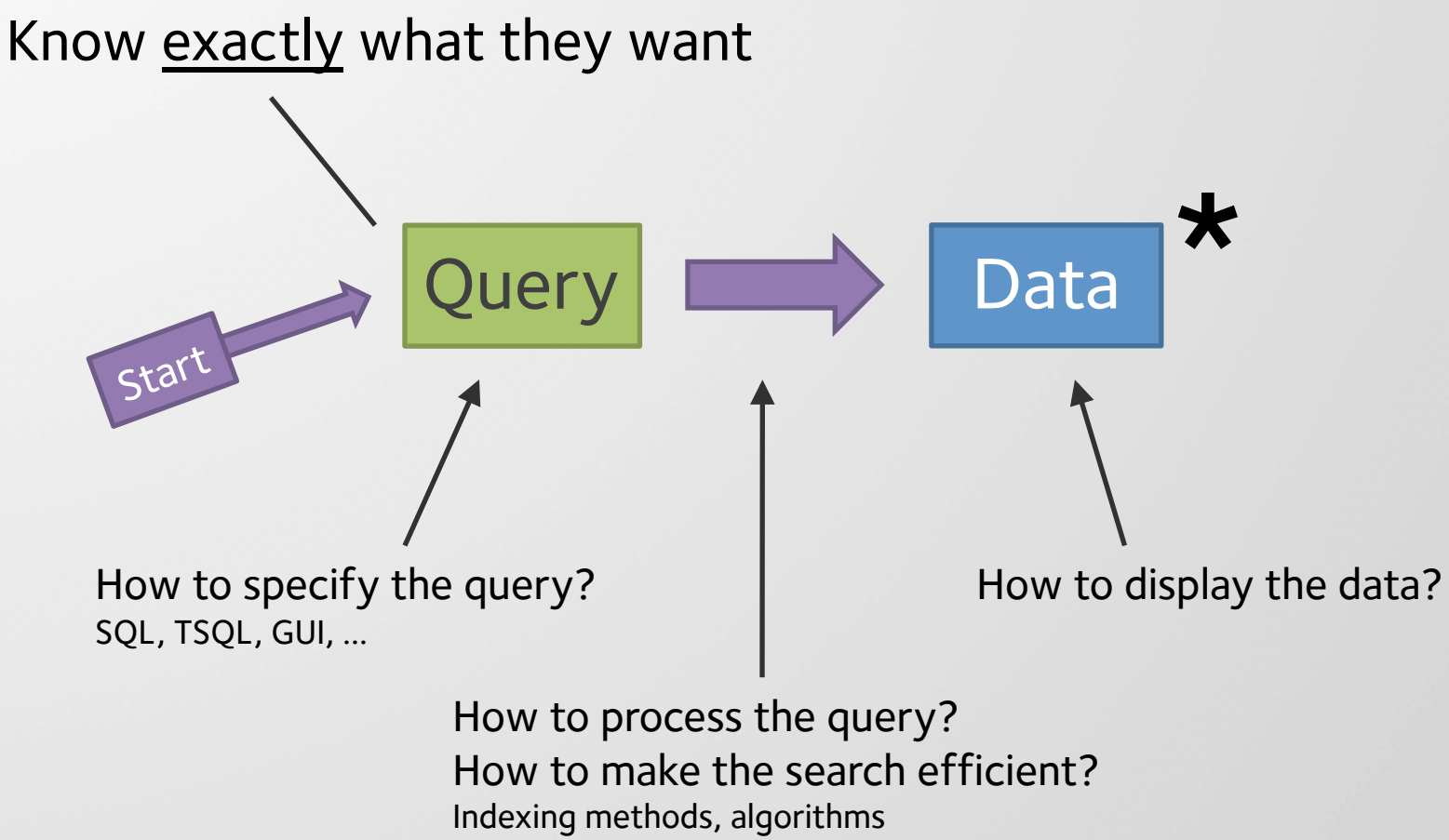
RESEARCH STATEMENT

"MY RESEARCH AIMS TO DESIGN
EFFECTIVE VISUALIZATION
AND INTERACTION TECHNIQUES
TO SUPPORT USERS IN EXPLORING
EVENT SEQUENCES IN TEMPORAL CATEGORICAL DATA."

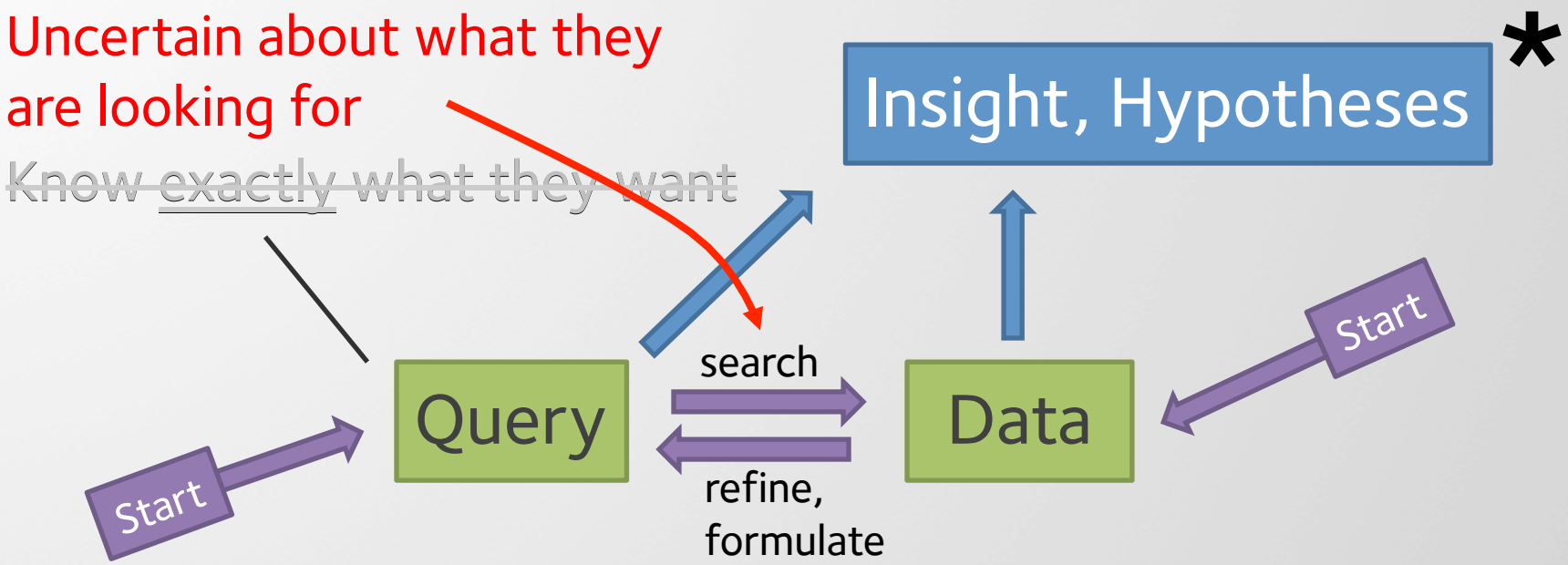
- A FLEXIBLE TEMPORAL SEARCH APPROACH
- A NOVEL VISUALIZATION THAT PROVIDES AN OVERVIEW OF MULTIPLE RECORDS


RESEARCH QUESTION#1
PRELIM. + PROPOSED WORK
MOTIVATION
CONCLUSION
RESEARCH QUESTIONS
RESEARCH QUESTION#2
PRELIM. + PROPOSED WORK

