Aligning Temporal Data by Sentinel Events:

Discovering Patterns in Electronic Health Records

Taowei Wang, Catherine Plaisant,
Alex Quinn, Roman Stanchak, Ben Shneiderman
University of Maryland

Shawn Murphy
Massachusetts General Hospital

LifeLines: Overview of Patient Record

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, Checku, AtrialFlutter, Flu, Pneumonia, KneePain, Fatigue>Diabetes, Diabe, Pregnancy

Hosps.: Appendectomy, Pneumonia, KneeSurgery

Tests: BloodEKG, EKG, Sonogr, Xray, Blood, Blood, Blood, Blood

Meds.: Prozac, Heartdrug, Ventolyn, Antib., Advil, Ad

Others: Ph, LowSaltFatDiet

Immun.: TBtest

92 93 94 95

load | Control Panel

LifeLine Control Panel

Layout | Label

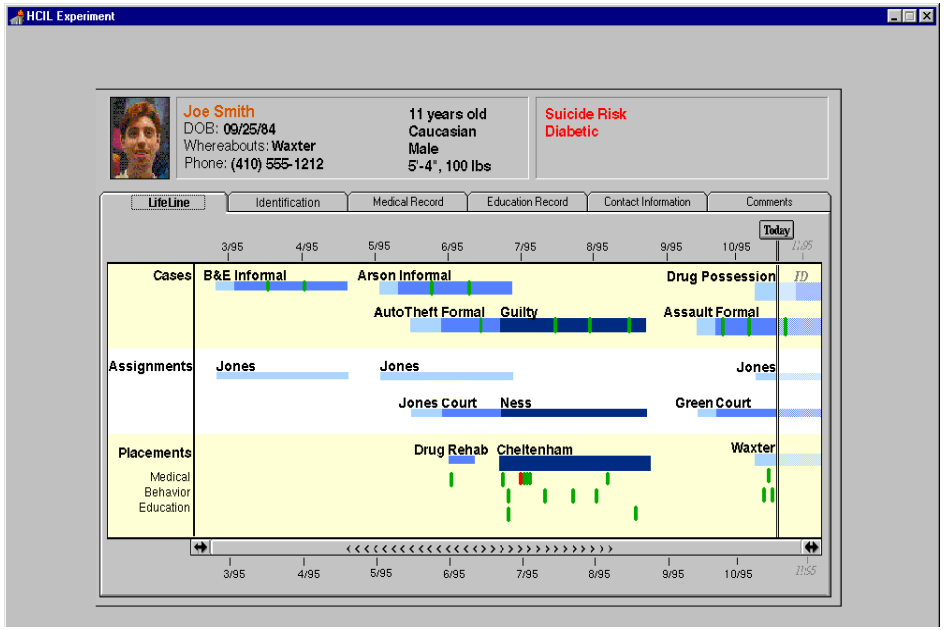
Default

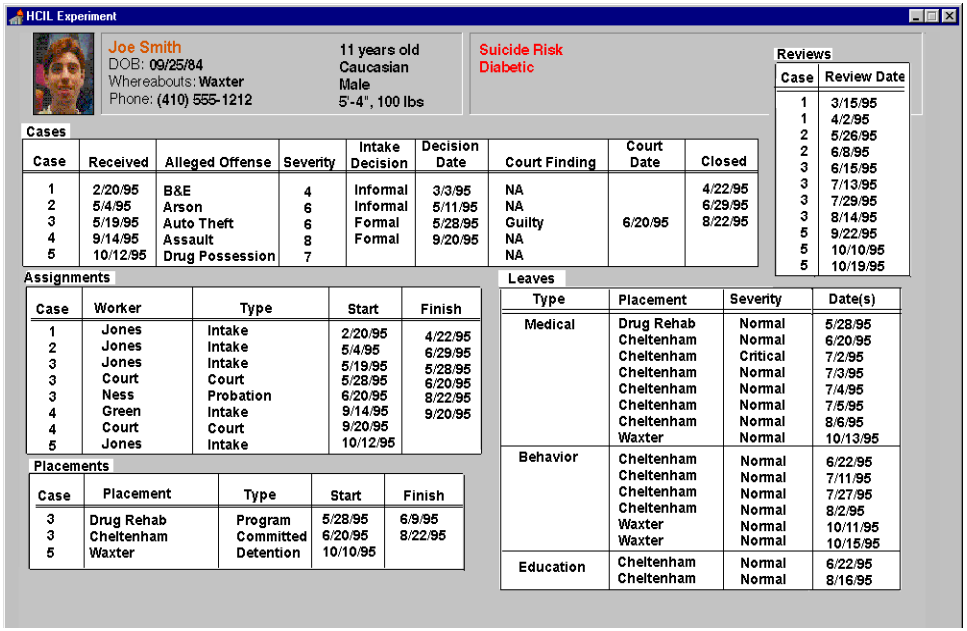
10 years ago

Lifelines
Single patient record

HCIL

Measured benefits over tabular display





www.cs.umd.edu/hcil/lifelines

Single record → Millions of records

- Large databases of Electronic Health Records (EHRs)

- Observational studies
  (i.e. using EHRs for clinical research - rather that clinical trials)
- Recruitment for clinical trails
- Hospital performance monitoring
- Alarm design and testing
- etc.

Often involve temporal comparison relative to an important event
(e.g. heart attack, start of a treatment, 1st diagnosis of cancer)

LifeLines: Overview of Patient Record

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, Checku, AtrialFlutter, Flu, Pneumonia, KneePain, Fatigue>Diabetes, Diabe, Pregnancy

Hosps.: Appendectomy, Pneumonia, KneeSurgery

Tests: BloodEKG, EKG, Sonogr, Xray, Blood, Blood, Blood, Blood

Meds.: Prozac, Heartdrug, Ventolyn, Antib., Advil, Ad

Others: Ph, LowSaltFatDiet

Immun.: TBtest

92 93 94 95

load | Control Panel

LifeLine Control Panel

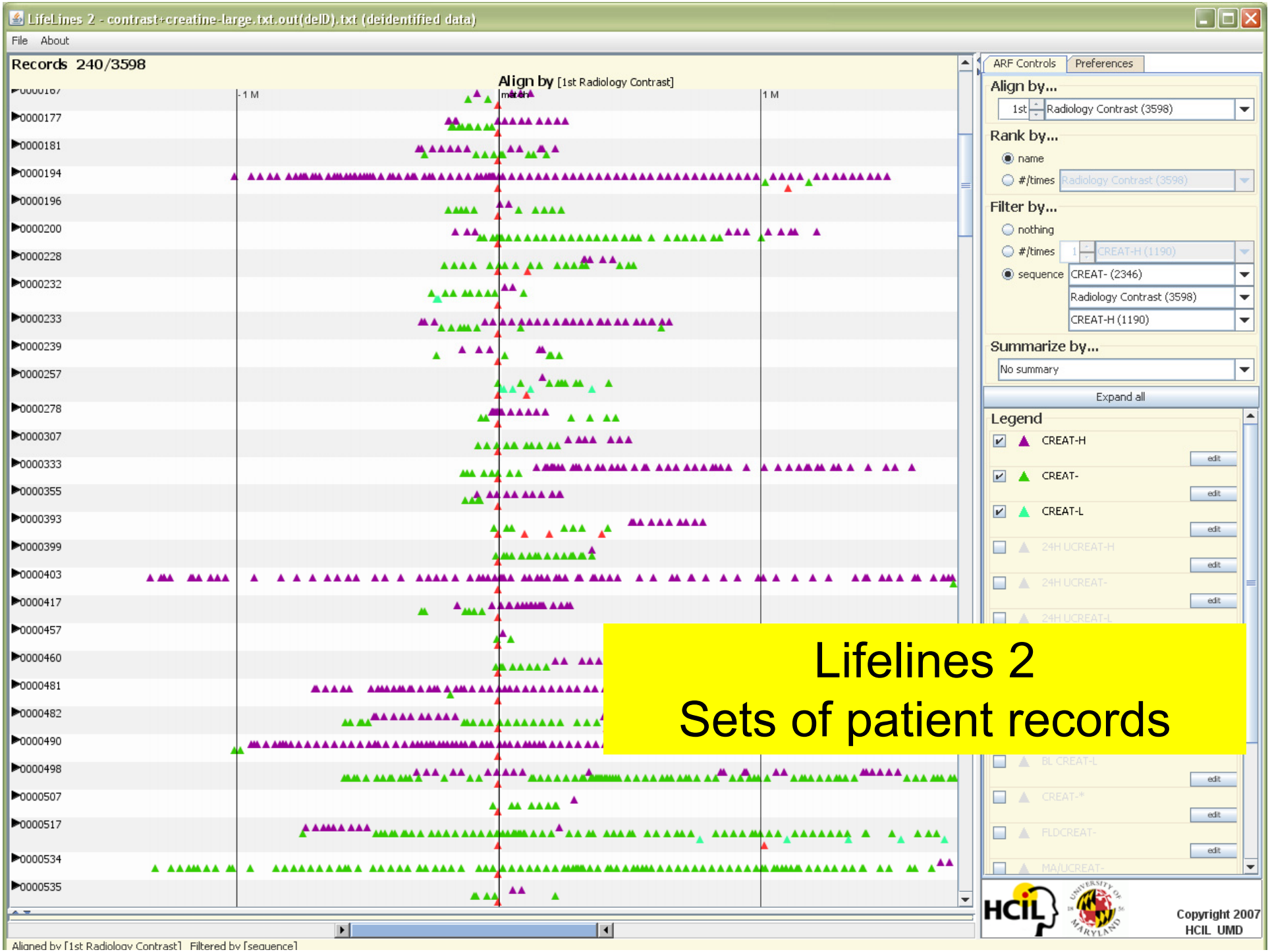
Layout | Label

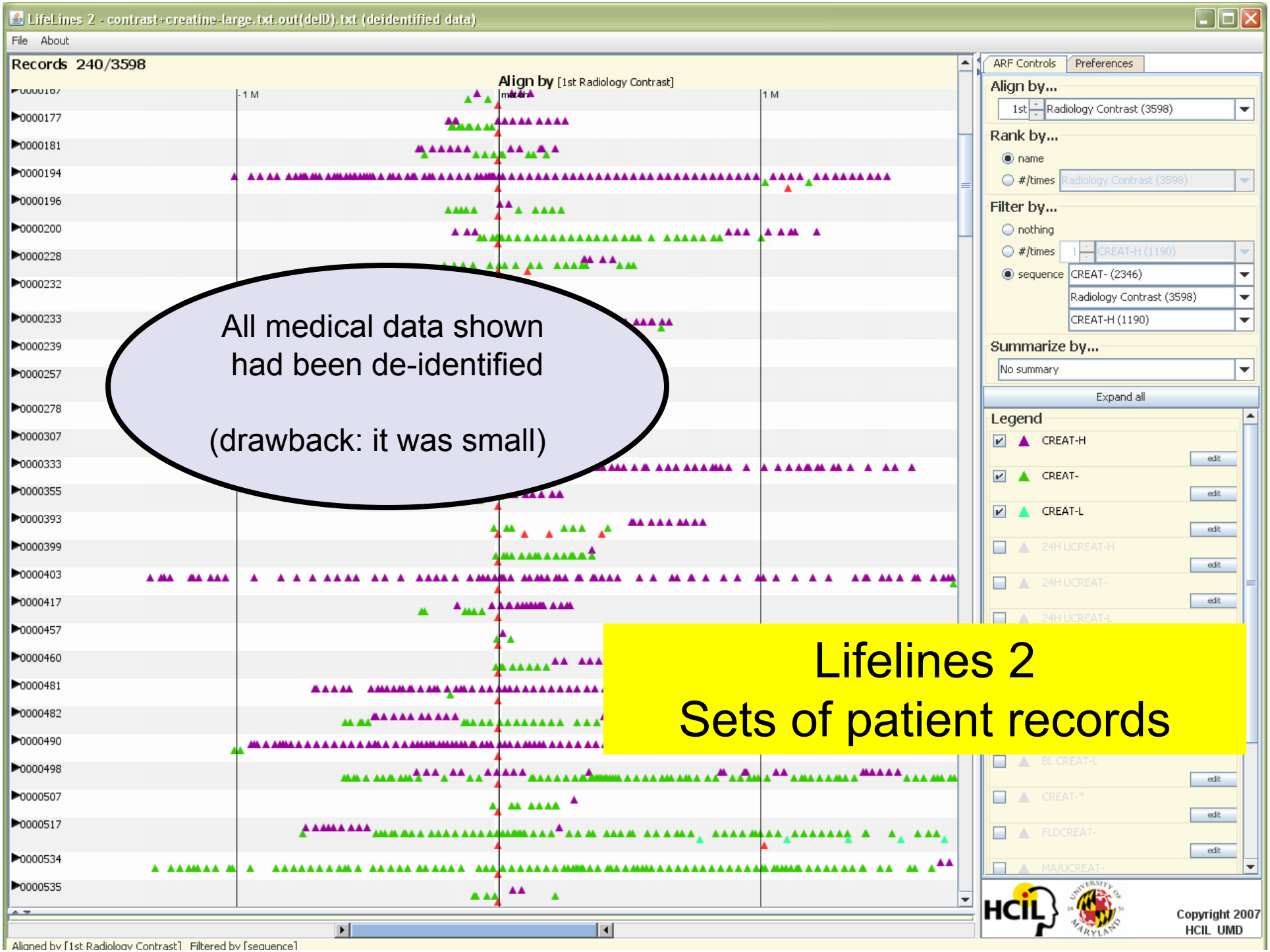
Default

10 years ago

Lifelines
Single patient record

HCIL


LifeLines 2 - contrast+creatine-large.txt.out(deID).txt (deidentified data)
File About
Records 240/3598
Align by [1st Radiology Contrast]
-1 M
1 M
ARF Controls
Preferences
Align by...
1st
Radiology Contrast (3598)
Rank by...
name
#/times
Radiology Contrast (3598)
Filter by...
nothing
#/times
CREAT-H (1190)
sequence
CREAT- (2346)
Radiology Contrast (3598)
CREAT-H (1190)
Summarize by...
No summary
Expand all
Legend
CREAT-H
CREAT-
CREAT-L
24H UCREAT-H
24H UCREAT-
24H UCREAT-L
BL CREAT-L
CREAT-*
FLDCREAT-
MA/UCREAT-
edit
Lifelines 2
Sets of patient records
HCIL
Copyright 2007
HCIL UMD
Aligned by [1st Radiology Contrast] Filtered by [sequence]



