personal informatics
collecting, integrating, and understanding personal data

anthony tang
university of calgary

roadmap

definition

part 1: “class stuff”

examples of personal informatics in everyday life

some example domains

properties of data types

basic tasks in personal informatics
roadmap

part 2: “research stuff”

a model for personal informatics

some inspiration

bigger issues
proclubboy Tony
So the one night I can't eat anything (blood test tomorrow morning), and everyone tells me about how good their snack food is.
1 hour ago

proclubboy Tony
Watching the #OWS stuff with interest. Will the protest be a means to an end, or is it an end in of itself?
15 hours ago

proclubboy Tony
Linesman with the best hit of the night. Clocked Letang when he wasn't looking
8 Oct

proclubboy Tony
After a promising first, #flames have given up 4 straight goals. Ouch
8 Oct

proclubboy Tony
Opening ceremonies with Phil Collins! yfrog.com/mep6oswtj
8 Oct

proclubboy Tony
Heading out to the #flames game, but the goal is to see the #penguins!
8 Oct

proclubboy Tony
University one month in: not sure if kids are planking, or napping in extremely awkward places/positions.
7 Oct
definition

*Personal Informatics* is a class of applications that help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge.

definition

*personal informatics* is a class of applications that help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge.
why personal informatics?

definition

why personal informatics?
why personal informatics?

understanding oneself / one’s behaviours, activities

guesses vs. reality

behaviour change

just a nerd

example domains

fitness
example domains

fitness

example domains

wellness / dieting

Food Log

<table>
<thead>
<tr>
<th>Breakfast</th>
<th>Fat</th>
<th>Kcal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quaker, Quaker Oat Bran Cereal</td>
<td>65</td>
<td>194.5</td>
</tr>
<tr>
<td>2 x 1.25 cup</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Lunch                         |     |      |
| Wild rice, cooked             | 0.8 | 165.6|
| 1 x 1 cup                     |     |      |
| Chicken, stewing, meat only, raw| 41.2| 287.1|
| 1 x 1 unit (yield from 1 lb chicken) | | |

| Dinner                        |     |      |
| Sandwich spread, pork, beef   | 2.6 | 35.3 |
| 1 x 1 tbsp                    |     |      |

Average daily calorie intake: 913.23
example domains

location tracking

example domains

“work”: email

http://flowingdata.com/2008/03/19/21-ways-to-visualize-and-explore-your-email-inbox/
example domains

“work”: email

http://flowingdata.com/2008/03/19/21-ways-to-visualize-and-explore-your-email-inbox/
example domains

“work”: where does my time go?

http://www.slifelabs.com/

example domains

asymmetry in instant messaging
properties and data types

concepts and issues

data collection
  » how is data collected
  » granularity of data

data types
  » traditional data & issues
  » non-traditional data

properties and data types: concepts & issues

data collection: how it’s collected

manual
  » *do it yourself, son!*

automated
  » *I am computer, I will collect it for you...*

semi-automated
  » *CPU collects, person labels*
properties and data types: concepts & issues

data collection: how it’s collected

manual
  » in-situ: collecting it as it happens
  » post-hoc: collecting it after the fact

advantages:

disadvantages:

- you can modify things on the go
- you forget to collect the data
data collection: how it’s collected

automated

advantages:

- lots of data
- all the parameters are captured for you
- don’t have to remember to collect

disadvantages:

- without adequate context, difficult to know what is collected
- difficult to “annotate”
properties and data types: concepts & issues

data collection: how it’s collected

**semi-automated**

advantages:

- tagging allows for people to annotate data (i.e. “humanize” it)

disadvantages:

- may still be difficult to modify what’s being collected
- can be time-consuming to tag/label data
properties and data types: concepts & issues

semi-automated collection example
properties and data types: concepts & issues

granularity of data

fine-grained: nuanced, detailed
coarse-grained: higher level, less detail

http://babygreendesign.blogspot.com
http://theenergycollective.com

traditional data types

nominal data
» categories without inherent ordering (e.g. “activities in a day”)

ordinal data
» “ordered” categories (e.g. attitude: “feeling tired”, “feeling energetic”)

scalar data
» data where the “scale” metric works, and comparisons between values are appropriate (e.g. body temperature)
what kind of data is being collected / presented here?

what type of data is being collected / presented here?
properties and data types: concepts & issues

traditional data: visualization implications

ordering
» this matters (or has implications) in ordinal/scalar data types

size
» this matters (or has implications) in scalar data types

“middle”
» “average” exists for scalar types, but not for ordinal or nominal types
» “median” works for ordinal data
» “median” works for nominal data

properties and data types: concepts & issues

non-traditional data types

location
» semantic location
» geo-location
» spatial implications for visualization

pictures
» rich data, but what does it all mean?