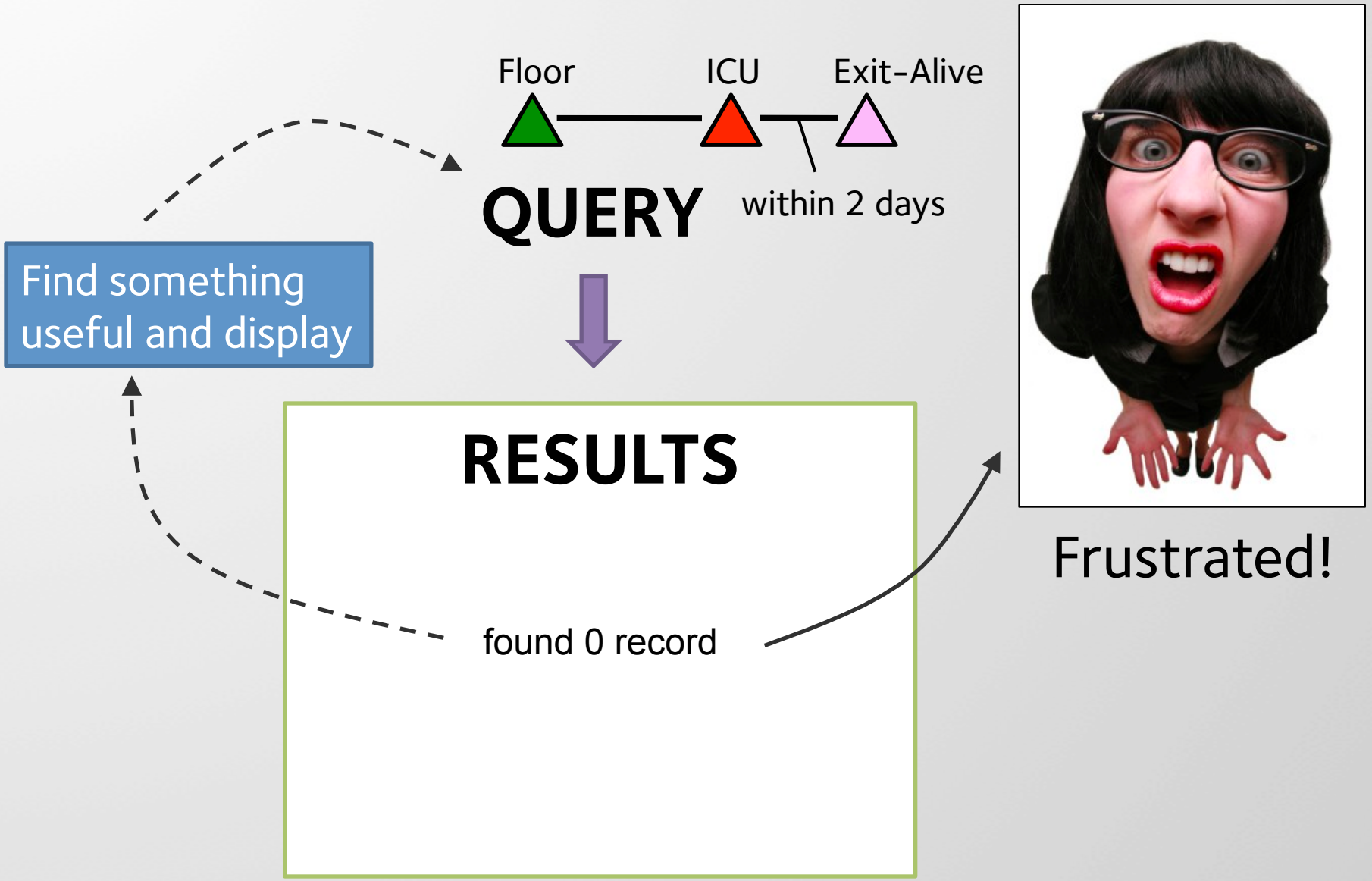
TRADITIONAL SEARCH MODEL



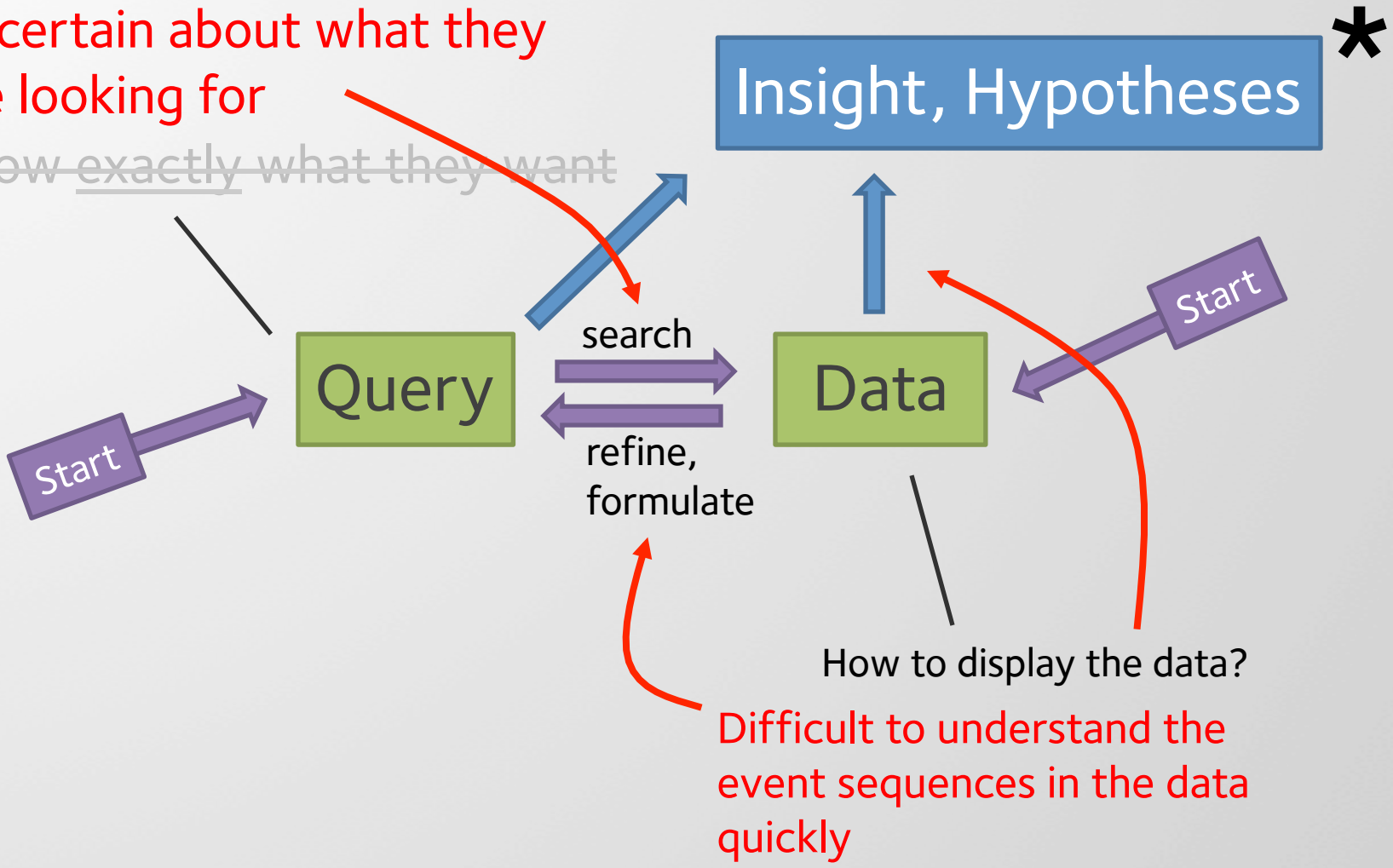
EXPLORATORY SEARCH MODEL



UNCERTAIN


Floor
ICU
Exit-Alive
QUERY
within 2 days
Find something useful and display
RESULTS
found 0 record
Frustrated!

EXPLORATORY SEARCH MODEL



NEEDS A STARTING POINT

Show overview or summary

Where should I start?

RESEARCH QUESTIONS

1. How to support the users when they are uncertain about what they are looking for?
2. How to provide an overview of event sequences for temporal categorical data?

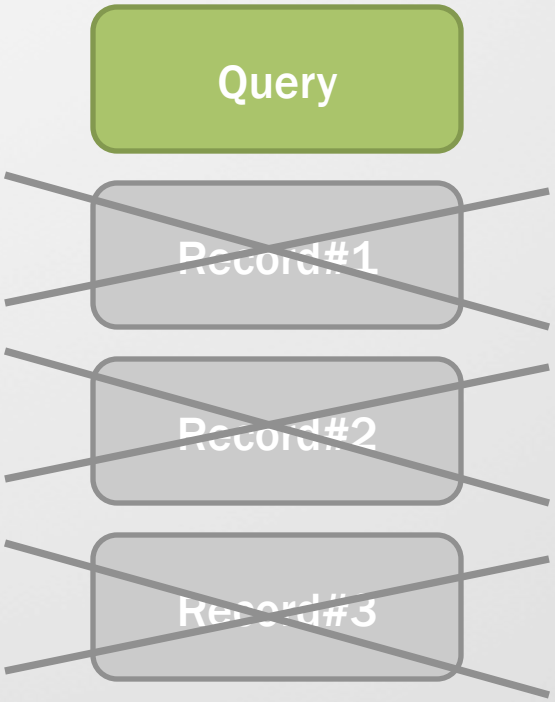
RESEARCH QUESTIONS

1. How to support the users when they are uncertain about what they are looking for?
2. How to provide an overview of event sequences for temporal categorical data?

APPROACHES

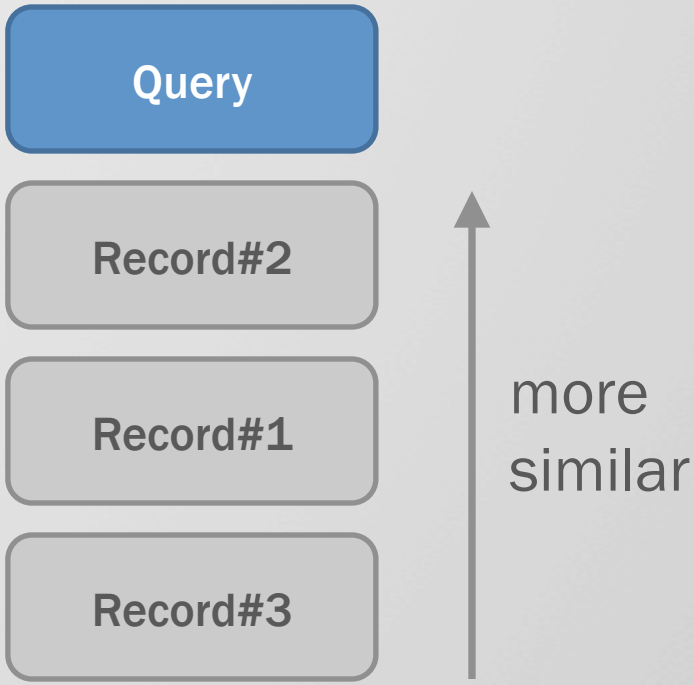
Exact Search

MUST have A, B, C

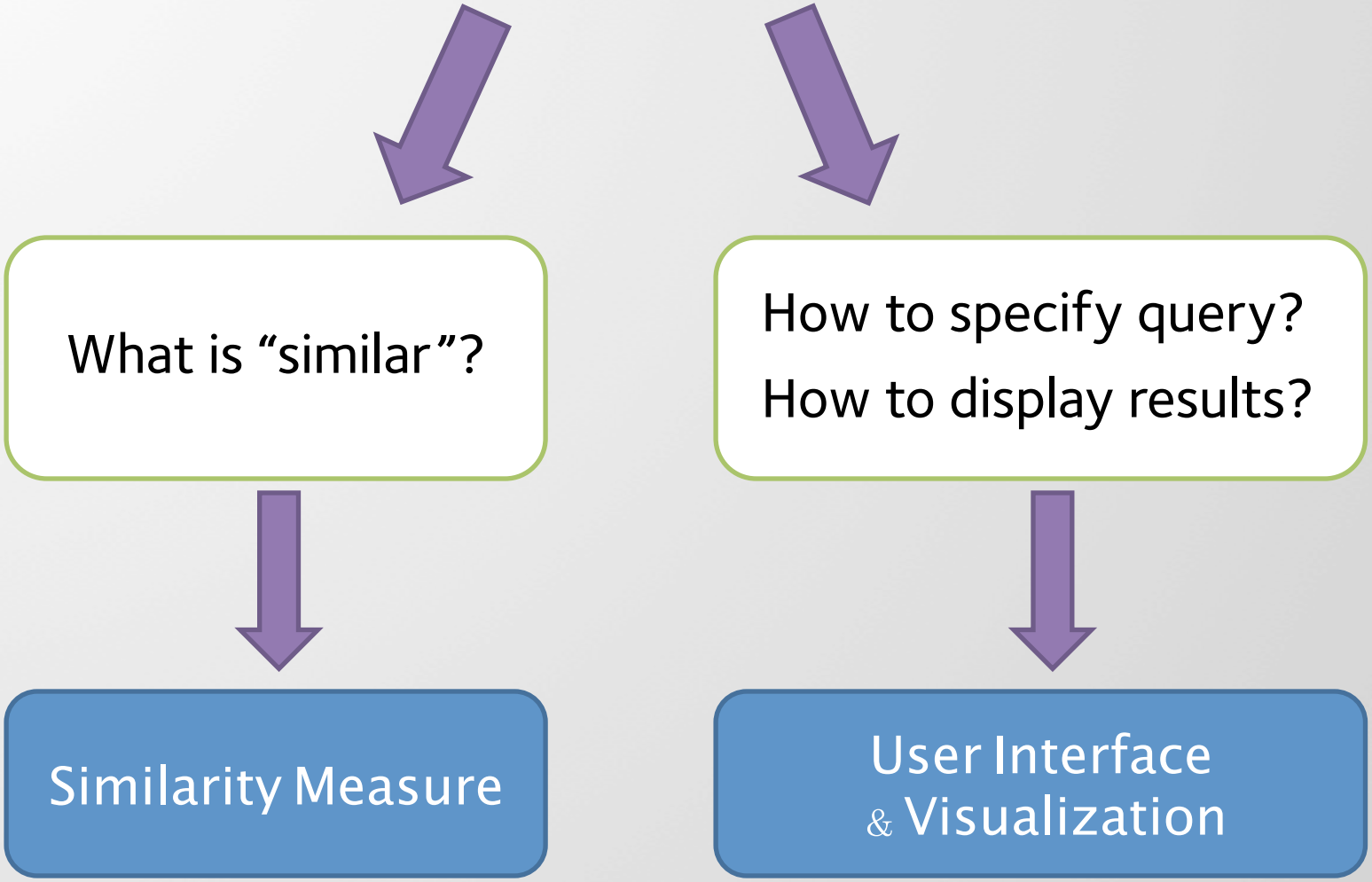


Similarity-based Search

SHOULD have A, B, C

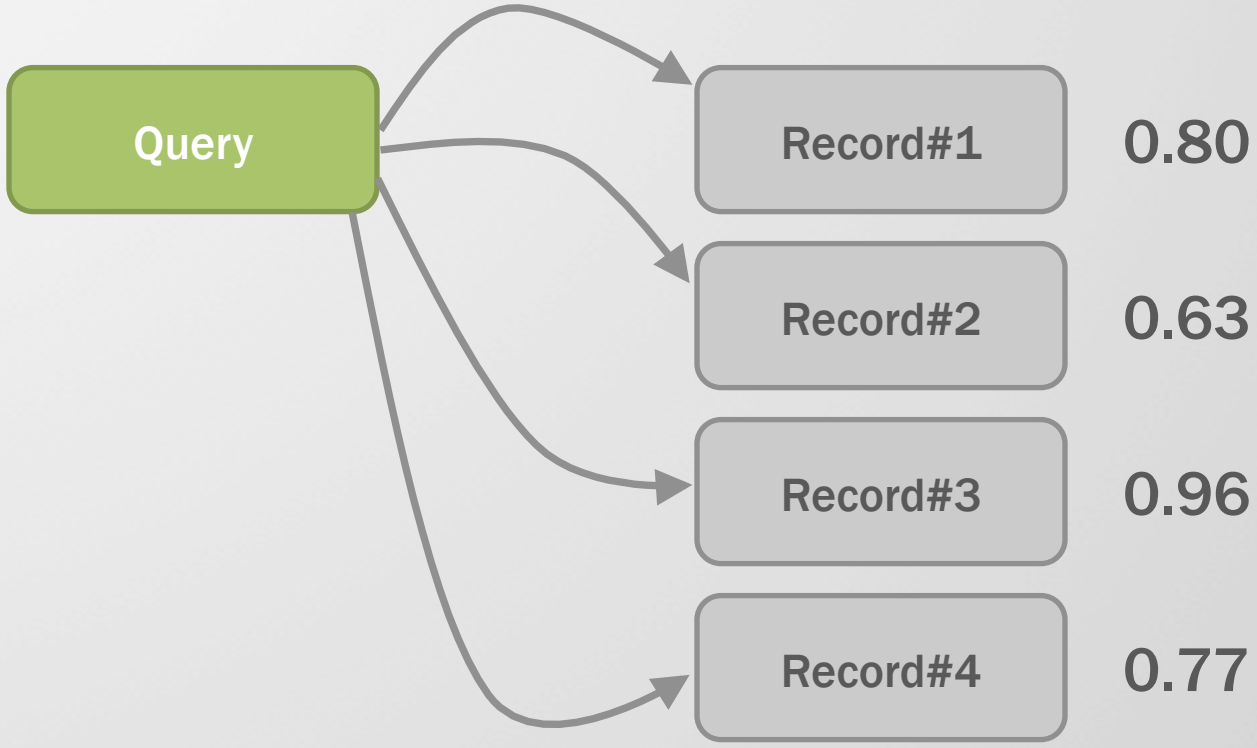


SIMILARITY-BASED SEARCH



SIMILARITY MEASURE

Gives a **score** that shows how much a record is similar to the query.
Min = 0 / Max = 1