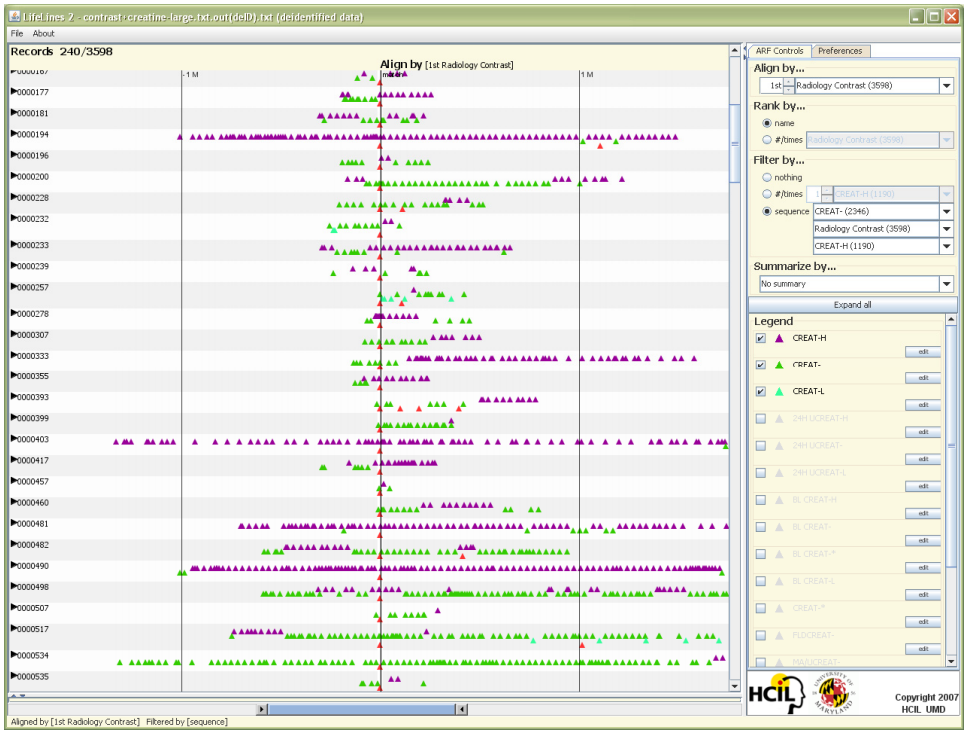
All medical data shown had been de-identified

(drawback: it was small)

Lifelines 2
Sets of patient records

Today

- Introduce powerful combination of simple operations **Align Rank and Filter**
  - Multiple records simultaneously visible
  - Align by sentinel events
  - Rank by frequency
  - Filter by events
- Measure **benefit of alignment**
- Explore representation of **intervals of validity**



Focus on categorical point data

- Examples
  - Diagnoses
  - Admission to hospital
  - Complaints: Shortness of breath
  - Tests (e.g. type: creatinine serum, low/normal/high, value)
  - Exams (e.g. type=xray, normal)

NOT:
Images
Full text of notes
Continuous numerical data (e.g. EKGs)

- Point data (not interval)

  even if some implicit interval of validity exist

Focus on categorical point data

- Examples
  - Diagnoses
  - Admission to hospital
  - Complaints: Shortness of breath
  - Tests (e.g. type: creatinine serum, low/normal/high, value)
  - Exams (e.g. type=xray, normal)

NOT:
Images
Full text of notes
Continuous numerical data (e.g. EKGs)

- Point data (not interval)

  even if some implicit interval of validity exist

Pneumonia

Focus on categorical point data

- Examples
  - Diagnoses
  - Admission to hospital
  - Complaints: Shortness of breath
  - Tests (e.g. type: creatinine serum, low/normal/high, value)
  - Exams (e.g. type=xray, normal)

NOT:
Images
Full text of notes
Continuous numerical data (e.g. EKGs)

- Point data (not interval)

  even if some implicit interval of validity exist

▲

Pneumonia

Focus on categorical point data

- Examples
  - Diagnoses
  - Admission to hospital
  - Complaints: Shortness of breath
  - Tests (e.g. type: creatinine serum, low/normal/high, value)
  - Exams (e.g. type=xray, normal)

NOT:
Images
Full text of notes
Continuous numerical data (e.g. EKGs)

- Point data (not interval)

  even if some implicit interval of validity exist

Pneumonia

Hospital discharge

Focus on categorical point data

- Examples
  - Diagnoses
  - Admission to hospital
  - Complaints: Shortness of breath
  - Tests (e.g. type: creatinine serum, low/normal/high, value)
  - Exams (e.g. type=xray, normal)

NOT:
Images
Full text of notes
Continuous numerical data (e.g. EKGs)

- Point data (not interval)

  even if some implicit interval of validity exist

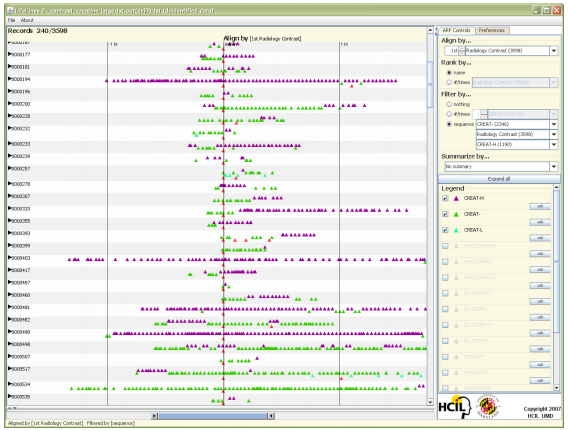
Pneumonia

Context

Search
millions of records

→ ←

Interactive
visualization of results



LifeLines2

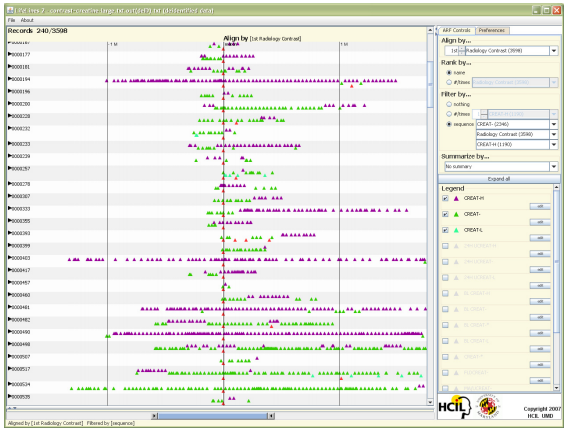
Selected subsets of the records of multiple patients

Context

Search
millions of records

Interactive
visualization of results

Writing SQL not an option!



LifeLines2

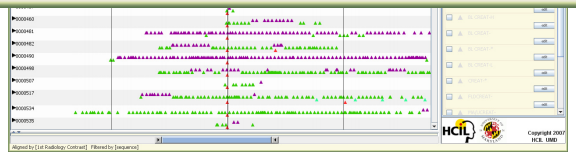
Selected subsets of the records of multiple patients

Context

```
SELECT TOP 200
 *
FROM
 contrast contrast_1,
 contrast contrast_2,
 contrast contrast_3
WHERE
 1=1 AND
 [contrast_1].[Lab Test] = 'CREAT' AND
 [contrast_2].[Value] BETWEEN '0.6' AND '1.2' AND
 [contrast_2].[Lab Accessioning Time] < [contrast_1].[Radiology Exam Time] AND [contrast_2].[Lab Accessioning Time] > dateadd(Day, -2, [contrast_1].[Radiology Exam
Time]) AND
 [contrast_1].[Account] = [contrast_2].[Account] AND
 [contrast_3].[Value] > ([contrast_2].[Value] * (1 + 50 * .01)) AND
 [contrast_3].[Value] > ([contrast_2].[Value] + 1) AND
 [contrast_3].[Lab Accessioning Time] > [contrast_1].[Radiology Exam Time] AND [contrast_3].[Lab Accessioning Time] < dateadd(Day, 5, [contrast_1].[Radiology Exam Time])
AND
 [contrast_2].[Account] = [contrast_3].[Account]
```

OK

Writing SQL not



LifeLines2

Selected subsets of the records of multiple patients

- Related work
- Quick demo
- Report on studies
- Ongoing & Future work

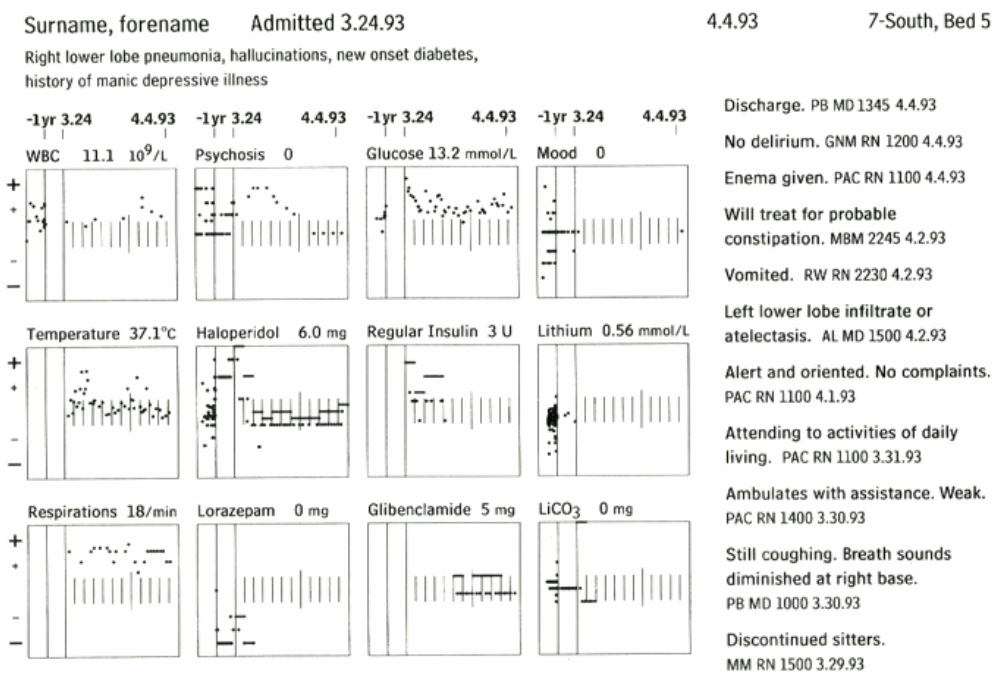
Sample of Related Work

- Many tools…
  because many types of time series, and needs

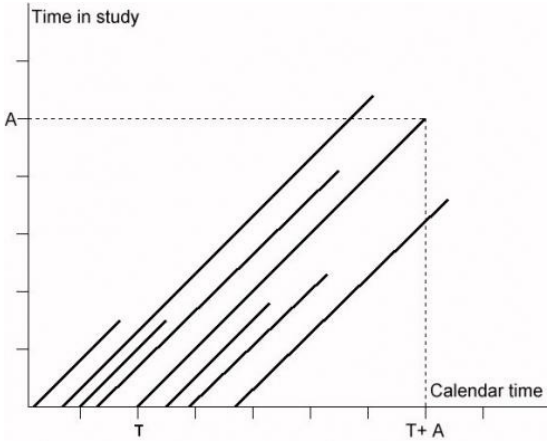
- Recent survey paper: Aigner et al, TVCG 2008

- Many static displays
- Numerical vs. Categorical data
  - Numerical
    - Single one (e.g. VizTree looking at patterns in an EKG)
    - Multiple series (e.g.TimeSearcher)
  - Categorical
    - Many modeling and query tools
    - Queries
- Alignment
  - Many examples of use
    - Manual specification of alignment point
    - Periodical data (e.g. spiral displays)

Static views



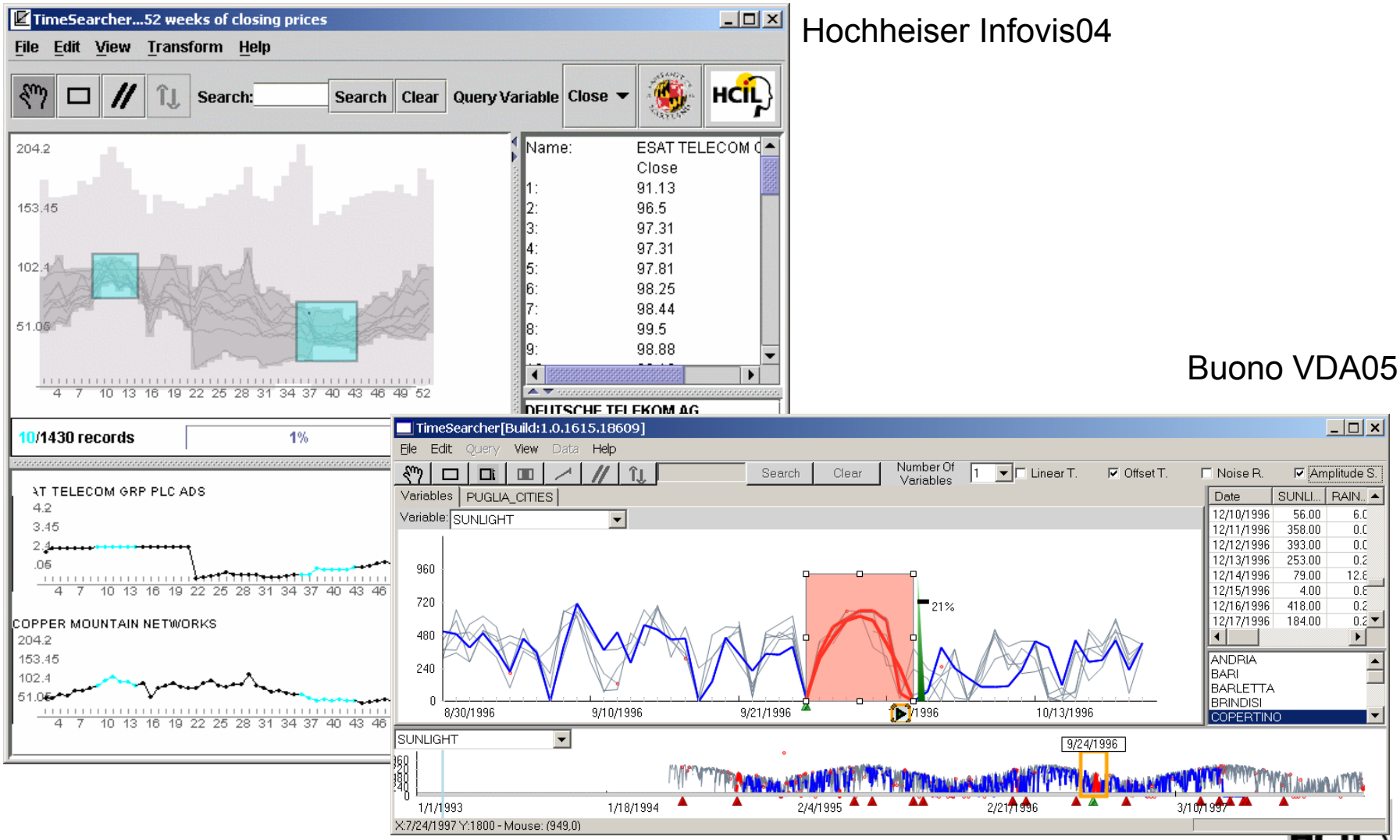
Powsner & Tufte, 1994



Lexis diagrams (Bertin)