time
» coarseness
» subjective vs. objective time
» cyclic vs. linear
why personal informatics visualizations?

basic tasks

exploring the data

comparing the data with something / someone else

identifying / looking for patterns
roadmap

part 1

examples of personal informatics in everyday life
definition
some more example domains
properties of data types
basic tasks in personal informatics

roadmap

part 2: “research stuff”

a model for personal informatics
some inspiration
bigger issues
a model for personal informatics

research question:

how does personal informatics actually work in practice?
what are people trying to do?
what problems do people encounter?


a model for personal informatics

survey research

quantified self people: hardcore enthusiasts

« 28-hour day guy

jeopardy guy »
five-stage iterative model

Preparation

The stage before people start collecting information.
- What information to record
- How to record the information
five-stage iterative model

Preparation Barriers
- Choosing the right information to collect
- Finding the right tool to use

Collection
The stage when people collect information about themselves (e.g., inner thoughts, behavior, social interactions, and their immediate environment).
Collection Barriers

- Using the tool
- Remembering
- Lack of time
- Motivation
- Finding data
- Accuracy

Integration

The stage when the information from the Collection stage is prepared, combined, and transformed for the user to reflect on.
a model for personal informatics

five-stage iterative model

Integration Barriers

- Organization
- Scattered visualizations
- Transcribing data
- Multiple inputs

Reflection

The stage when people reflect on their personal information.

- Users may reflect immediately (short-term)
- Or after several days or weeks (long-term)
a model for personal informatics

five-stage iterative model

Reflection Barriers
- Lack of time
- Self-criticism
- Visualization
- Interpretation
- Sparse data
- No context

Action
The stage when people choose what they are going to do with their new-found understanding of themselves.
a model for personal informatics

some lessons

**barriers cascade**: problems in earlier stages affect later stages

**stages are iterative**: people discover that they want to integrate/incorporate new types of data as they go

integration is a beast

some inspiration

nick felton’s stuff is super cool

http://feltron.com/
THIRTY-EIGHT THOUSAND FIVE

An exhaustive compendium of travel and activity in 2008, including:
- 366 days of walking
- 254 subway trips
- 107 taxis
- 12 flights
- 19 buses
- 1,333 miles of car travel
- 84 miles of running
- 84 miles of cycling
- 200 miles of driving
- 40 trips with other drivers
- 61 miles in the gym
- 1,500 miles swimming
- 2,000 miles biking
- 65 hours of skier

HUNDRED AND TWENTY-FOUR

FOUR

MICHAEH CLAYTON

SEVEN

14

11

ONE

54.7

T

96'

20

31

SEVEN

31

30

DINING

S reset menu and $100 meals at 10 restaurants, 13 homes and 4 restaurants.

140

SEVEN

573

ASSORTED

SEVEN

573

SIERRA NEVADA ESB

ASSORTED

STELLA ARTOIS

ONE

ONE

FIG, MINT & TEQUILA

MIRACLE FRUIT

AMERICAN

ONE

FOG, FROG LEGS & GOAT
some inspiration

IMPACT

Problem:
- incorporate physical activity into daily life
- step counts are not enough: low insight value

Approach:
- provide insight into patterns
- use semi-automated data collection
- synchronize data collection

mylifebits project

MylifeBits is a project (led by MSR’s Gordon Bell) to have a lifetime store of everything:

- full-text search, text & audio annotations, and hyperlinks
- articles, books, cards, CDs, letters, memos, papers, photos, pictures, presentations, home movies, videotaped lectures, and voice recordings
- phone calls, IM transcripts, television, radio

some inspiration
related to the mylifebits project

vicon revue née microsoft sensecam

- 3MP photos (every few seconds)
- temperature sensor
- light colour / intensity sensor
- infrared motion detection
- accelerometer
- 3-axis compass
- ...

mylifebits project

The Microsoft Research Cambridge SenseCam viewer application allows playback and review of Sensecam images and associated sensor readings.

http://research.microsoft.com/
vicon revue data visualized... (one way)

integrating time, location, image data...


more mylifebits

more mylifebits


bigger issues

behaviour monitoring and change

personalization

recognition: speakers, speech, people, activity

privacy / security / ownership
roadmap

part 2: “research stuff”

a model for personal informatics
some inspiration
bigger issues

what to take away from today...

we have mountains of data

a lot of it is personal data

**personal informatics is coming:** people want to understand and make use of personal data