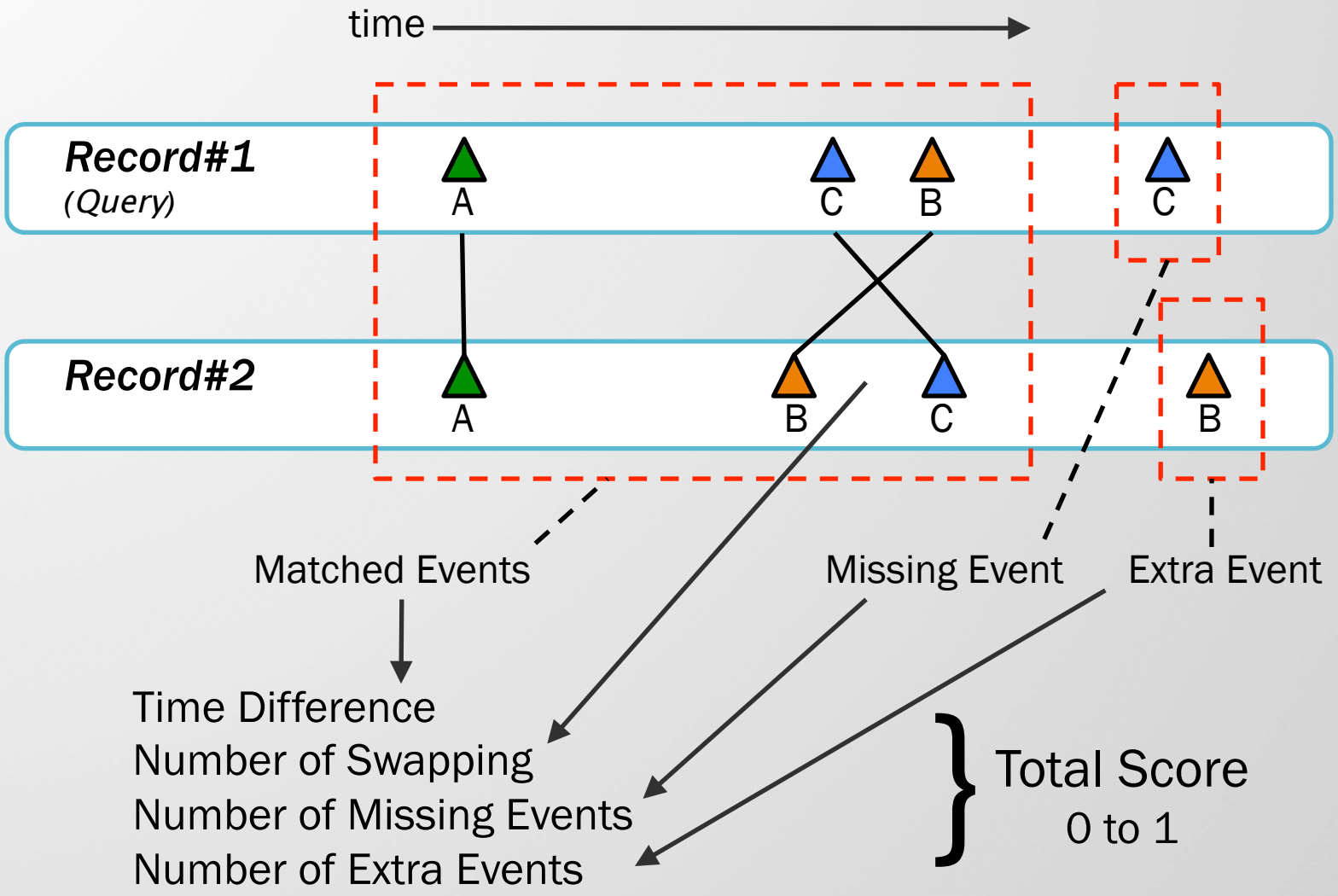


SIMILARITY MEASURE (2)

CHALLENGE

- What is “similar”? depends on users/tasks

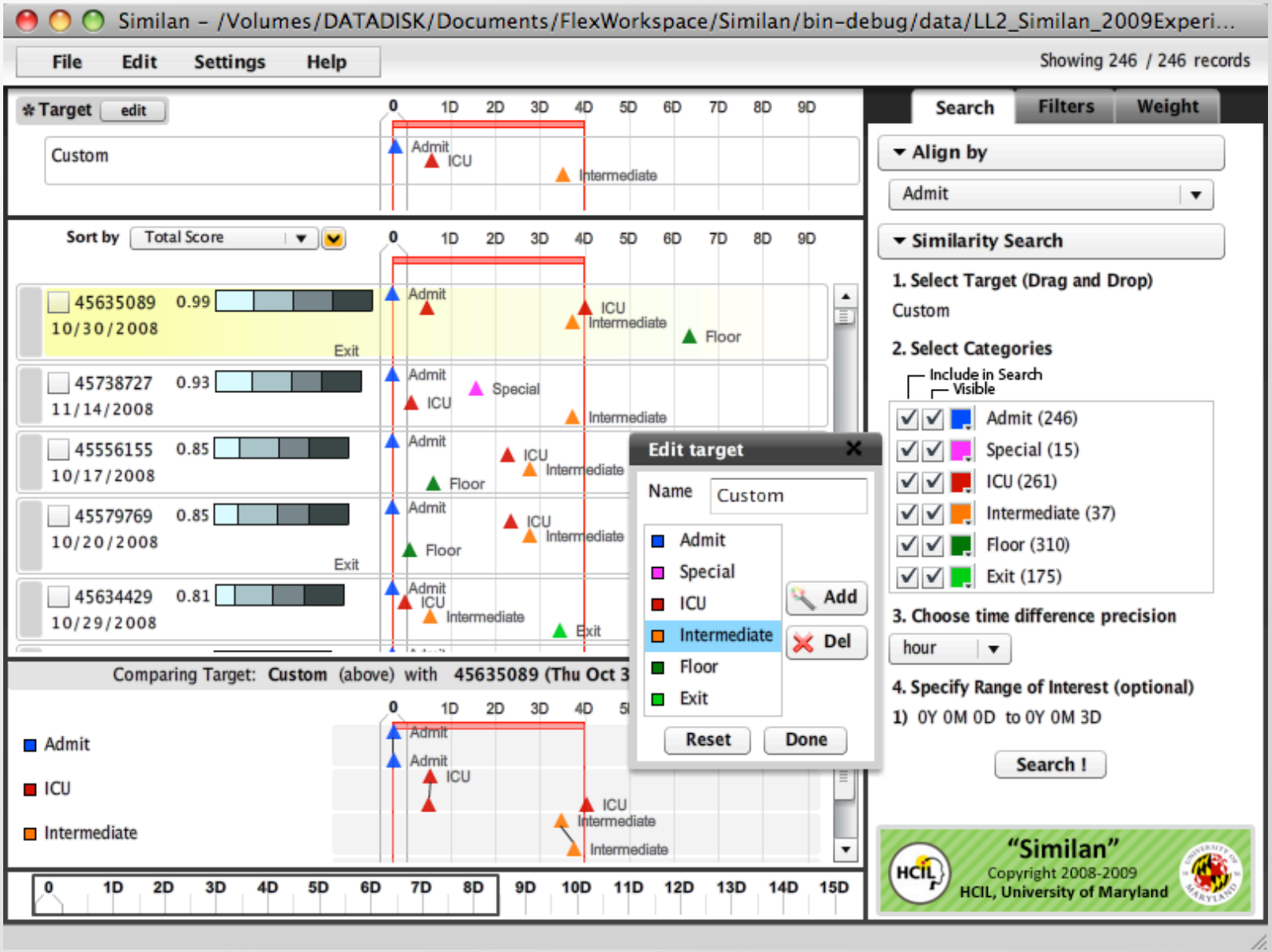
MATCH & MISMATCH MEASURE



RELATED WORK

- Similarity measure
  - For numerical time series
    - [Ding et al. 2008], [Liu et al. 2006], [Berndt and Clifford 1994], …
  - Edit distance
    - [Levenshtein 1966], [Winkler 1999], [Chen 2004], …
  - Biological sequence searching
    - [Pearson and Lipman 1988], [Altschul et al. 1990], …
  - Temporal Categorical Data
    - [Sherkat 2006], …

PROTOTYPE: SIMILAN



DEMO – DATA

- Patient transfers in the emergency department
- Real data from hospital (Jan – Mar 2010)

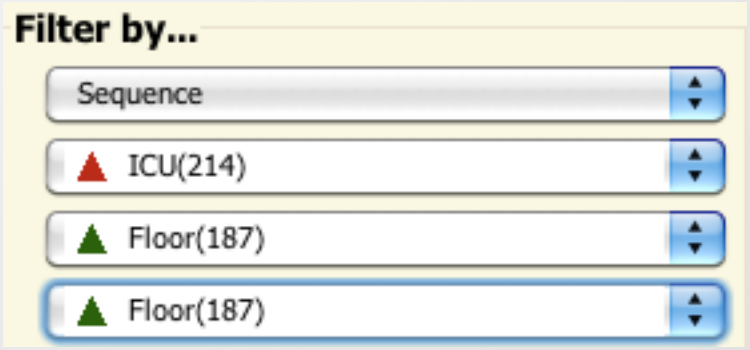
ADMIT	Admission time
EMERGENCY	Emergency room
ICU	Intensive Care Unit
INTERMEDIATE	Intermediate Medical Care
FLOOR	Normal room
EXIT-ALIVE	Leave the hospital alive
EXIT-DEAD	Leave the hospital dead

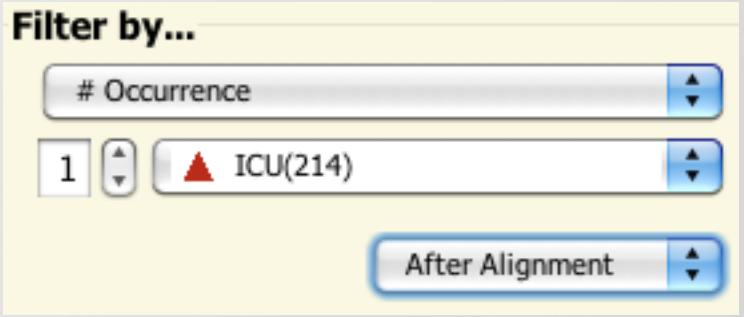
DEMO – SIMILAN

This is when bugs are more likely to occur.