TimeSearcher

Dynamic queries on numerical temporal data



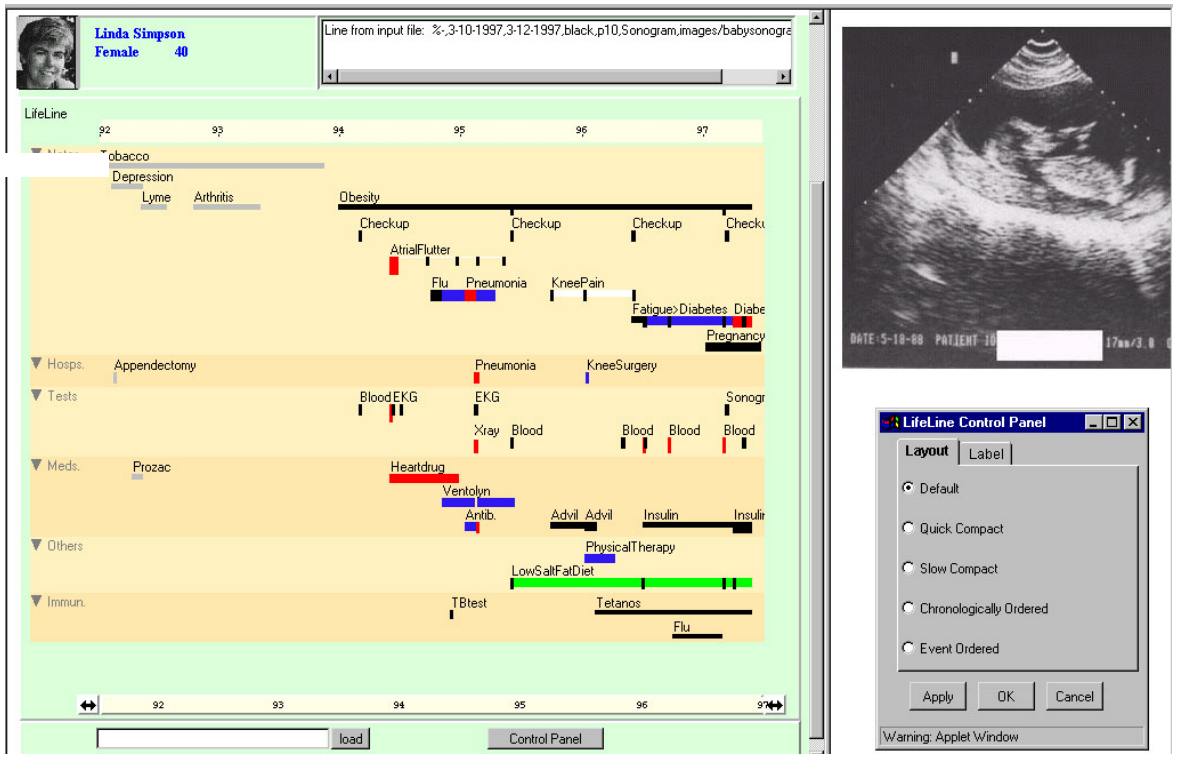
Hochheiser Infovis04

Buono VDA05

www.cs.umd.edu/hcil/timesearcher

Lifelines and improvements

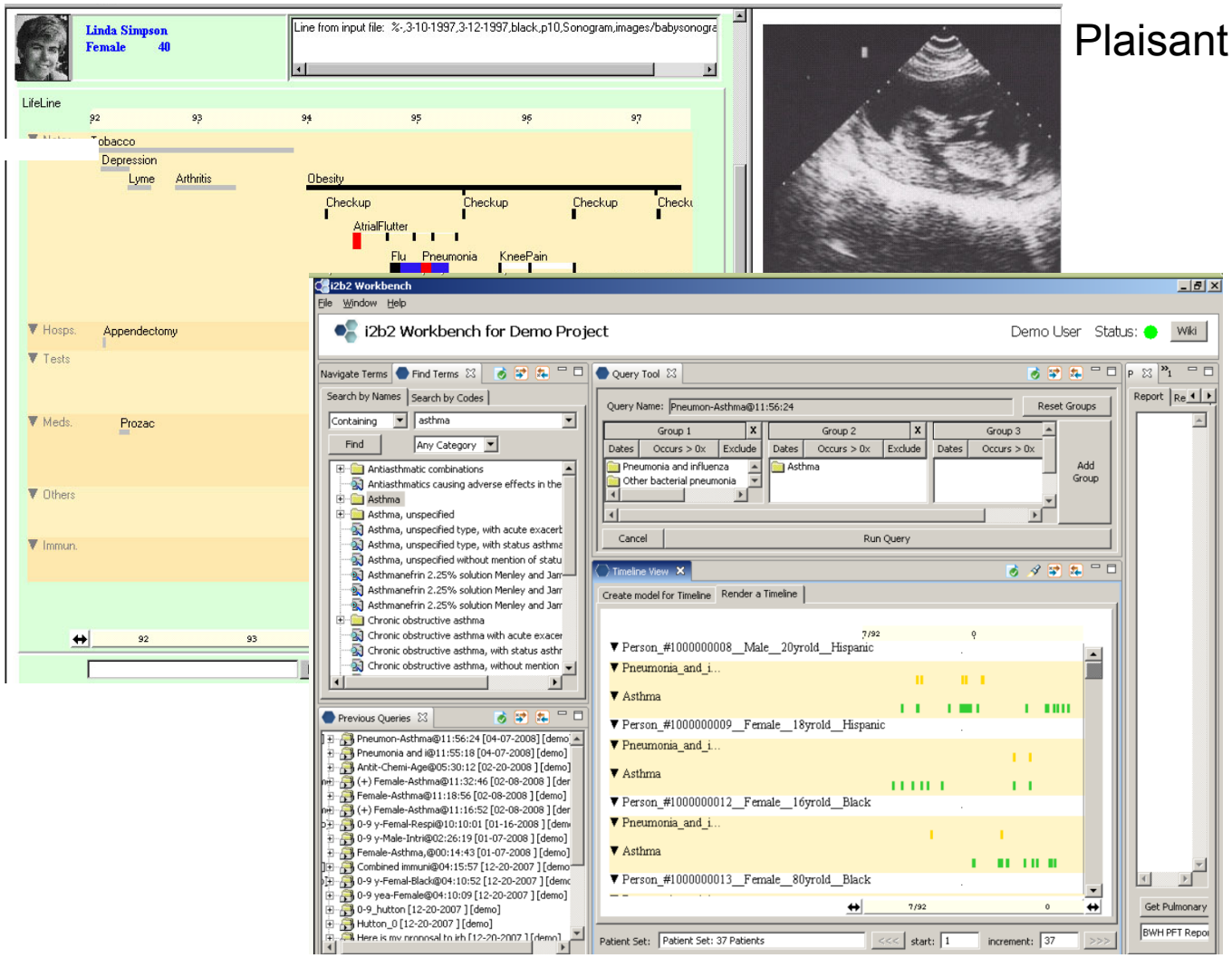
Overview of categorical and/or numerical data (semantic zoom)



Plaisant CHI 96, AMIA 98

Lifelines and improvements

Overview of categorical and/or numerical data (semantic zoom)

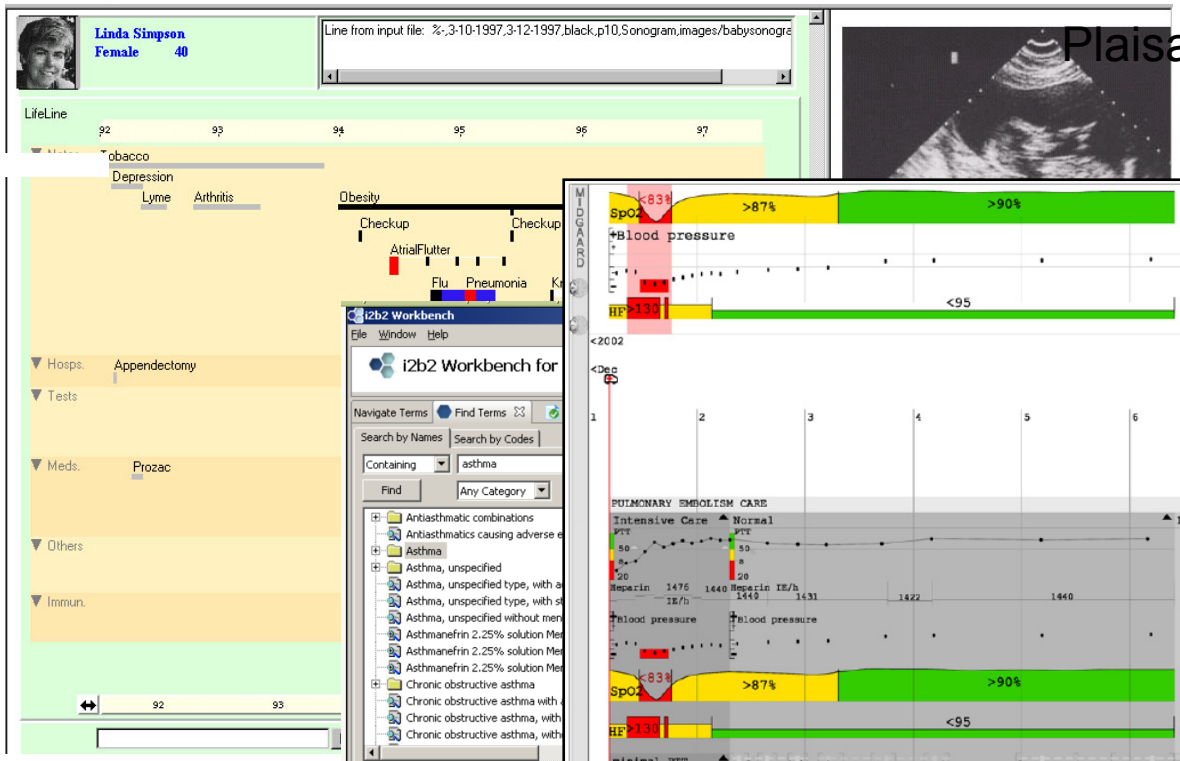


Plaisant CHI 96, AMIA 98

I2b2 (Murphy AMIA 07)

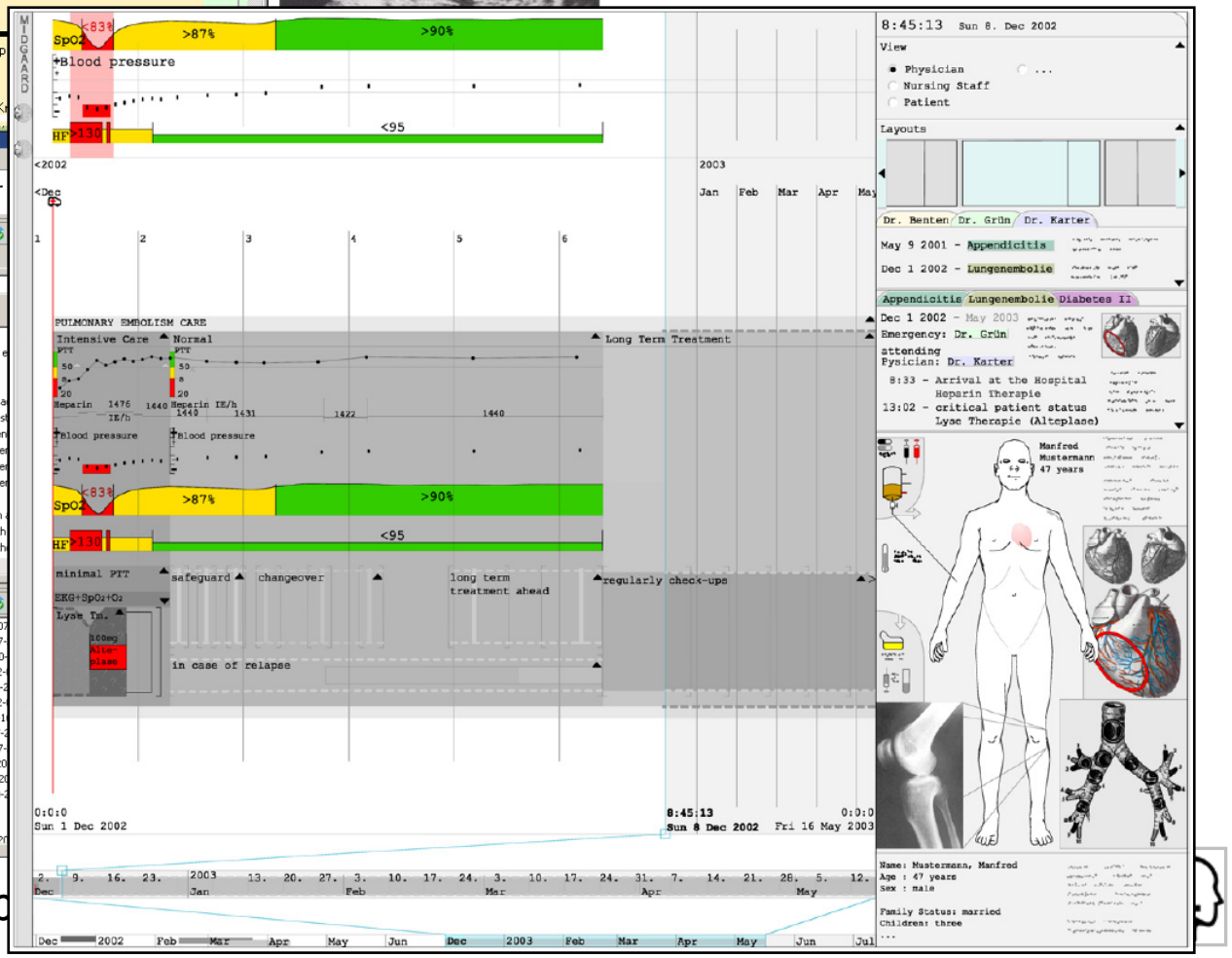
Lifelines and improvements

Overview of categorical and/or numerical data (semantic zoom)