CONTROLLED EXPERIMENT

- 18 participants
- 2 interfaces by 5 tasks
  - Similarity-based Search (Similan)
  - Exact Search (LifeLines2)





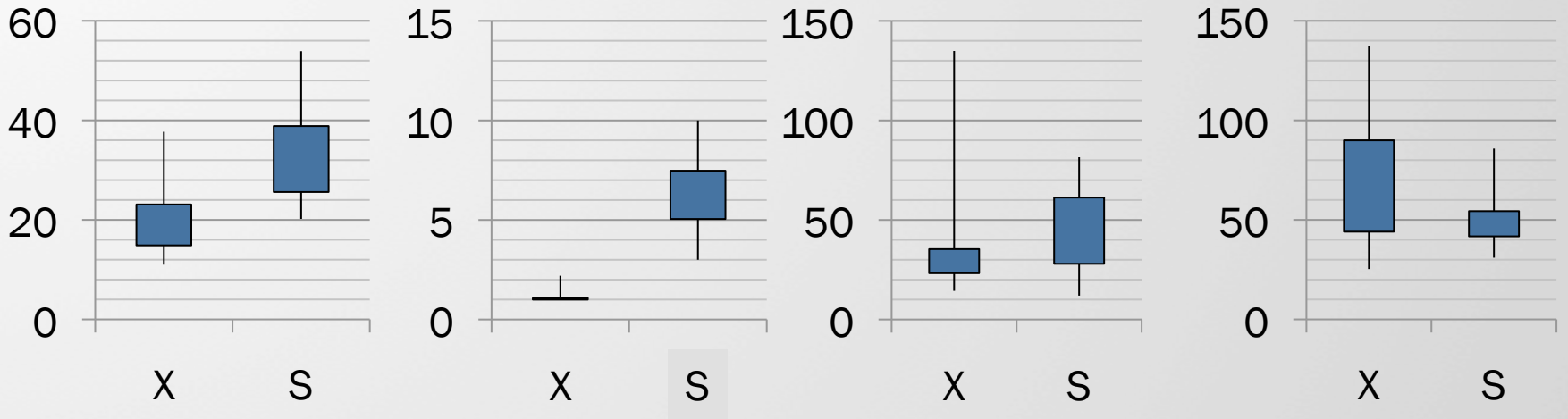
- Time, error, preference

RESULTS

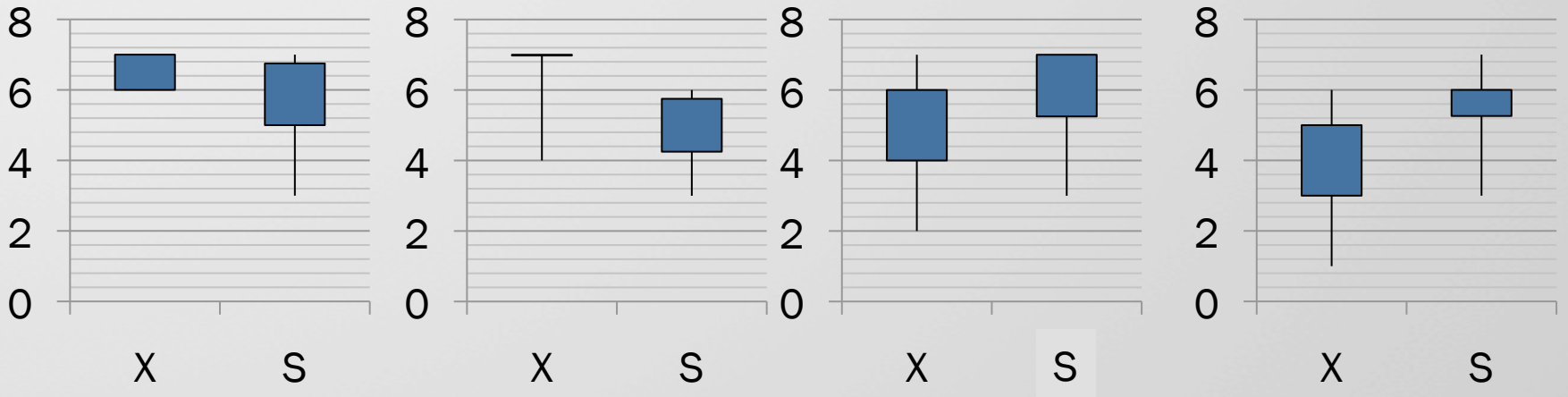
X = Exact Search , S = Similarity-based Search

Sequence | Counting | Time const. | Uncertainty

Speed (seconds)



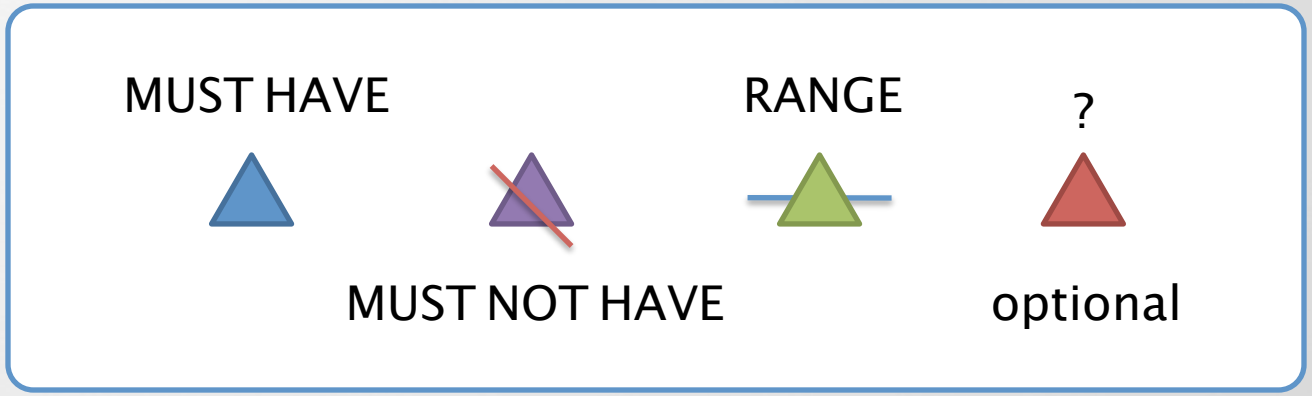
Preference (7-points likert scale)



FLEXIBLE TEMPORAL SEARCH (FTS)

PROPOSED WORK: FTS

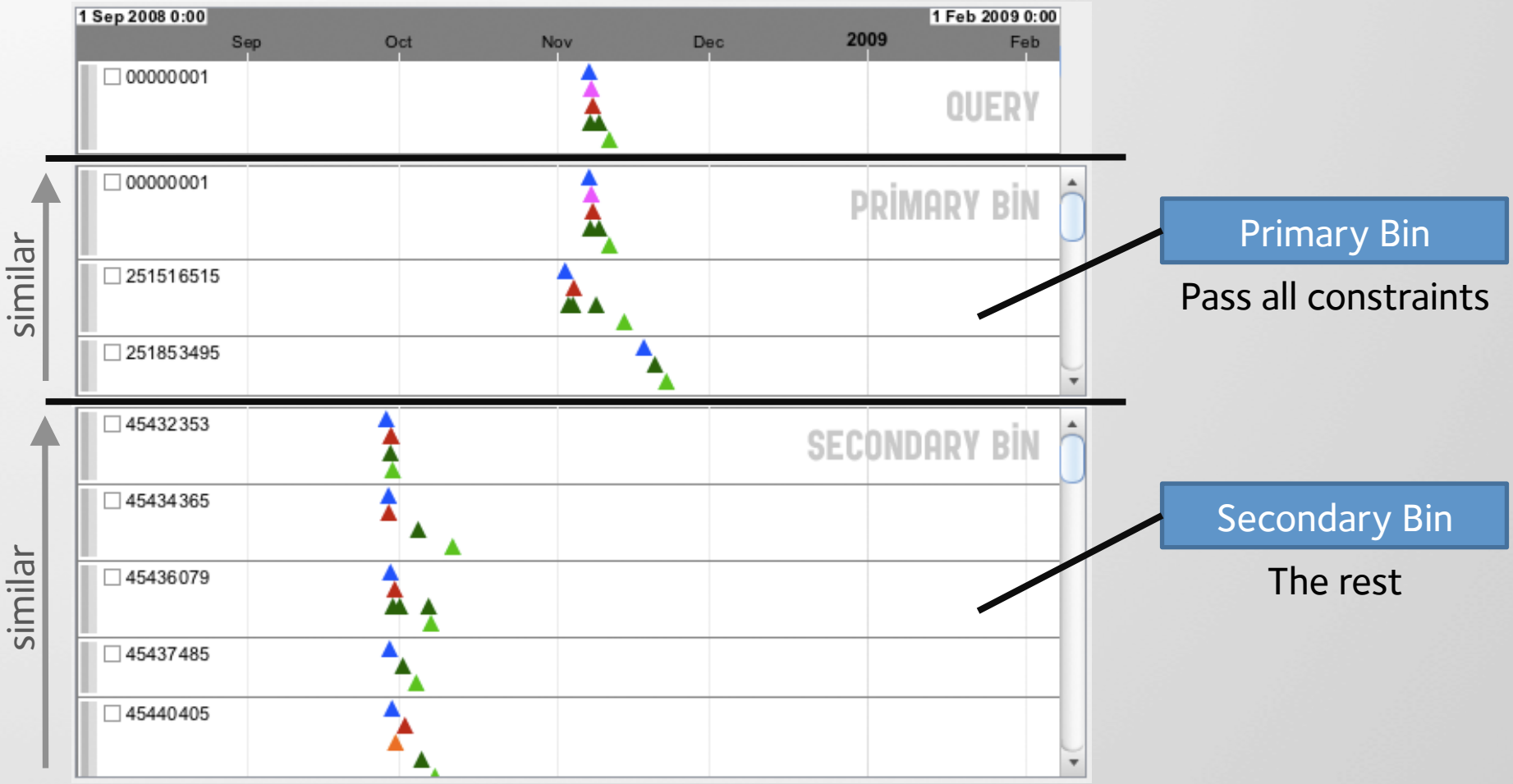
- Draw an example to query
- Specify flexible and inflexible constraints



- Algorithms

PROPOSED WORK: FTS (2)

- Combine exact and similarity-based search



RESEARCH QUESTIONS

1. How to support the users when they are uncertain about what they are looking for?
2. How to provide an overview of event sequences for temporal categorical data?

WHEN YOU READ A PAPER…

- What is it about?

Abstract

A useful starting point for designing advanced graphical user interfaces is the Visual Information-Seeking Mantra: overview first, zoom and filter, then details on demand. But this is only a starting point in trying to understand the rich and varied set of information visualizations that have been proposed in recent years. This paper offers a task by data type taxonomy with seven data types (one-, two-, three-dimensional data, temporal and multi-dimensional data, and tree and network data) and seven tasks (overview, zoom, filter, details-on-demand, relate, history, and extracts).

keys), are being pushed aside by newer notions of information gathering, seeking, or visualization and data mining, warehousing, or filtering. While distinctions are subtle, the common goals reach from finding a narrow set of items in a large collection that satisfy a well-understood information need (known-item search) to developing an understanding of unexpected patterns within the collection (browse) (Marchionini, 1995).

Exploring information collections becomes increasingly difficult as the volume grows. A page of information is easy to explore, but when the information becomes the size of a book, or library, or even larger, it may be difficult to locate known items or to browse to gain an overview.

INFORMATION VISUALIZATION MANTRA
OVERVIEW FIRST,
ZOOM AND FILTER,
THEN DETAILS ON DEMAND

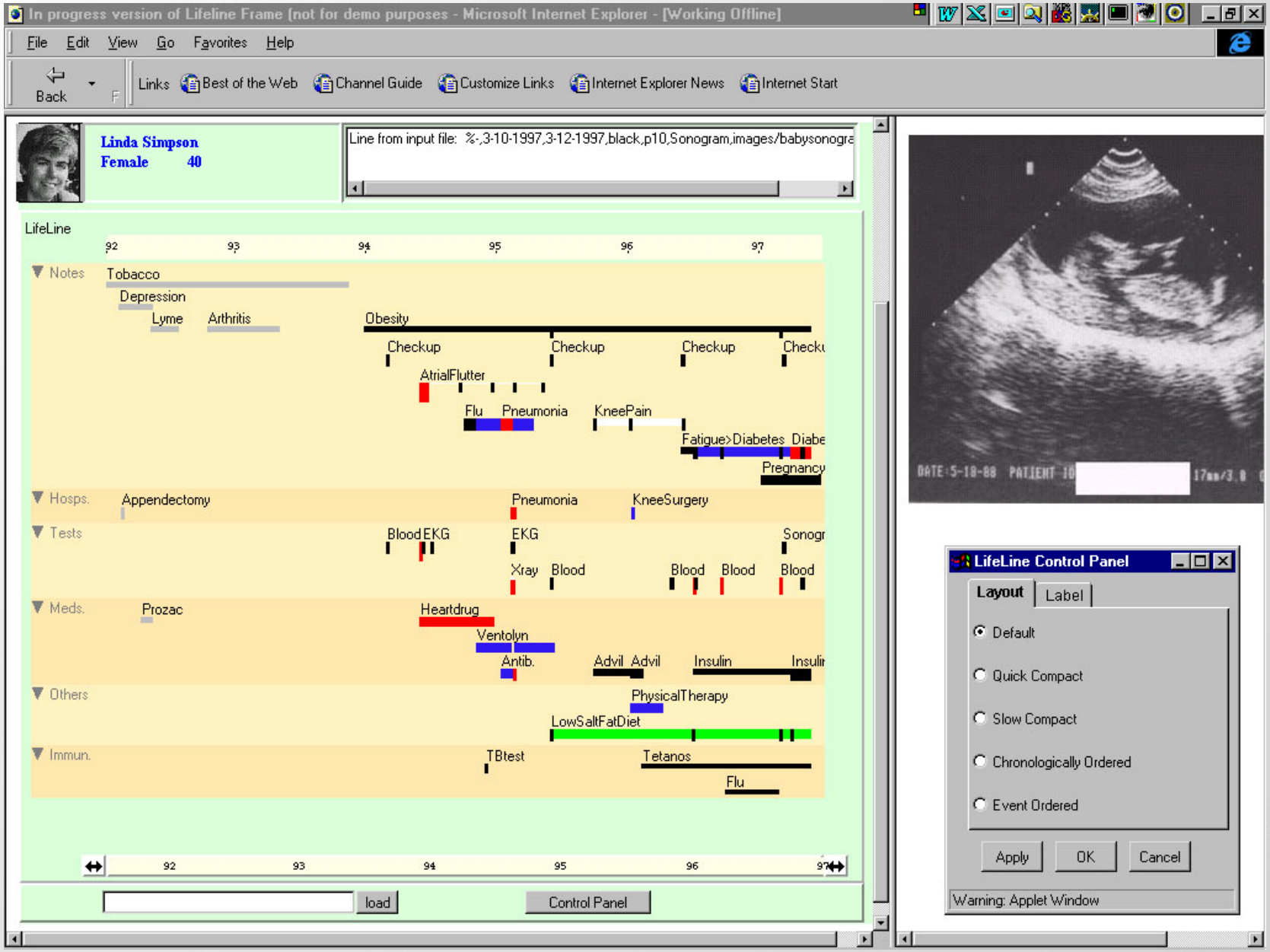
RELATED WORK

- Single-record visualizations

Patient ID: 45851737	
12/02/2008 14:26	Admit
12/02/2008 14:26	Emergency
12/02/2008 22:44	ICU
12/05/2008 05:07	Floor
12/08/2008 10:02	Floor
12/14/2008 06:19	Exit

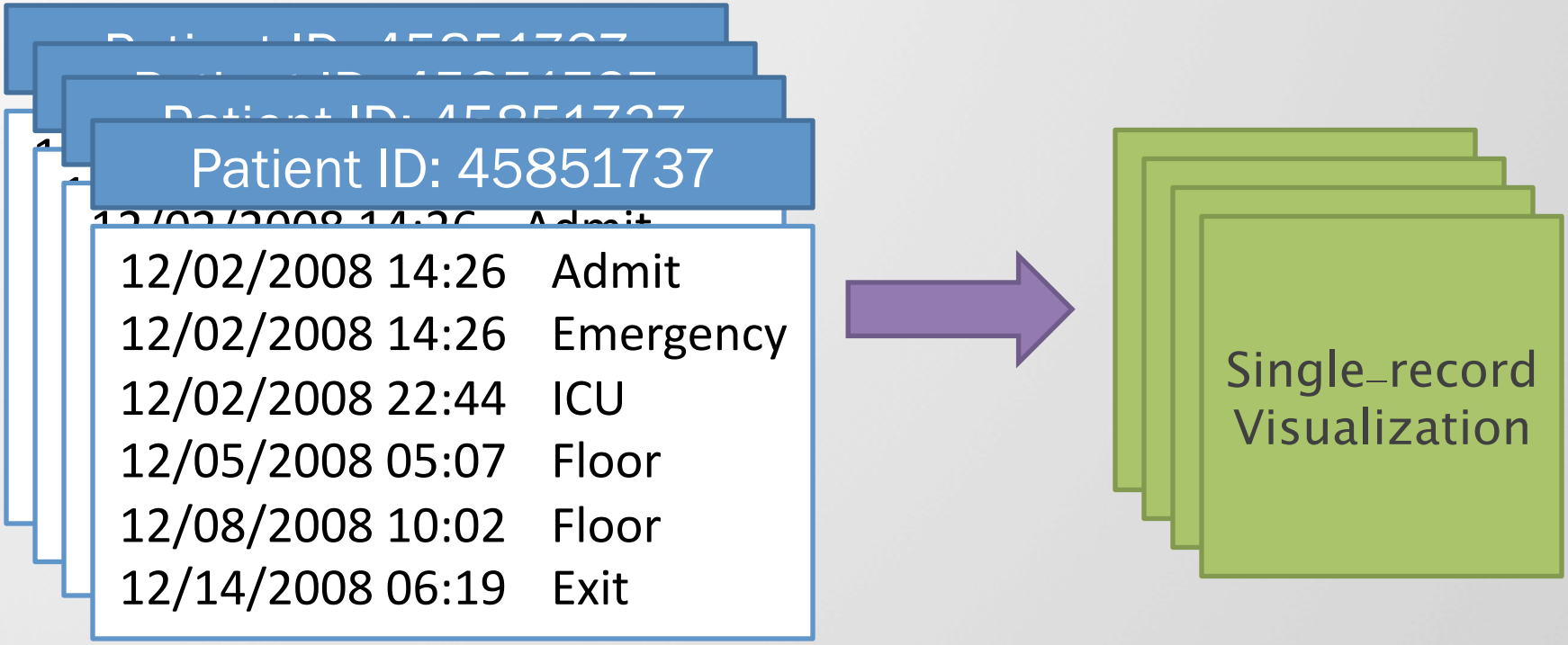
Single_record Visualization

- E.g. LifeLines, MIDGAARD, etc.

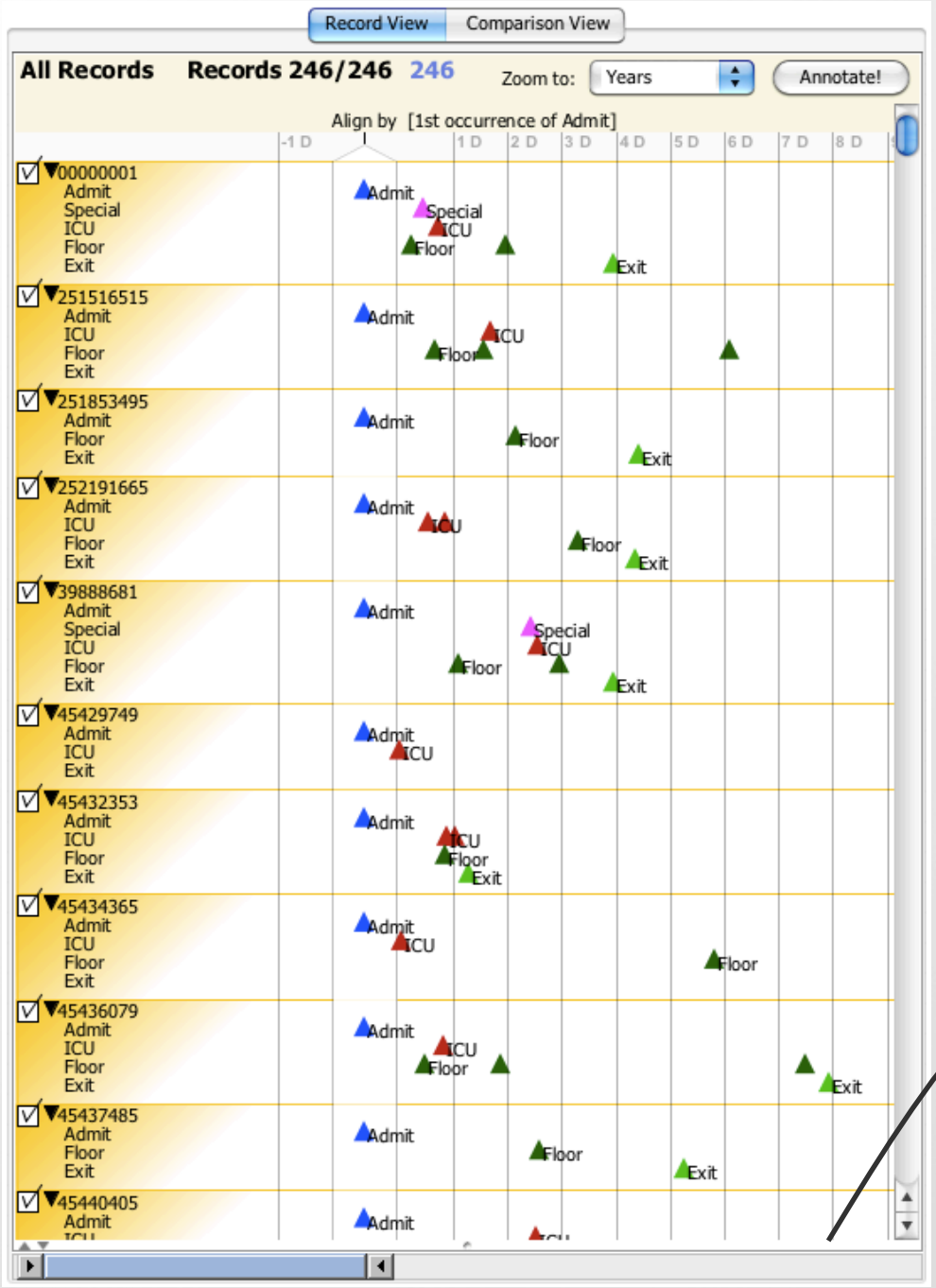

In progress version of Lifeline Frame (not for demo purposes - Microsoft Internet Explorer - [Working Offline]
File Edit View Go Favorites Help
Back
Links Best of the Web Channel Guide Customize Links Internet Explorer News Internet Start
Linda Simpson
Female 40
Line from input file: %-,3-10-1997,3-12-1997,black,p10,Sonogram,images/babysonogra
LifeLine
92 93 94 95 96 97
Notes
Tobacco
Depression
Lyme
Arthritis
Obesity
Checkup
Checkup
Checkup
AtrialFlutter
Flu
Pneumonia
KneePain
Fatigue>Diabetes
Pregnancy
Hosps.
Appendectomy
Pneumonia
KneeSurgery
Tests
Blood EKG
EKG
Xray
Blood
Blood
Blood
Blood
Meds.
Prozac
Heartdrug
Ventolyn
Antib.
Advil
Advil
Insulin
Others
PhysicalTherapy
LowSaltFatDiet
Immun.
TBtest
Tetanos
Flu
load
Control Panel
LifeLine Control Panel
Layout
Label
Default
Quick Compact
Slow Compact
Chronologically Ordered
Event Ordered
Apply
OK
Cancel
Warning: Applet Window

RELATED WORK (2)

- Multiple instances of single-record visualizations



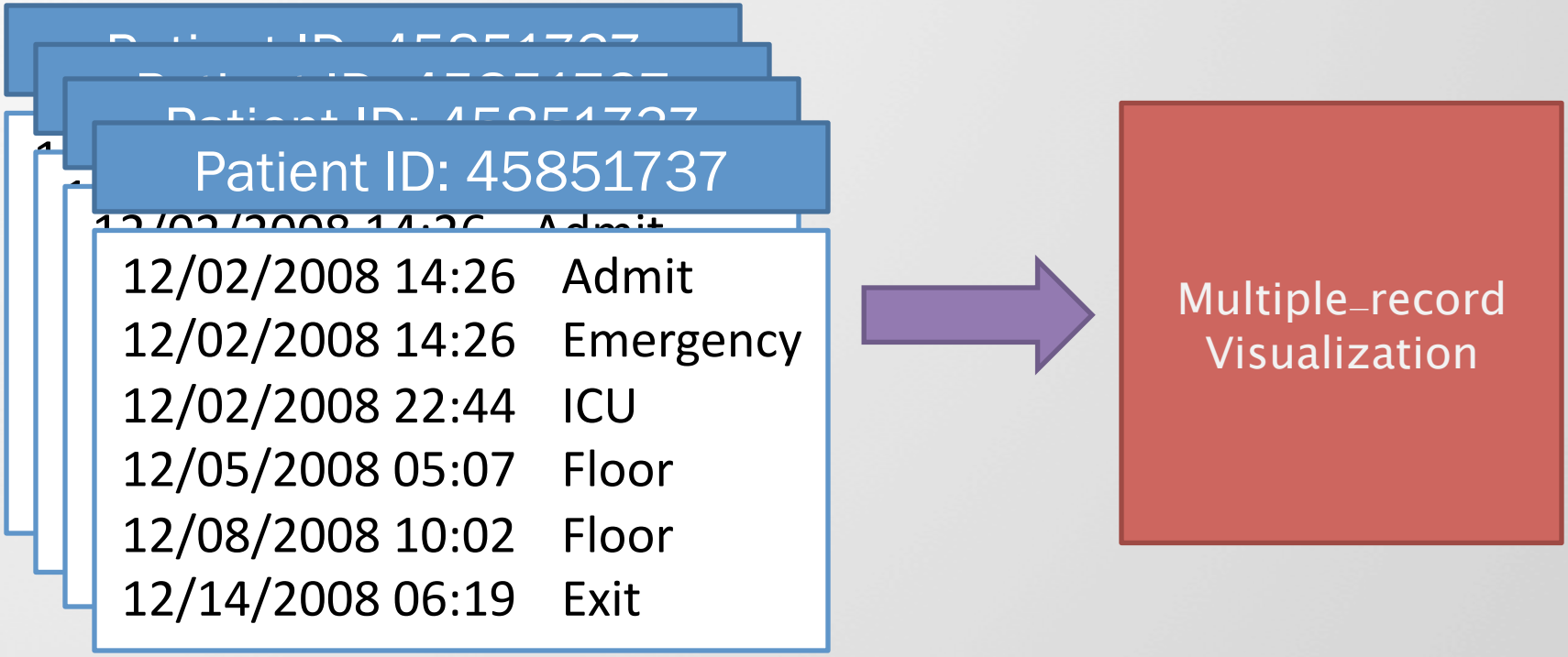
- E.g. LifeLines2, Continuum, ActiviTree, etc.