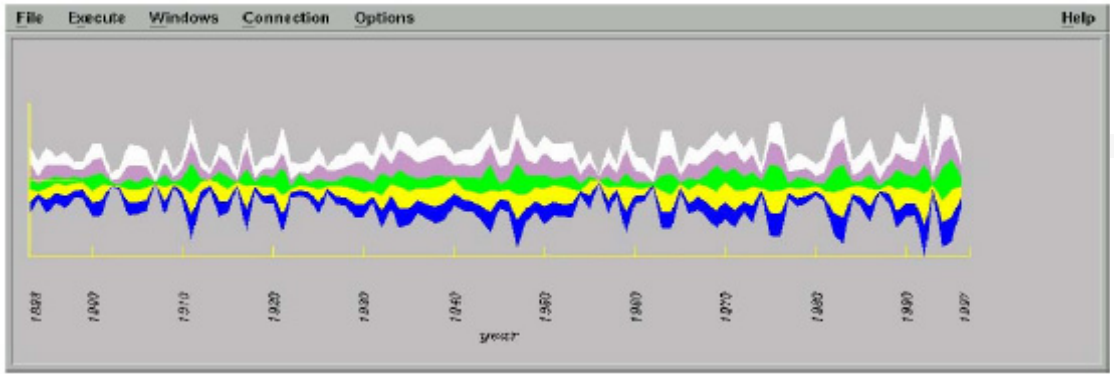


Plaisant CHI 96, AMIA 98

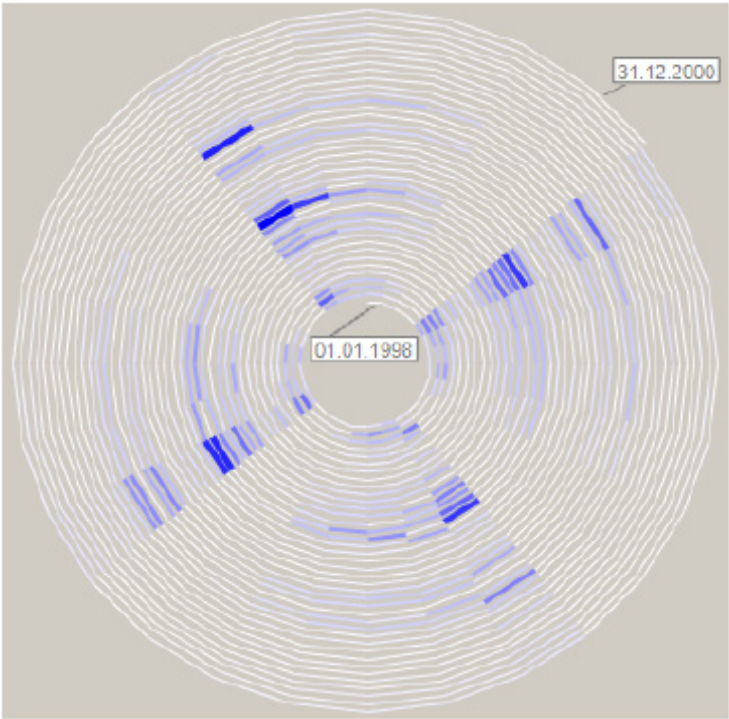
Bade CHI 2004



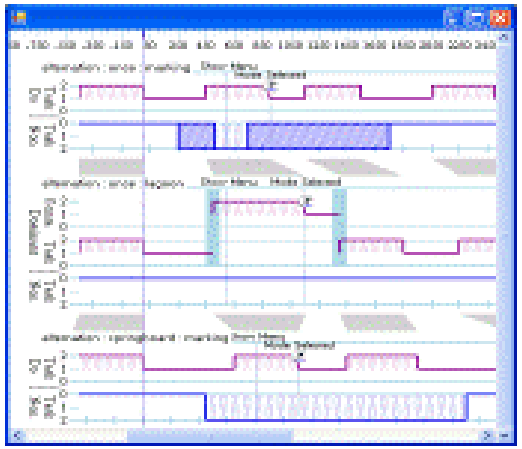
I2b2 (Murp



ThemeRiver (Havre, Infovis00)

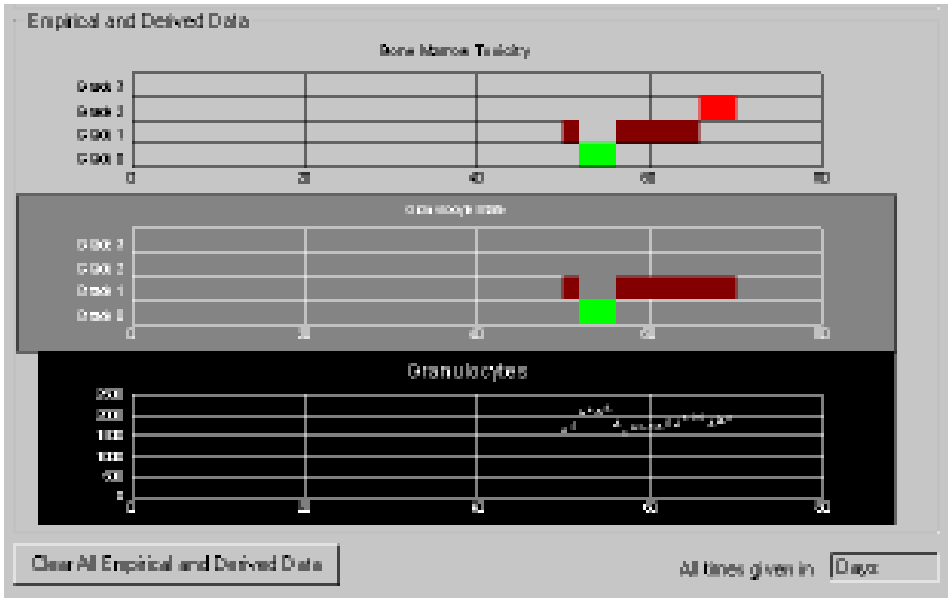


Spiral Graph: Weber 01 (based on Carlis UIST 89)
Periodic data



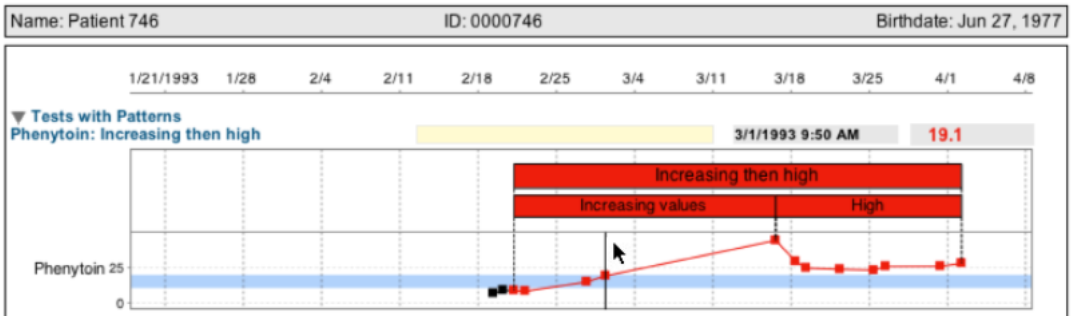
Experiscope (Guimbretiere, CHI07)
One of many example of manual alignment

Specification of temporal abstractions
To reason/query with them



Shahar 1999

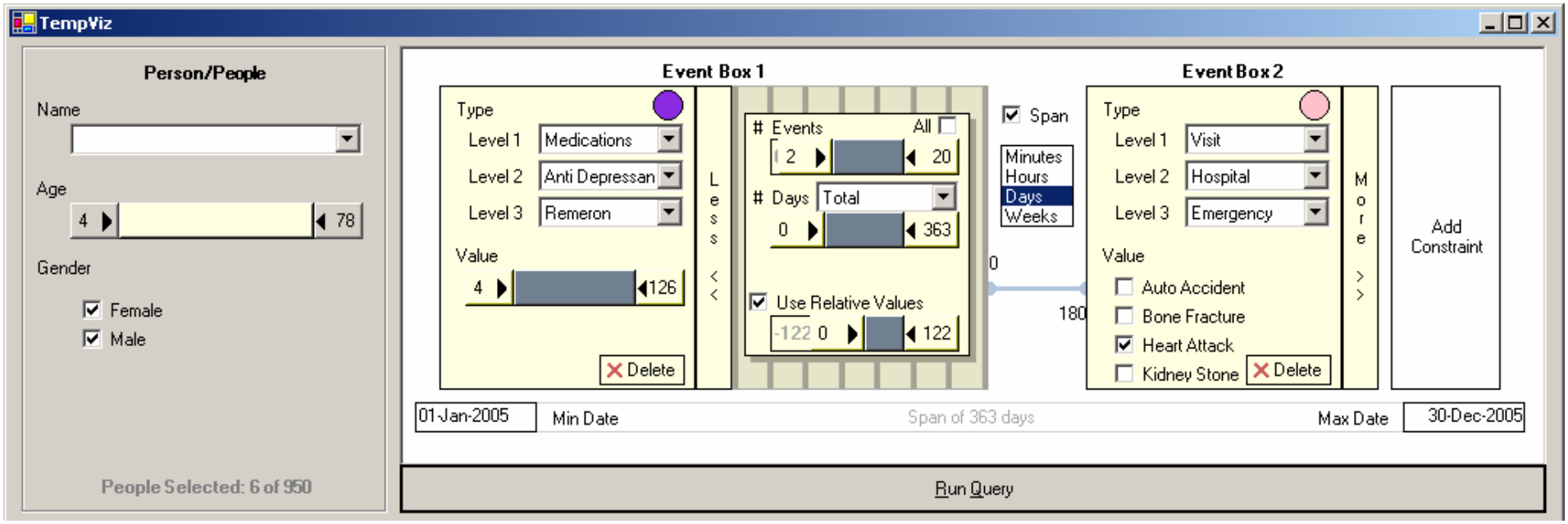
Post 2007



No focus on UI, or presenting results

PatternFinder

Specification of complex temporal queries on categorical data



[Fail et al. VAST06]

Patients with increasing dosages of Remeron followed by a heart attack within 180 days

www.cs.umd.edu/hcil/patternfinder **VAST 06 paper**

PatternFinder

Specification of complex temporal queries on categorical data

TempViz

Person/People

Name

Age 4 78

Gender

Female

Male

People Selected: 6 of 950

Event Box 1

Type

Level 1 Medications

Level 2 Anti Depressan

Level 3 Remeron

Value 4 126

Delete

Less <<

Events All 2 20

Days Total 0 363

Use Relative Values -122 0 122

Span

Minutes

Hours

Days

Weeks

0

180

Event Box 2

Type

Level 1 Visit

Level 2 Hospital

Level 3 Emergency

Value

Auto Accident

Bone Fracture

Heart Attack

Kidney Stone

Delete

More >>

Add Constraint

01-Jan-2005 Min Date Span of 363 days Max Date 30-Dec-2005

Run Query

Visualization Table 6 matches found

Braham, Burton

Elsas, Darnell

Fuchs, Layton

Greif, Gage

Kern, Leola

Williams, Azura

Span: 171 days

Starting Event: 05-May-2005 - Medications.Anti Depressants.Remeron -> 100

Ending Event: 23-Oct-2005 - Medications.Anti Depressants.Remeron -> 120

Show All Events 01-Jan-2005 30-Dec-2005

[Fail et al. VAST06]

Ball and chain display of matches

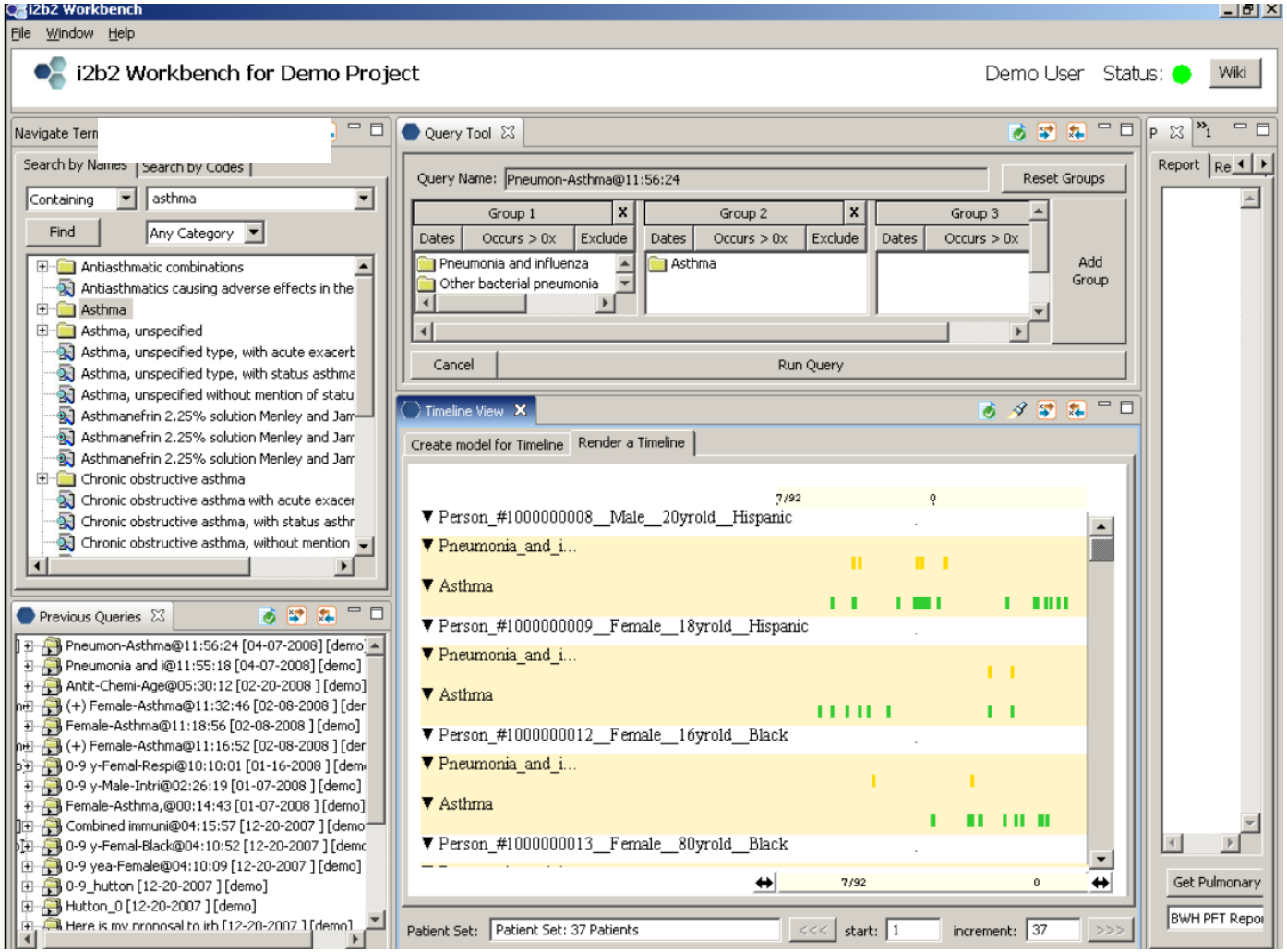
- Related work
- **Quick demo LifeLines2**
- Report on studies
- Ongoing & Future work

Scenario: Study relationship between asthma and pneumonia

1) Run query

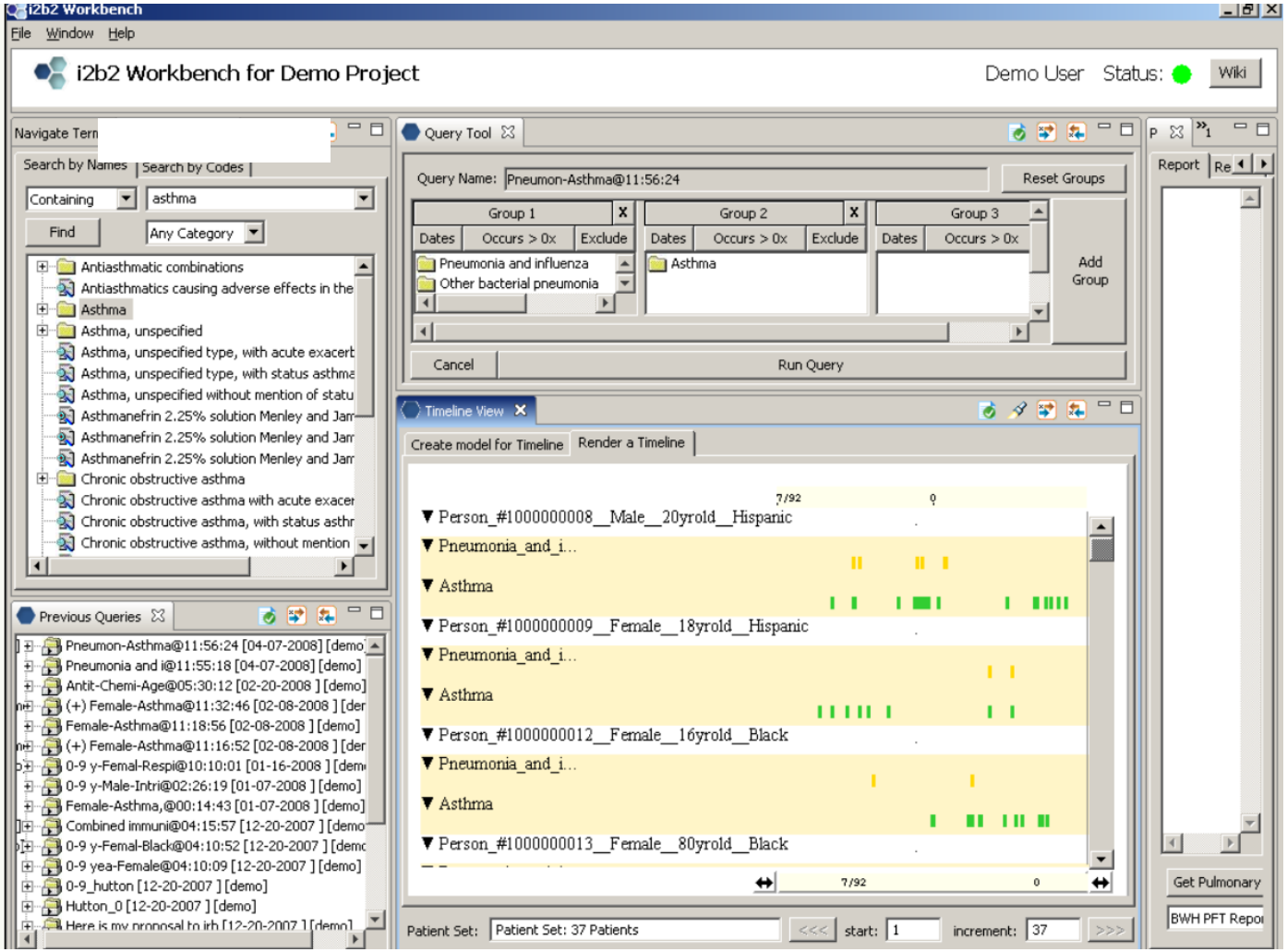
Scenario: Study relationship between asthma and pneumonia

1) Run query **Find all patients who have both Asthma and Pneumonia diagnoses**



Scenario: Study relationship between asthma and pneumonia

1) Run query **Find all patients who have both Asthma and Pneumonia diagnoses**



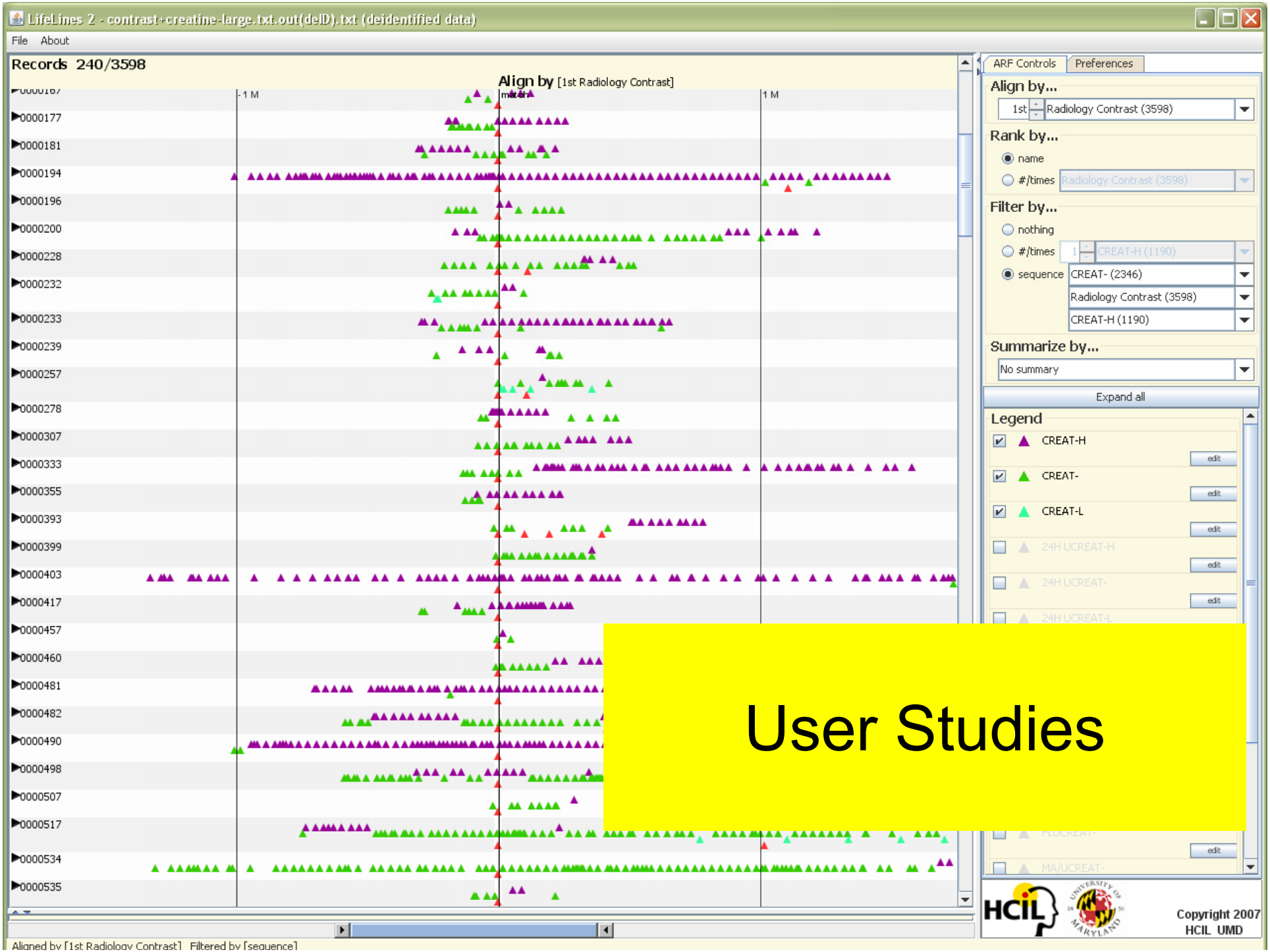
2) **Review results**

- Lifelines2

elements
scroll
open close
align
rank
filter
sequence

interval of validity

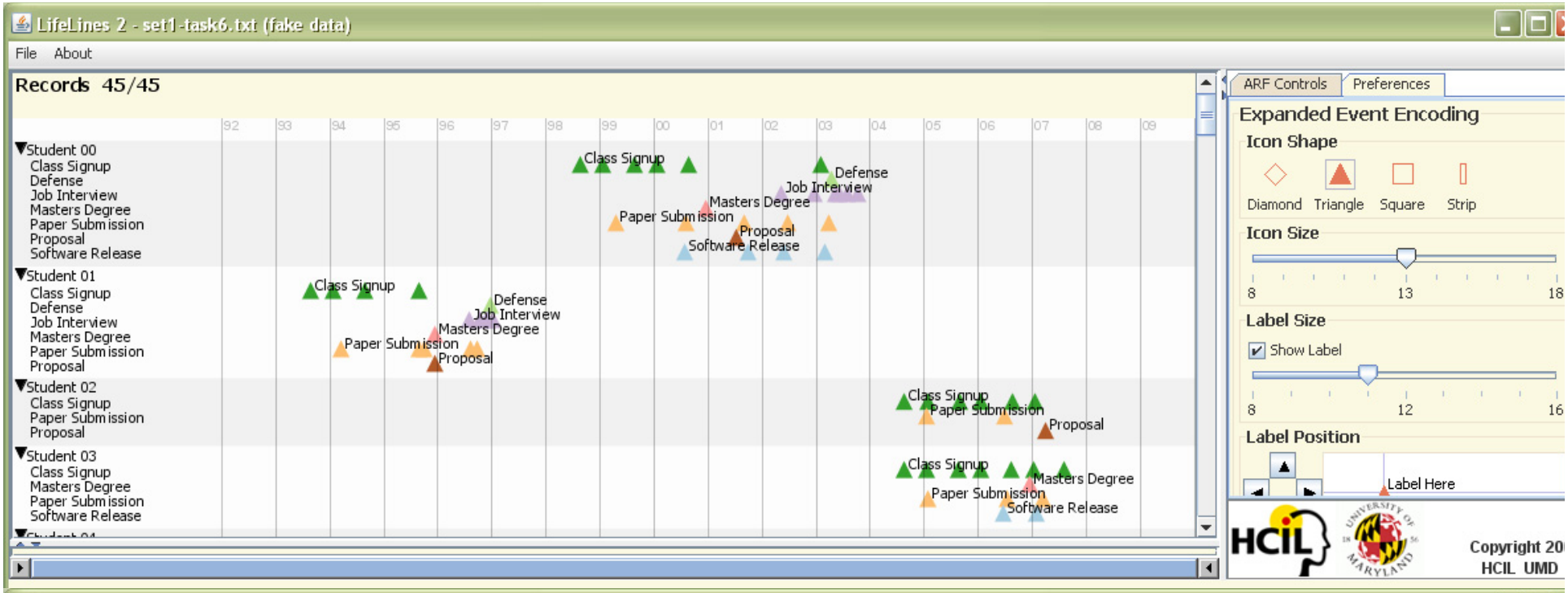
switch to demo

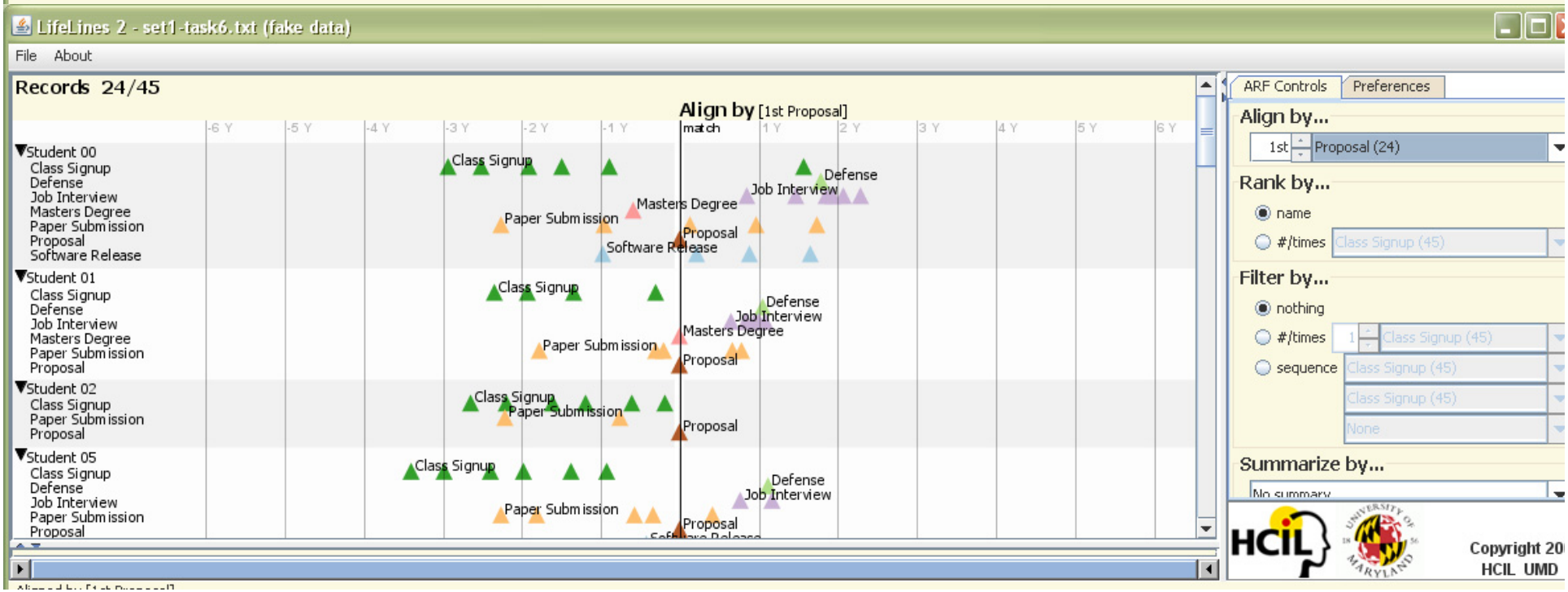


User Studies

Two user studies

- Controlled experiment
  (some training, measure speed and error)
  - Benefit of alignment **YES** (between 0 to 60% improvement)
  - Benefit of interval of validity **NO**
  - 20 participants: grad students
    Data: synthetic student record data
- Domain expert qualitative study
  (no training, think aloud, discussion)
  -
  -
  -
  -