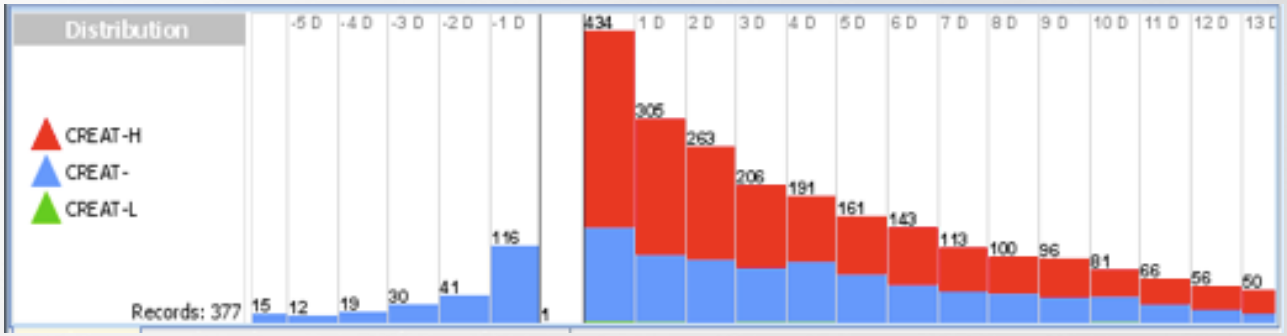
More space please....

RELATED WORK (3)

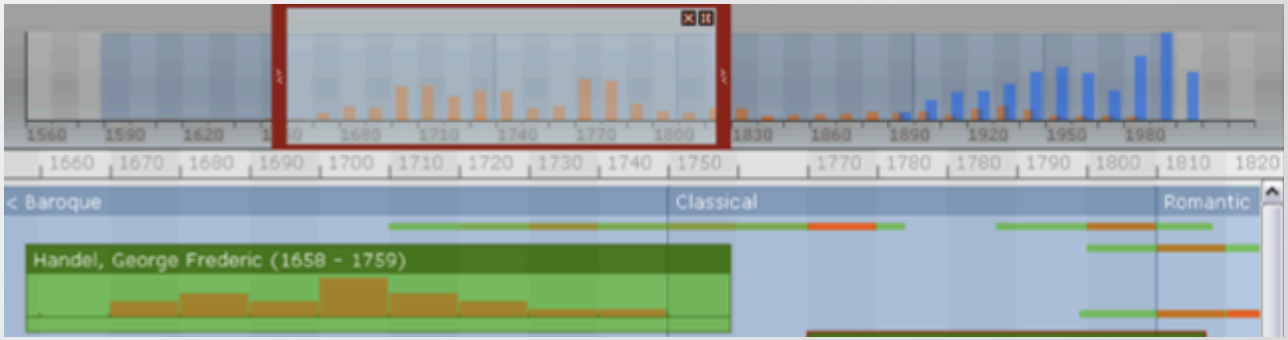
- Multiple-record visualizations



- E.g. LifeLines2, Continuum



LifeLines2's Temporal Summary [Wang et al. 2009]



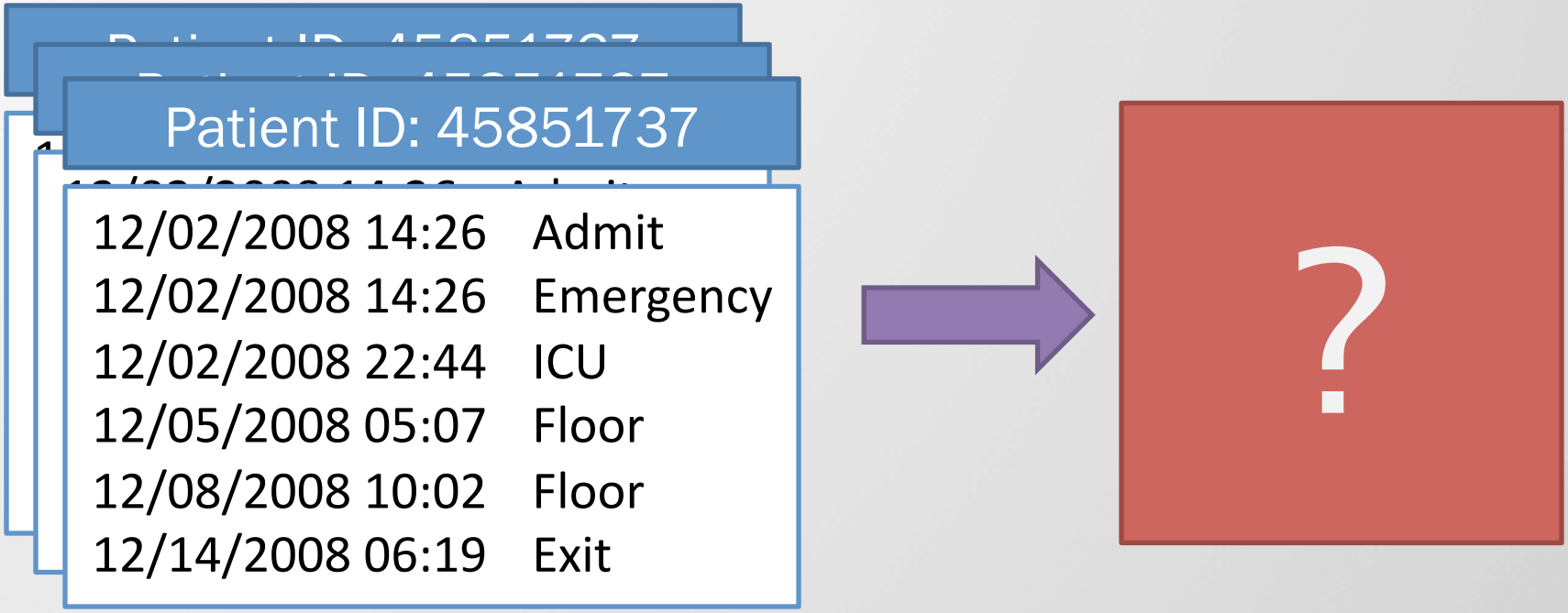
Continuum's Histogram [Andre 2007]

Not so useful with many event types

Does not help exploring event sequences

OVERVIEW OF EVENT SEQUENCES

- Scalability vs. Loss of information



- Goal: Show common trends, outliers

LIFEFLOW

AGGREGATE

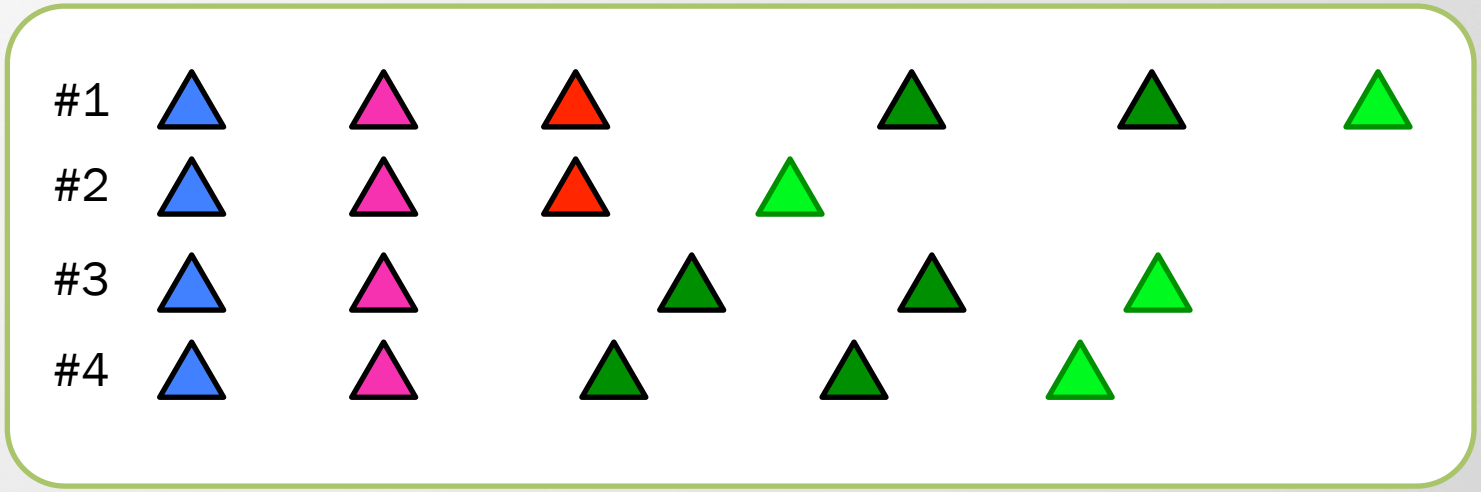
Merge multiple records into tree

VISUALIZE

Display the tree

AGGREGATE

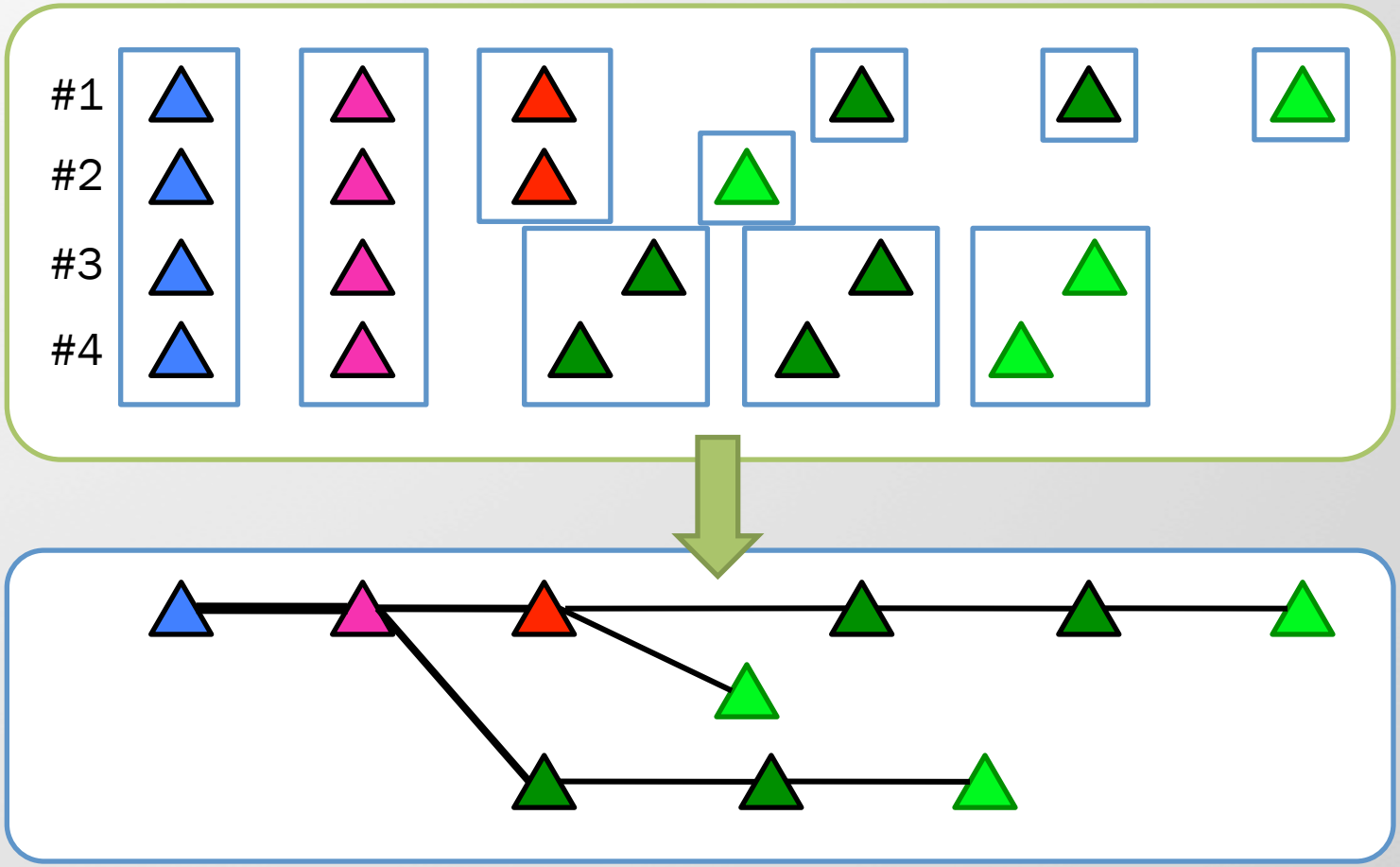
- Aggregate by prefix



Example with 4 records

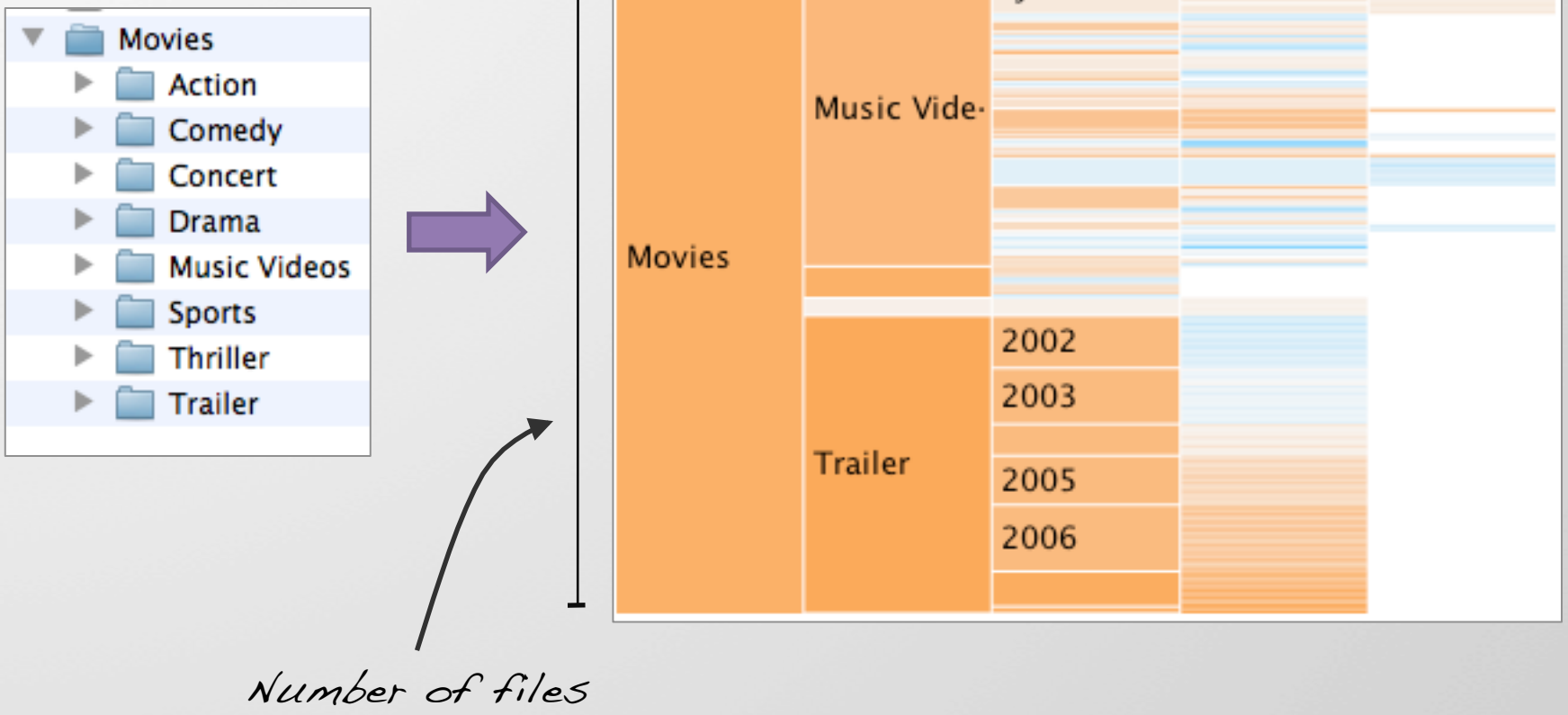
AGGREGATE

- Aggregate by prefix



VISUALIZE

- Inspired by the Icicle tree



VISUALIZE (2)

- Modify: Use horizontal axis to represent time
- Video

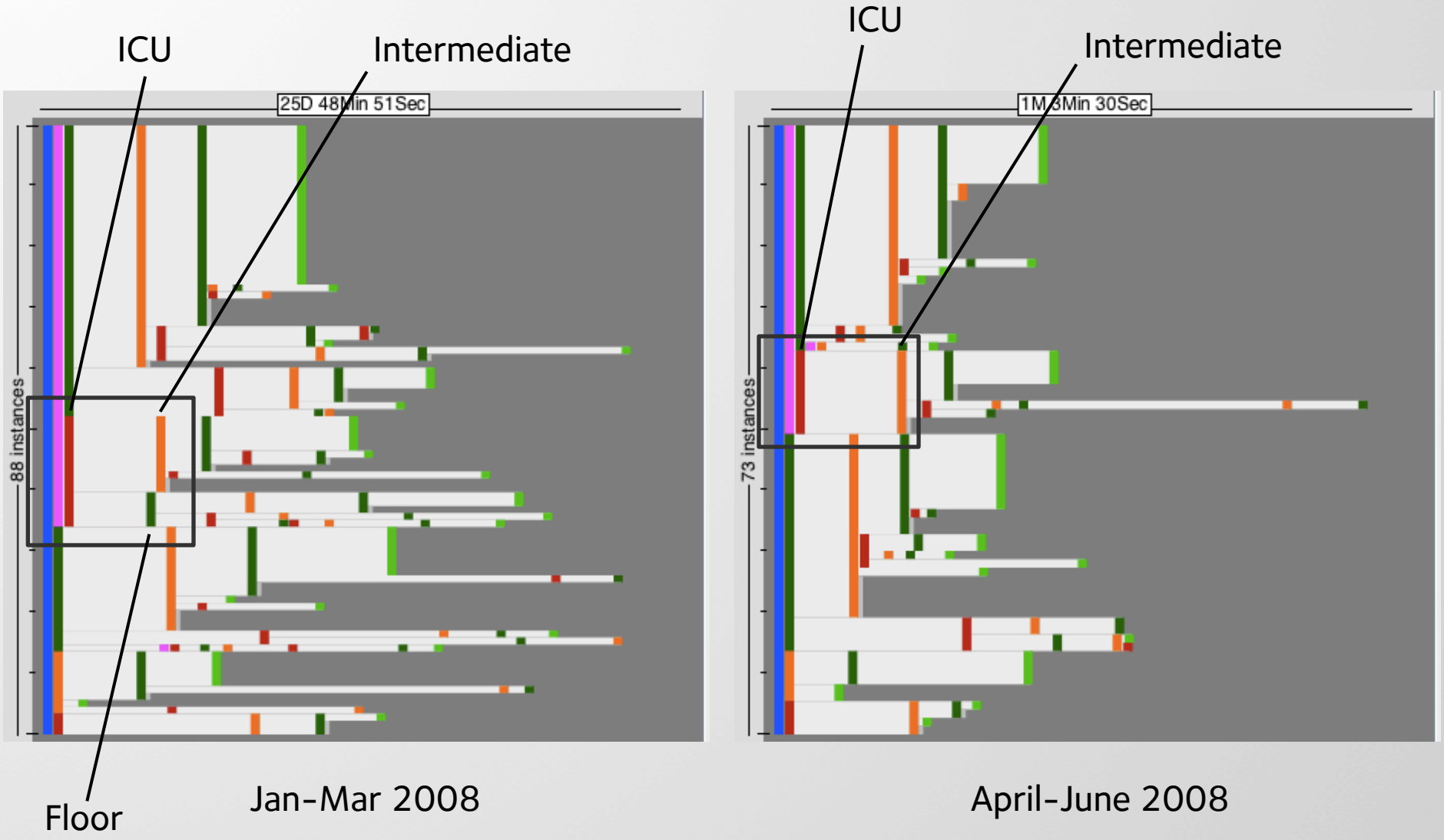
DEMO – LIFEFLOW

When the lines are combined into flow

LIFEFLOW

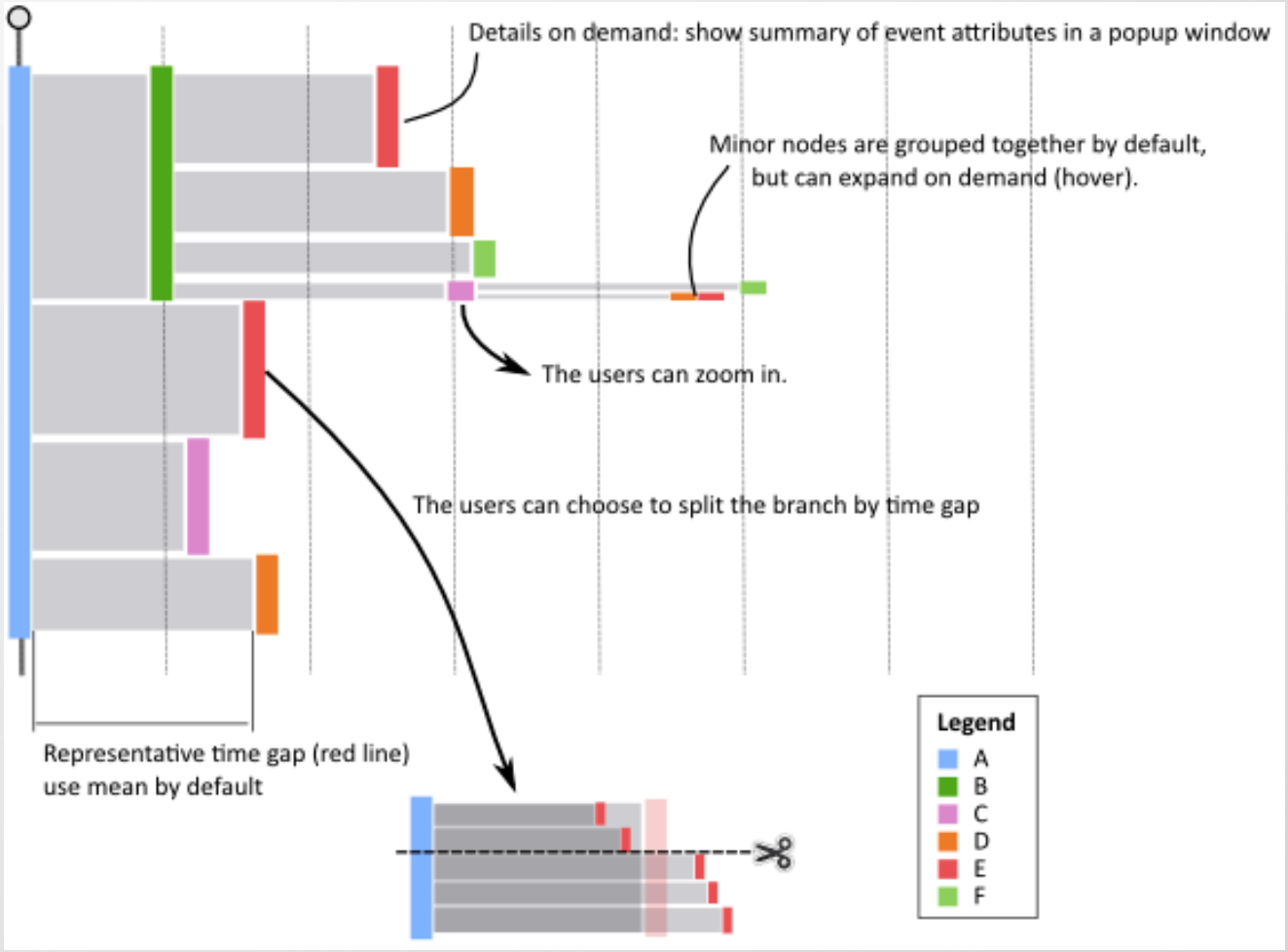
- Overview
- See trends
- Spot outliers
- Query
- Compare data