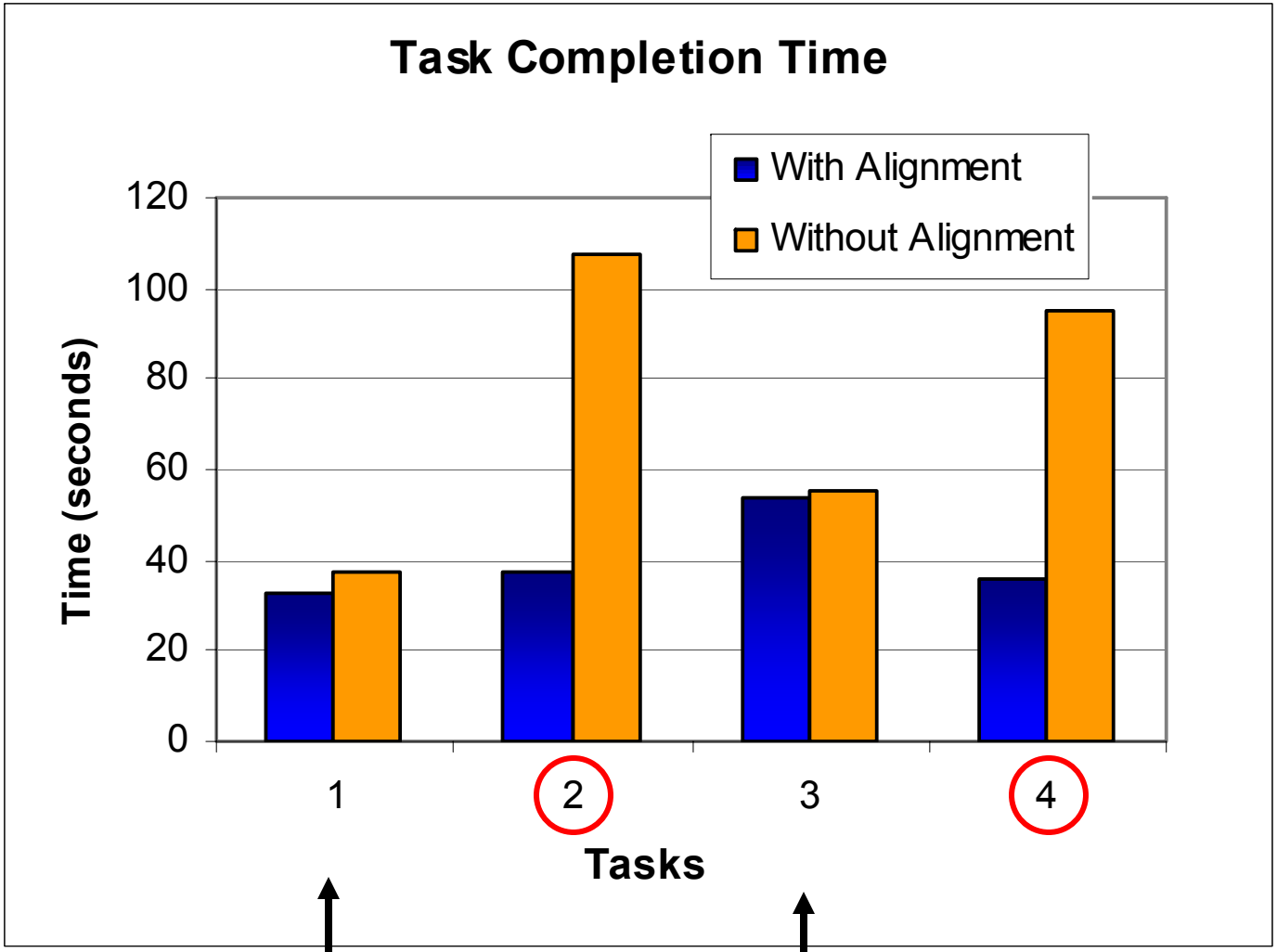

LifeLines 2 - set1-task6.txt (fake data)
File About
Records 45/45
92 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09
Student 00
Class Signup
Defense
Job Interview
Masters Degree
Paper Submission
Proposal
Software Release
Student 01
Student 02
Student 03
ARF Controls
Preferences
Expanded Event Encoding
Icon Shape
Diamond
Triangle
Square
Strip
Icon Size
8
13
18
Label Size
Show Label
12
16
Label Position
Label Here
HCIL
Copyright 20
HCIL UMD


LifeLines 2 - set1-task6.txt (fake data)
File About
Records 24/45
Align by [1st Proposal]
-6 Y -5 Y -4 Y -3 Y -2 Y -1 Y match 1 Y 2 Y 3 Y 4 Y 5 Y 6 Y
Student 00
Class Signup
Defense
Job Interview
Masters Degree
Paper Submission
Proposal
Software Release
Student 01
Student 02
Student 05
ARF Controls
Preferences
Align by...
1st
Proposal (24)
Rank by...
name
#/times
Class Signup (45)
Filter by...
nothing
sequence
None
Summarize by...
No summary
HCIL
Copyright 20
HCIL UMD

Example tasks

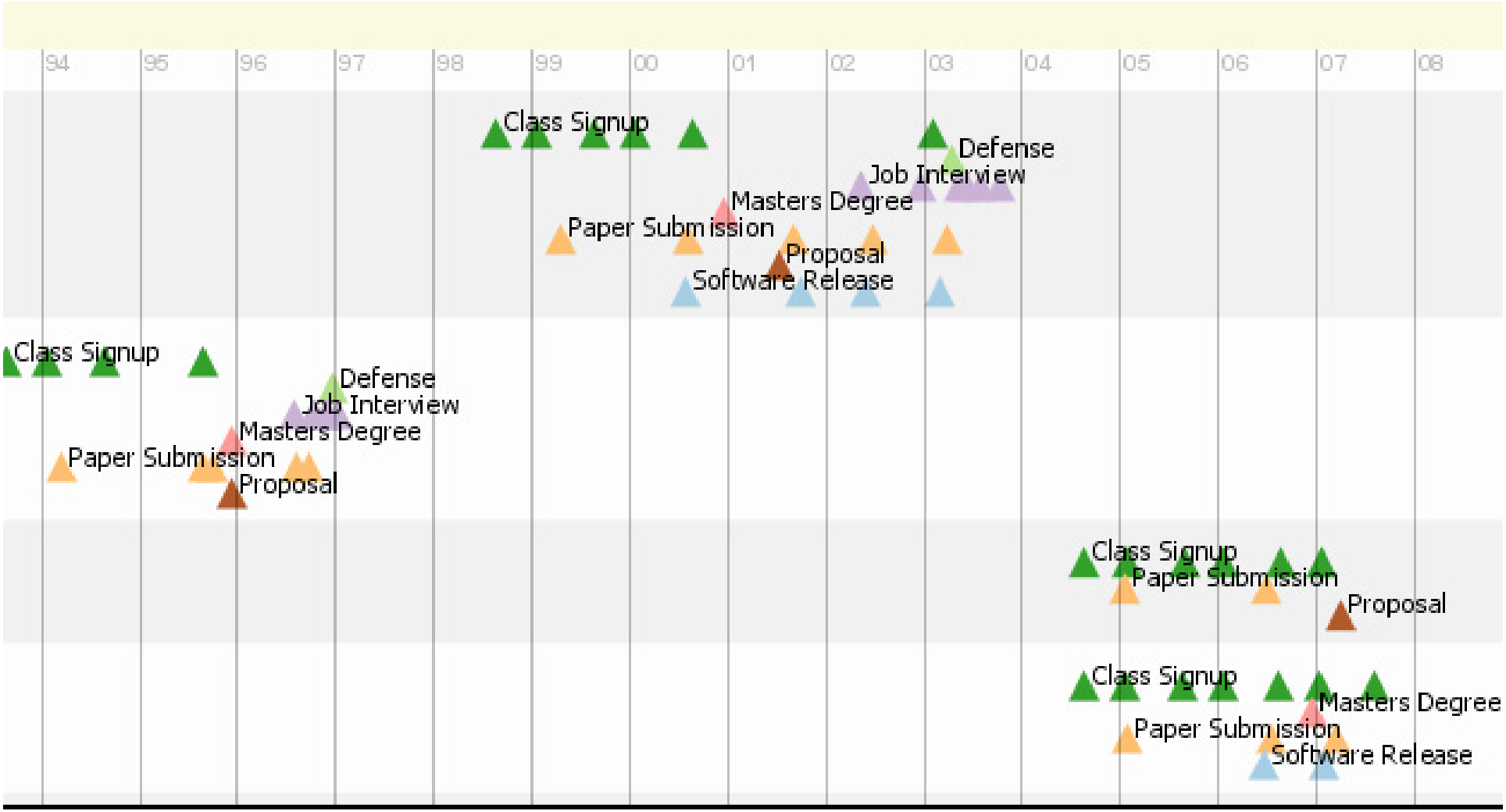
- **1**: How many students submitted a paper within 1 month after proposal? 5 records
- **2**: same with more 20 records
- **3**: How many students published at least 3 papers between PhD proposal and defense?
- **4**: What occurred most often within a month of a student's 1st paper submission?

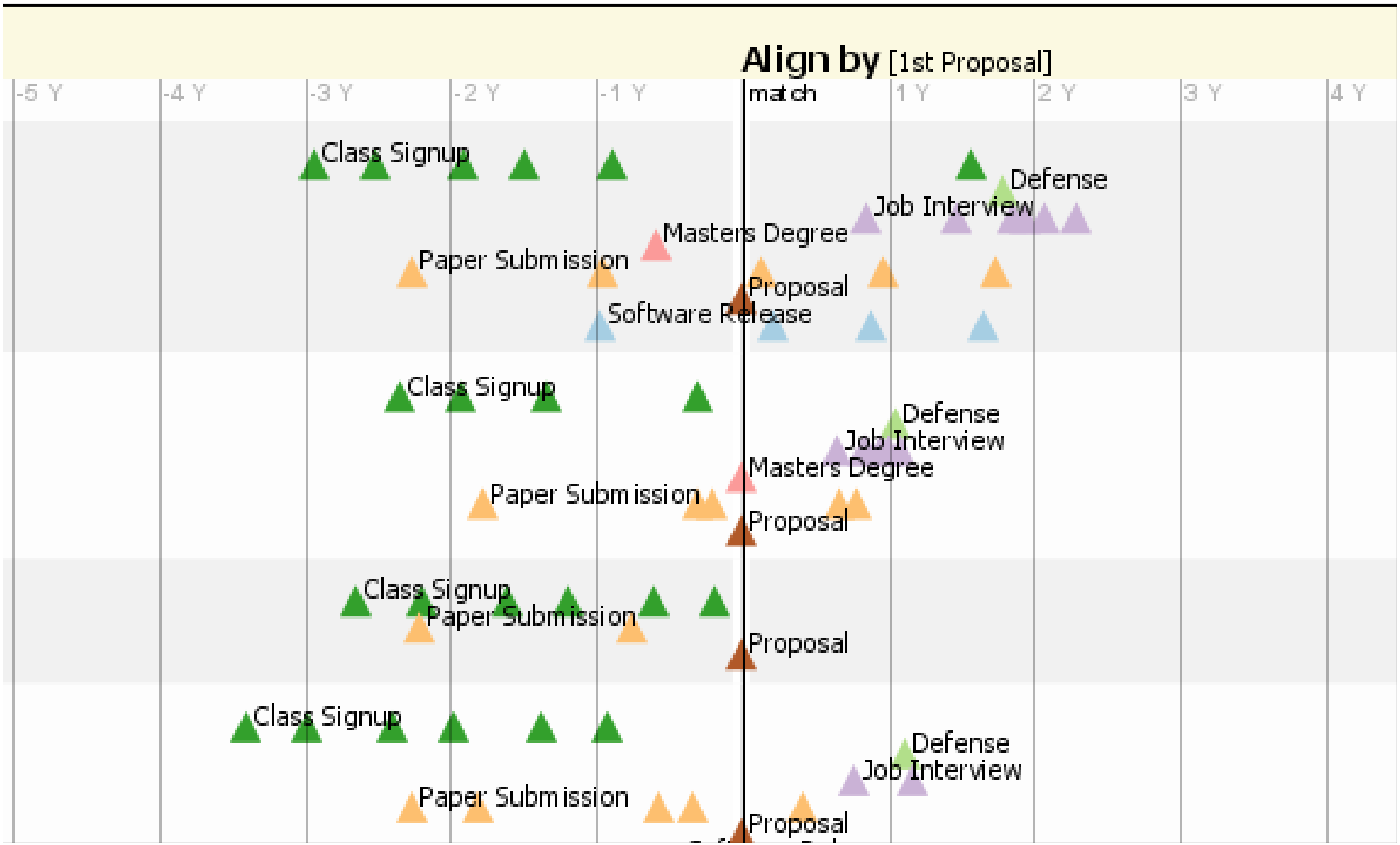
Task completion time



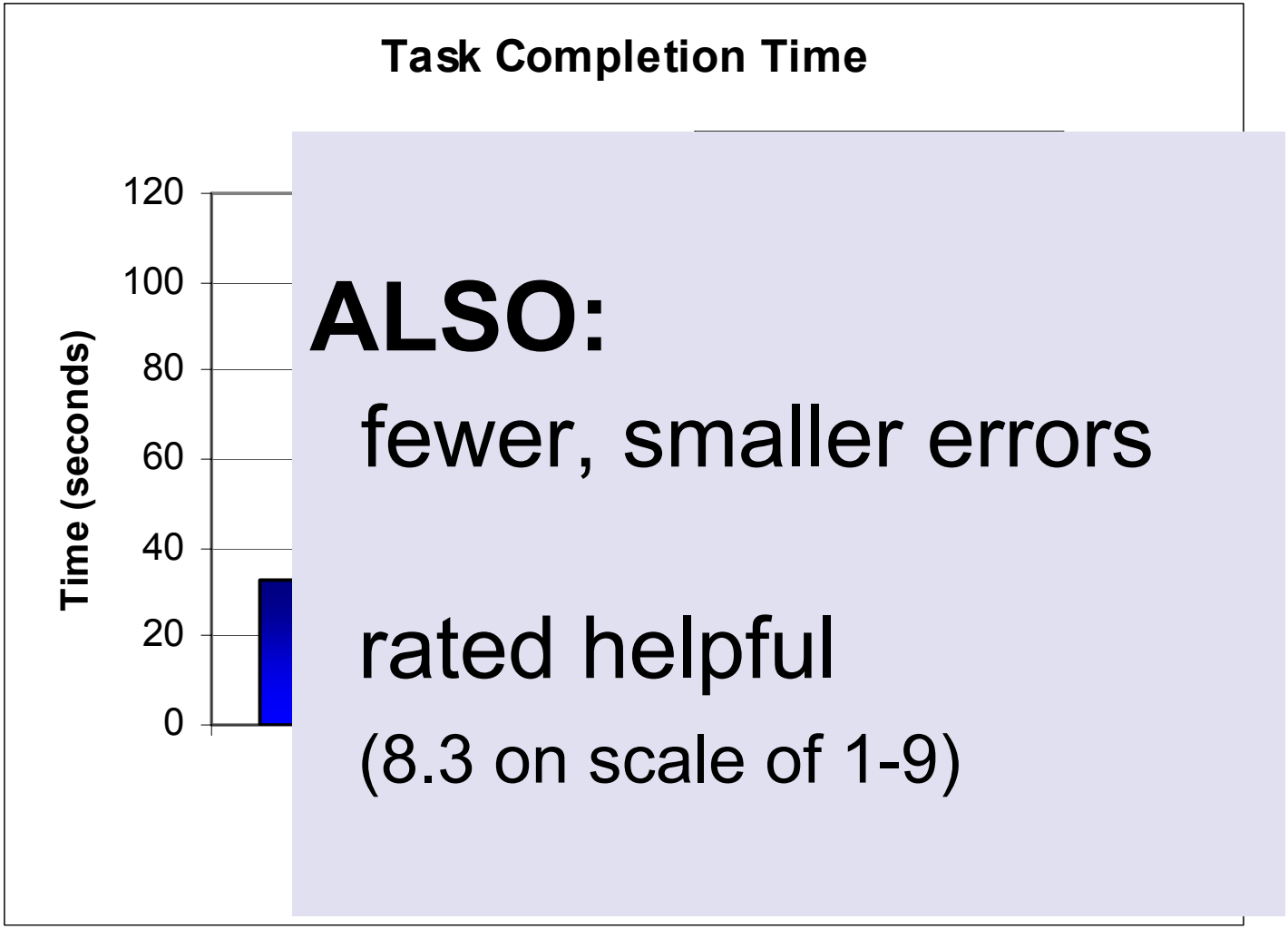
Not really useful if data fit in one screen

or if still learning i.e. best strategy not immediate


94
95
96
97
98
99
00
01
02
03
04
05
06
07
08
Class Signup
Defense
Job Interview
Masters Degree
Paper Submission
Proposal
Software Release
Class Signup
Defense
Job Interview
Masters Degree
Paper Submission
Proposal
Class Signup
Paper Submission
Proposal
Class Signup
Masters Degree
Paper Submission
Software Release


Align by [1st Proposal]
-5 Y
-4 Y
-3 Y
-2 Y
-1 Y
match
1 Y
2 Y
3 Y
4 Y
Class Signup
Defense
Job Interview
Masters Degree
Paper Submission
Proposal
Software Release
Class Signup
Defense
Job Interview
Masters Degree
Paper Submission
Proposal
Class Signup
Paper Submission
Proposal
Class Signup
Defense
Job Interview
Paper Submission
Proposal

Task completion time



ALSO:

fewer, smaller errors

rated helpful

(8.3 on scale of 1-9)

Observations

- In general with alignments users
  - First reduce data with filters
  - Align
  - Eye-ball temporal relationships
  - Zoom/Inspect data with potential

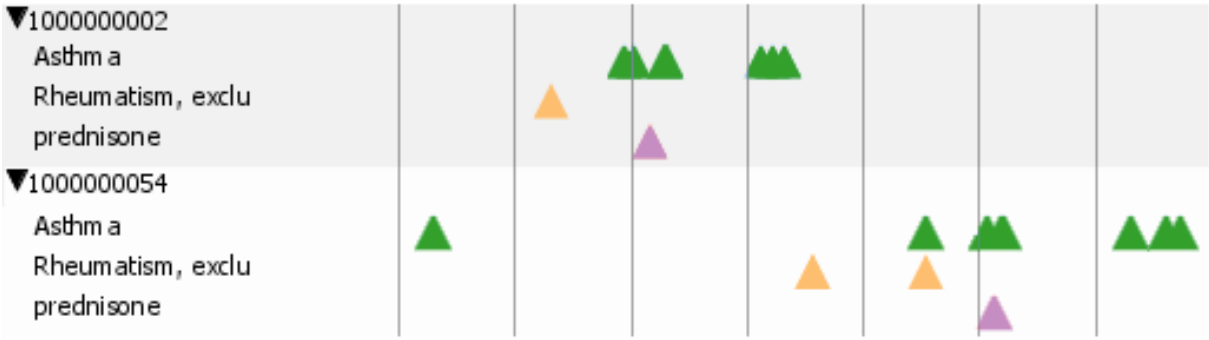
  - Best strategy not found for task 3
    (they aligned on proposal instead of defense)

    When no alignment they filtered by sequence:

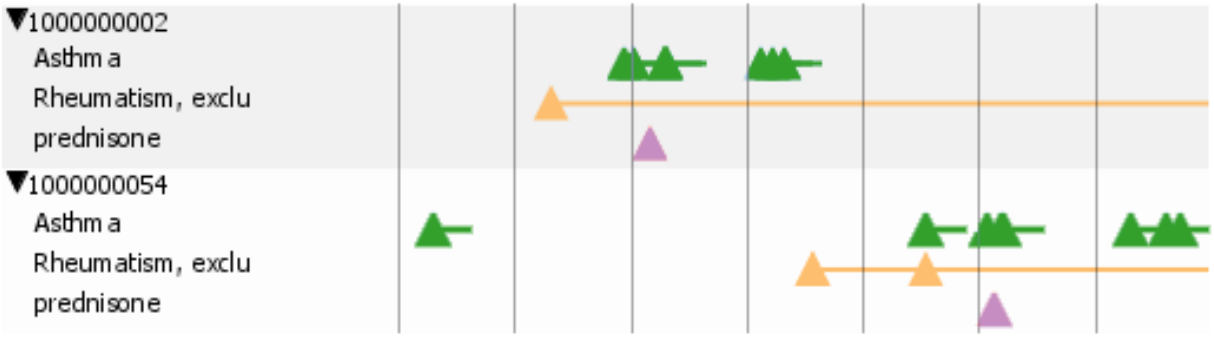
    "proposal < paper < defense"

Intervals of validity: NO Significant differences

(1) Without intervals of validity



(2) With intervals of validity



Two user studies

- Controlled experiment
  (some training, measure speed and error)
  - Benefit of alignment **YES** (between 0 to 60% improvement)
  - Benefit of interval of validity **NO**
  - 20 participants: grad students
    Data: synthetic student record data
    Tasks checked as domain independent
- Domain expert qualitative study
  (no training, think aloud, discussion)
  - Learnability (visual presentation, UI)
  - General feedback and suggestions
  - 4 participants: nurse, physician, 2 prof. of nursing
  - All experienced with EHR and medical research

Domain expert qualitative study

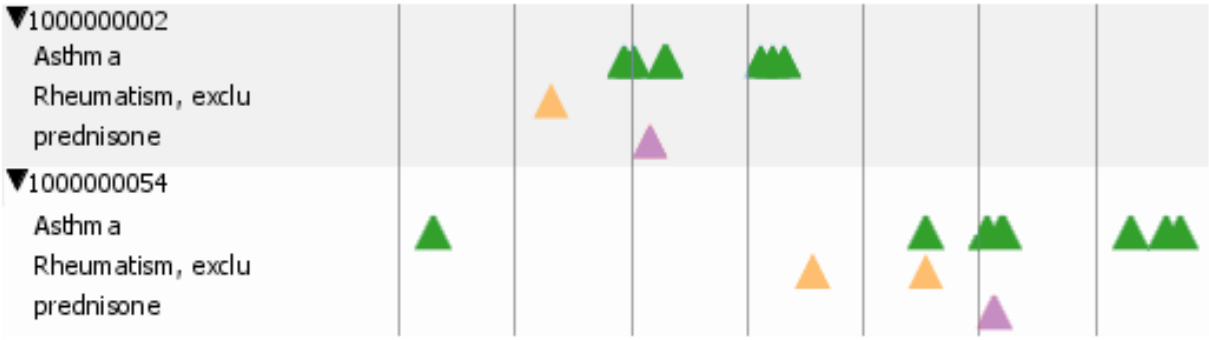
- 3 scenarios
  - Picking subjects for a clinical trials
    - Patients who took steroids for asthma
  - Trends
    - Asthma attacks seem more likely to follow or precede pneumonia?
    - What events seem to co-occur with acute myocardial infraction?
- Zero training
- Review data
- Think aloud

Domain expert qualitative study

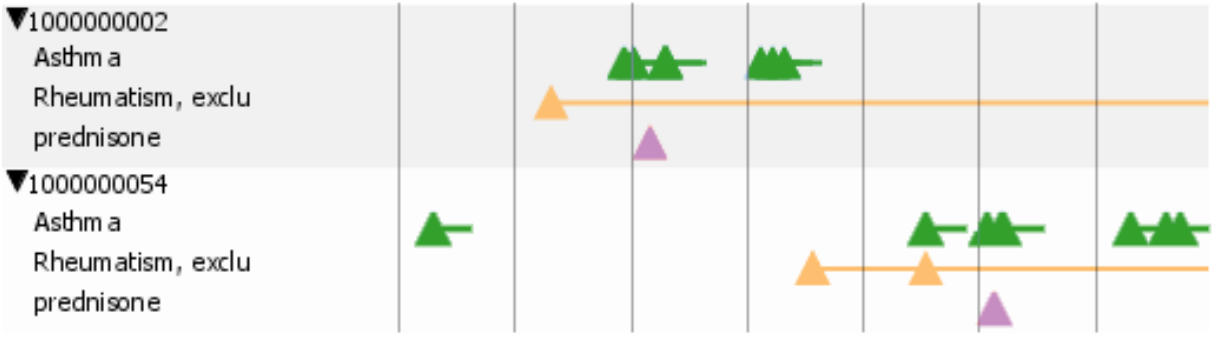
- Visual representation and Align-rank-filter understood without training
  - One user had initial problem seeing that 1 facet = 1 patient (but figured out on his own)
  - One user learned every single feature entirely on her own, others asked demonstration of some features.
- But even experts get confused about data available (e.g. asthma interpreted a flare)

Interpreting intervals of validity

(1) Without intervals of validity



(2) With intervals of validity



Uncertainty?
Or certain duration?

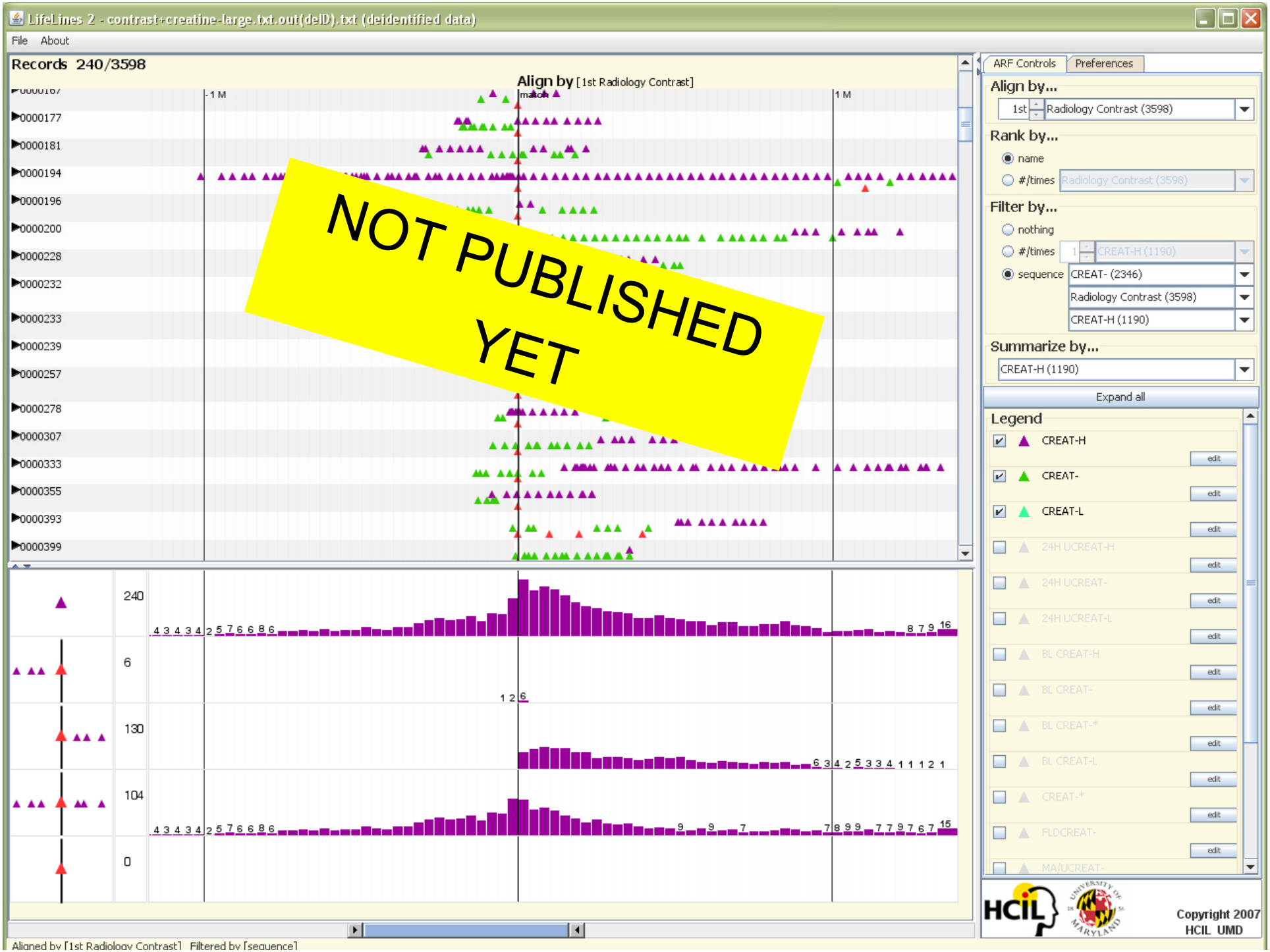
fading

May only add more confusion?
Is it realistic to count on adequate training?