COMPARISON


ICU
Intermediate
25D 48Min 51Sec
88 instances
Floor
Jan-Mar 2008
ICU
Intermediate
1M 3Min 30Sec
73 instances
April-June 2008

PROPOSED WORK: LIFEFLOW

- Designing interaction

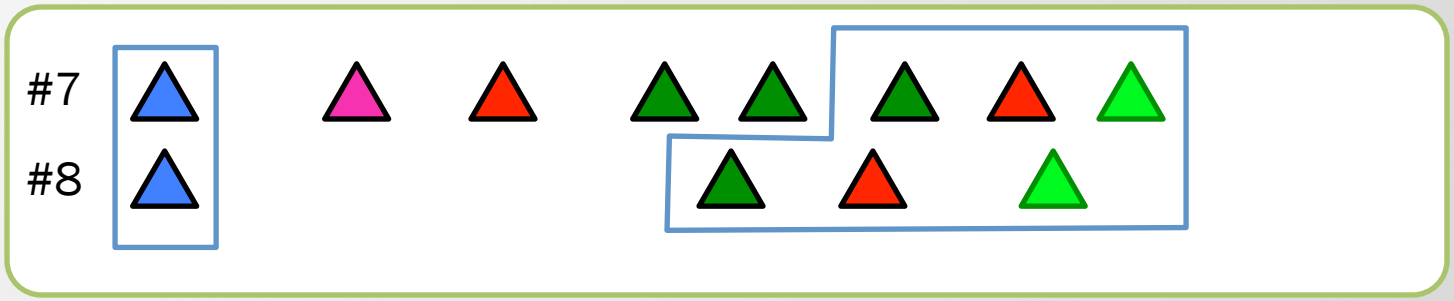


PROPOSED WORK: LIFEFLOW

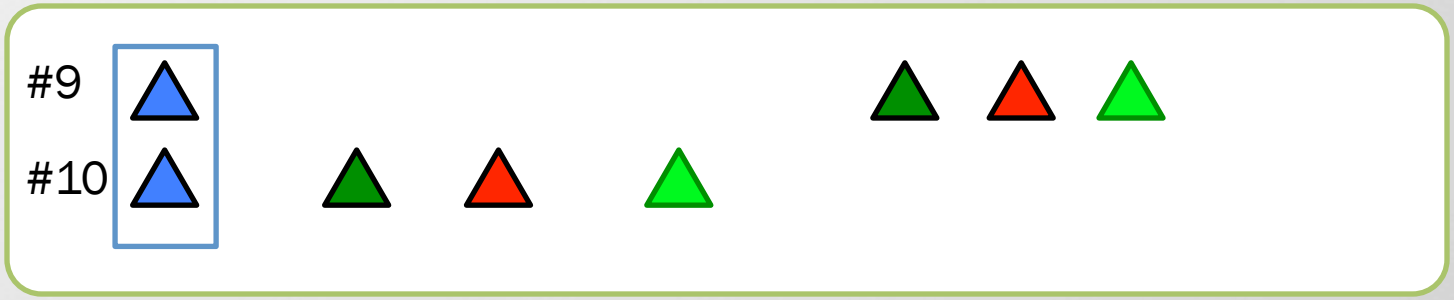
- Primary
  - Including non-temporal attributes
    - Age, Gender, etc.
  - Link with Flexible Temporal Search
- Optional
  - Rank-by-feature
  - Complex aggregation
  - Multiple alignments

PROPOSED WORK: LIFEFLOW

- Complex aggregation



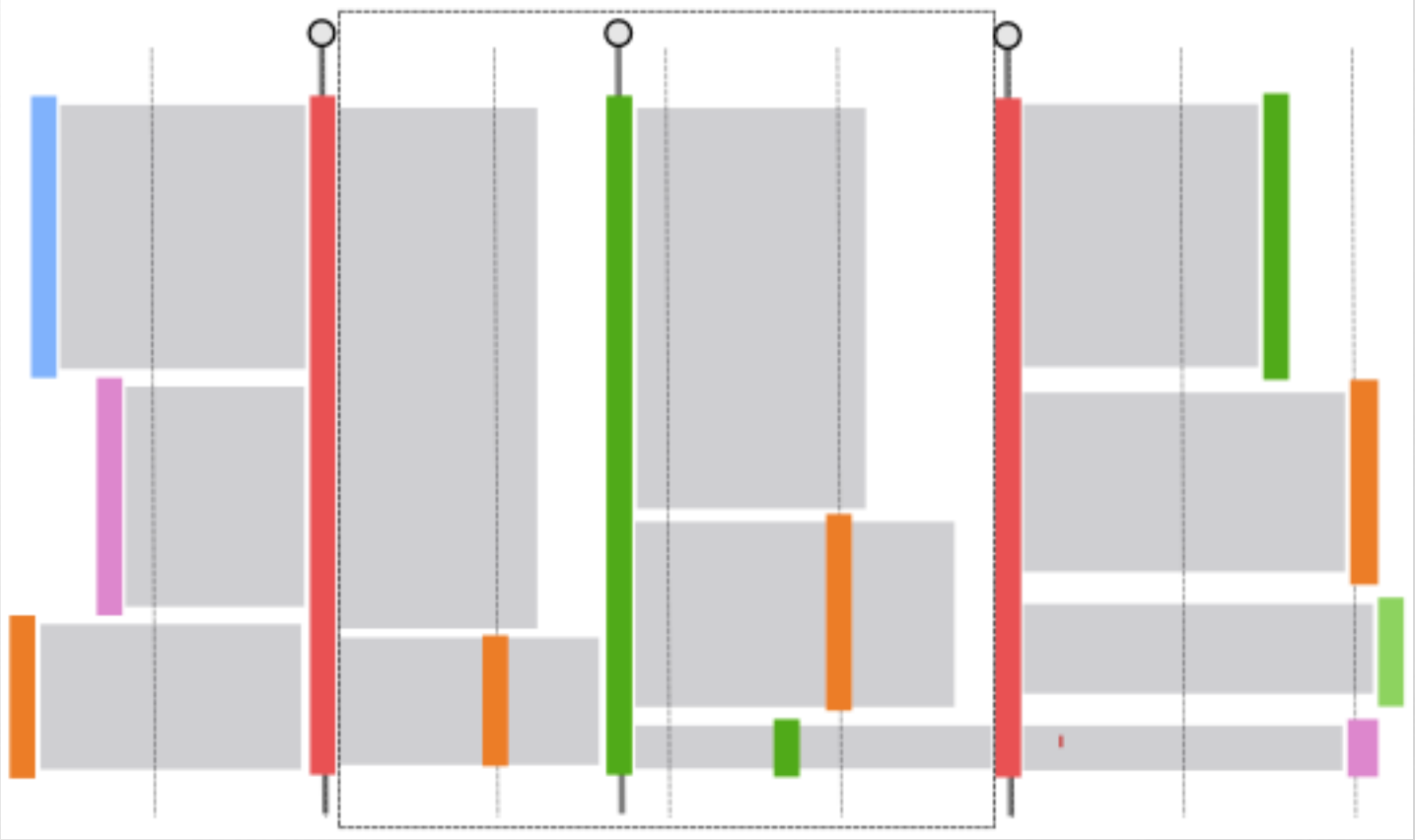
Not just merge by prefix



Do not merge if the time gap is too different

PROPOSED WORK: LIFEFLOW

- Multiple alignments



EVALUATION PLAN

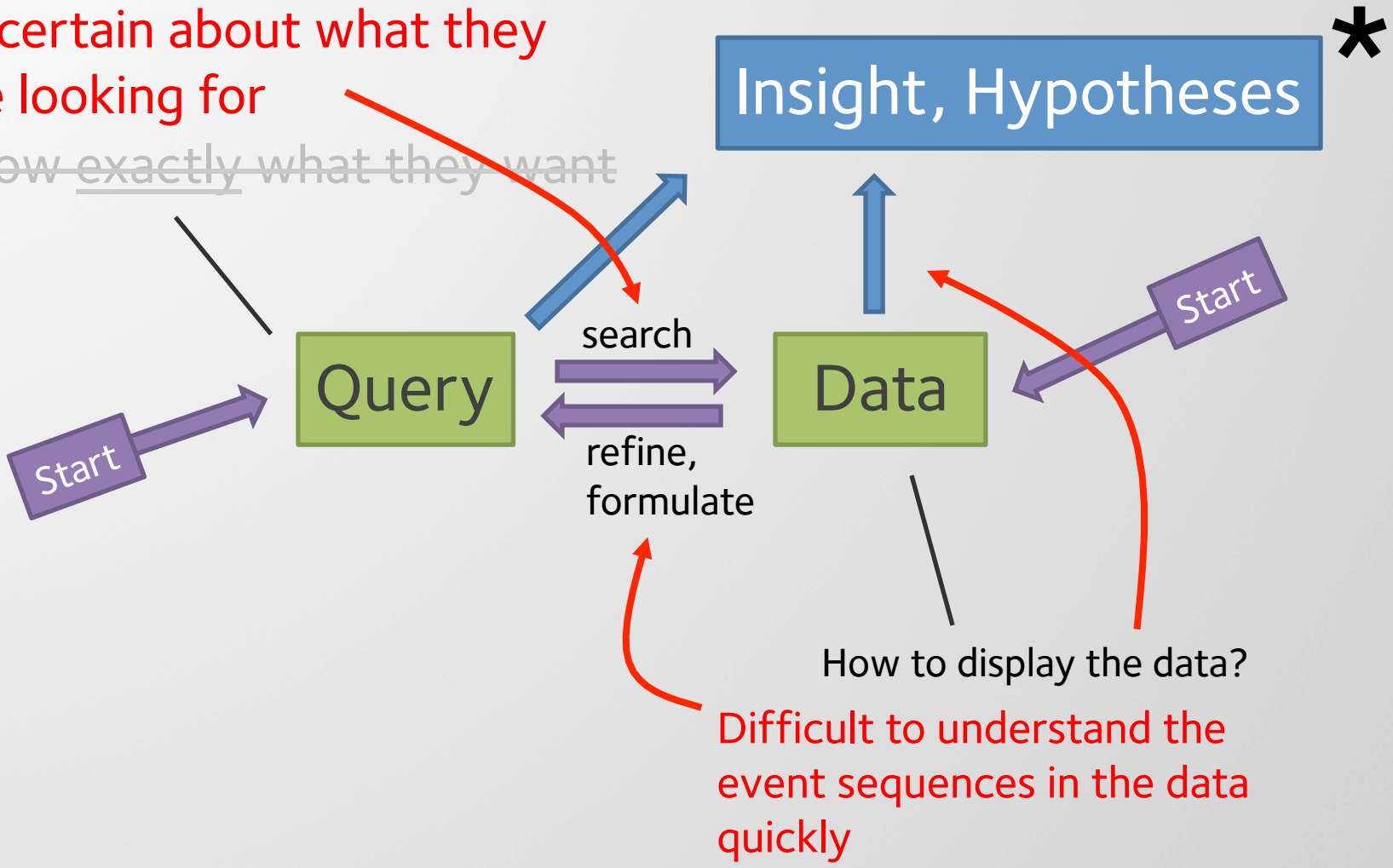
- Flexible Temporal Search
  - Controlled Experiment
    - 18-20 participants / 60-90 mins
- LifeFlow
  - Usability Study
    - 15-20 participants / 60-90 mins
  - MILCs
    - 3-5 domain experts / 5-12 weeks



CONCLUSION

Summary and expected contributions

EXPLORATORY SEARCH MODEL



RESEARCH QUESTIONS

1. How to support the users when they are uncertain about what they are looking for?

   Flexible Temporal Search

2. How to provide an overview of event sequences for temporal categorical data?

   LifeFlow: an overview visualization

EXPECTED CONTRIBUTIONS

1. Design of visual representations, user interfaces and interaction techniques
2. Algorithms for flexible temporal search
3. Evaluation results
4. Open new directions for exploring temporal categorical data

ACKNOWLEDGEMENT

Dr. Phuong Ho, Dr. Mark Smith, David Roseman

Washington Hospital Center

http://www.whcenter.org

National Institutes of Health (NIH)

http://www.nih.gov

Michael Pack, Michael VanDaniker

Center for Advanced Tranportation Technology Lab

(CATT Lab)

http://www.cattlab.umd.edu

Q&A

Questions?

THANK YOU

Thank you