Future work

- Summaries:
  - temporal distribution of event type(s)

- Comparison of populations grouped by
  - patient characteristic (e.g. men/women)
  - presence/absence (e.g. had stroke or not)
  - ordering relative to sentinel


LifeLines 2 - contrast+creatine-large.txt.out(deID).txt (deidentified data)
File
About
Records 240/3598
Align by [1st Radiology Contrast]
-1 M
match
1 M
NOT PUBLISHED YET
ARF Controls
Preferences
Align by...
1st
Radiology Contrast (3598)
Rank by...
name
#/times
Radiology Contrast (3598)
Filter by...
nothing
#/times
CREAT-H (1190)
sequence
CREAT- (2346)
Radiology Contrast (3598)
CREAT-H (1190)
Summarize by...
CREAT-H (1190)
Expand all
Legend
CREAT-H
CREAT-
CREAT-L
24H UCREAT-H
24H UCREAT-
24H UCREAT-L
BL CREAT-H
BL CREAT-
BL CREAT-*
BL CREAT-L
CREAT-*
FLDCREAT-
MA/UCREAT-
edit
HCIL
Copyright 2007
HCIL UMD
Aligned by [1st Radiology Contrast] Filtered by [sequence]

Future work

- Summaries:
  - temporal distribution of event type(s)

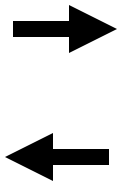
- Comparison of populations grouped by
  - patient characteristic (e.g. men/women)
  - presence/absence (e.g. had stroke or not)
  - ordering relative to sentinel

- Integration in operational systems

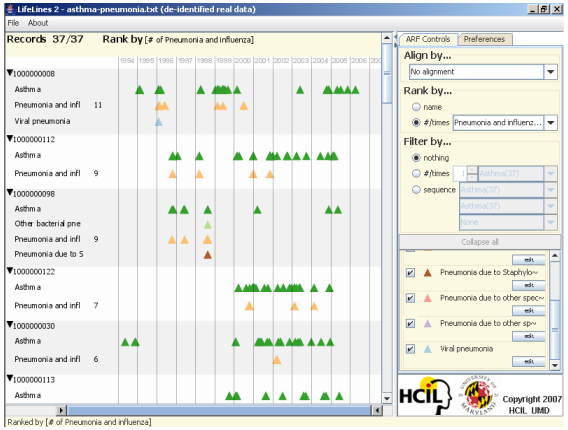
end

Context

Search
millions of records



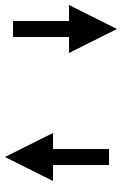
Interactive
visualization of results



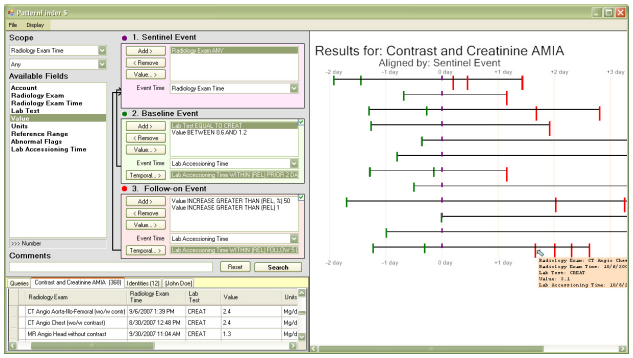
LifeLines2

Context

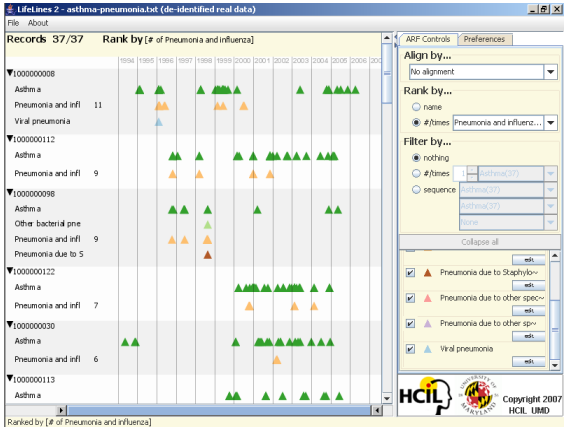
Search
millions of records



Interactive
visualization of results



PatternFinder
in Amalga



LifeLines2

PatternFinder (in Azyxxi/Amalga)

Complex temporal queries:

- Patients discharged from ER admitted again within a week
- **Patients with radiology exam using contrast, and a normal creatinine level at most 2 days before the exam, and creatinine increase of at least 50% and more than 0.1mg within 5 days of the exam**

Combine alignment with PatternFinder

PatternFinder S

File Display

Scope

Radiology Exam Time

Any

Available Fields

Account
Radiology Exam
Radiology Exam Time
Lab Test
Value
Units
Reference Range
Abnormal Flags
Lab Accessioning Time

>>> Number

Comments

1. Sentinel Event

Add > | Radiology Exam ANY
< Remove
Value... >
Event Time | Radiology Exam Time

2. Baseline Event

Add > | Lab Test EQUAL TO CREAT
Value BETWEEN 0.6 AND 1.2
< Remove
Value... >
Event Time | Lab Accessioning Time
Temporal... > | Lab Accessioning Time WITHIN (REL) PRIOR 2 DA

3. Follow-on Event

Add > | Value INCREASE GREATER THAN (REL, %) 50
Value INCREASE GREATER THAN (REL) 1
< Remove
Value... >
Event Time | Lab Accessioning Time
Temporal... > | Lab Accessioning Time WITHIN (REL) FOLLOW 5 D

Reset | Search

Queries | Contrast and Creatinine AMIA (368) | Identities (12) | [John Doe]

Radiology Exam	Radiology Exam Time	Lab Test	Value	Units
CT Angio Aorta-Illo-Femoral (wo/w contr)	9/6/2007 1:39 PM	CREAT	2.4	Mg/d
CT Angio Chest (wo/w contrast)	8/30/2007 12:48 PM	CREAT	2.4	Mg/d
MR Angio Head without contrast	9/30/2007 11:04 AM	CREAT	1.3	Mg/d

Results for: Contrast and Creatinine AMIA

Aligned by: Sentinel Event

-2 day | -1 day | 0 day | +1 day | +2 day | +3 day

Radiology Exam: CT Angio Che
Radiology Exam Time: 10/6/200
Lab Test: CREAT
Value: 3.1
Lab Accessioning Time: 10/8/2

Paper submitted (ask me)

- radiology exam with contrast
- normal creatinine level at most 2 days before the exam
- creatinine increase of at least 50% and more than .1mg within 5 days of the exam)

HCIL

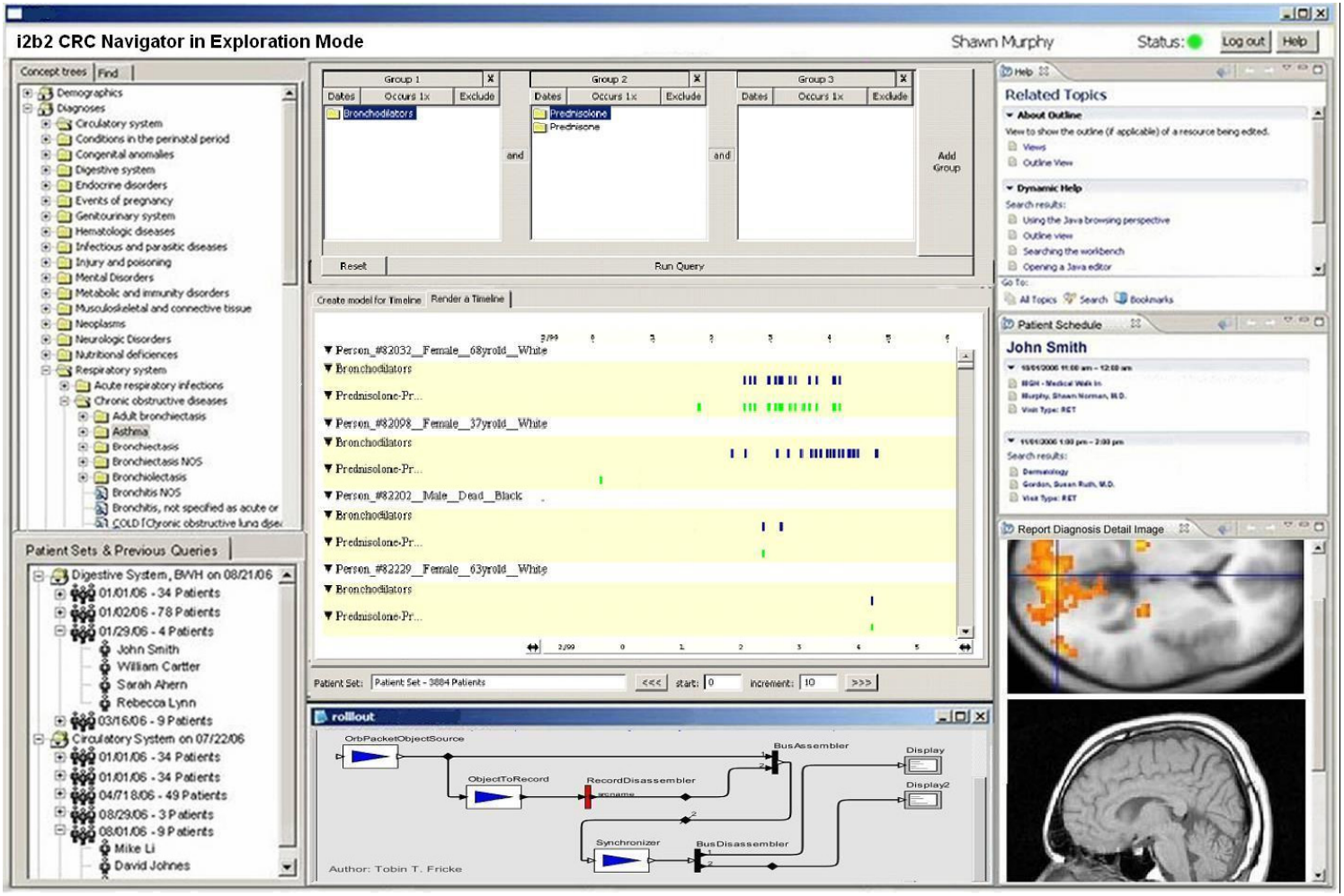
Combine alignment with PatternFinder

Collaboration with
Washington Hospital Center
(original developers of Azyxxi, now Microsoft Amalga)

- Alignment integrated
- Put on desk of first users last week

- Connected to real time database
- Search UI was designed to match style of existing search interface

Integrate Align-Rank-Filter in i2b2



by early summer

In summary…

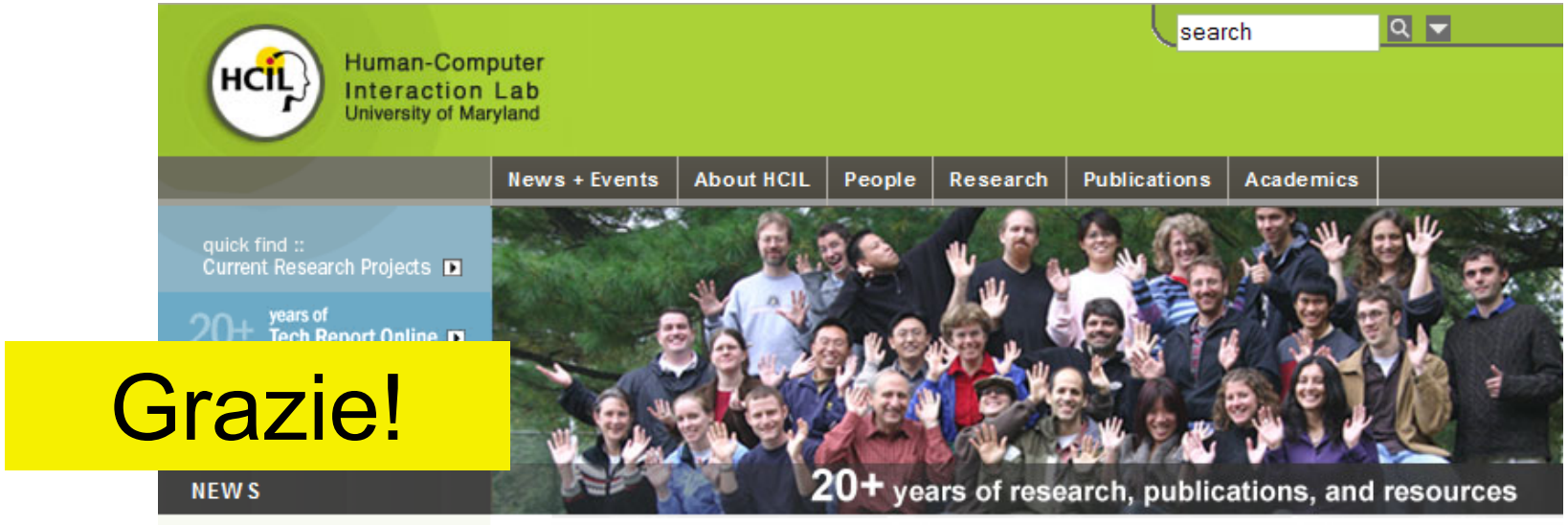
- **Align Rank and Filter**
  Powerful combination of simple operations to explore temporal categorical data

- **Performance benefit of alignment** significant

- Integration in 2 large operational EHR systems

- Many applicable domains:
  - Highway incident log
  - Student records
  - Web logs
  - Vehicle fleet records

- Interval of validity? Not clear



Grazie!

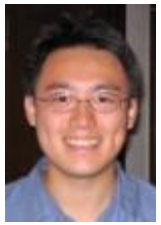
www.cs.umd.edu/hcil/lifelines2

plaisant@cs.umd.edu

All medical data shown had been de-identified

(drawback: it was small)

Still a few spots at our workshop:
"Interactive Visual Exploration of Electronic Health Records"
at HCIL Symposium (May 30) www.cs.umd.edu/hcil/soh



Special thanks to:
David Wang - who could not come ☹
Washington Hospital Center (and others) for financial support