Asking the Internet Questions

Michael Layzell

Supervisors

Rick Salay
Marsha Chechik
People are signing up with fake/invalid Names and Emails!

FIX IT!

Manager:

People are signing up with fake/invalid Names and Emails!

FIX IT!
Requirements

1. Name has both First & Last Name
2. Email is correctly formatted
3. Email belongs to the given name
“Unbounded”

Requirements cannot be fulfilled by the standard Algorithm-Database approach.
Unbounded?

Needs human judgement, understanding or creativity
  e.g., Ensure the user is real, provide a good model layout., etc.

Needs access to many/changing sources
  e.g., Getting ratings from the public.

Has many examples, but can’t be easily generalized
  e.g., Provide grammar checking capability.

Inherently uncertain
  e.g., Provide a function to determine whether a storm is coming.

Data required is incomplete or hard to obtain
  e.g. Determine how many people in Toronto are regular theatre goers.
Ask the INTERNET!
Demo
What was that?

1. Two Sources
2. Google/Bing "$NAME AND $EMAIL"
3. Sources were Combined
Simple... but Unreliable
More Sources?

Google/Bing "$NAME AND $EMAIL"
  How many results?

Check Github/Facebook etc...
  Is there an account with the name and email?

Send an email
  Do they click the confirmation link?
More Reliable... but more Complex
Contributional Implementation
A Programming Paradigm for improving the satisfaction of Unbounded requirements
Inputs

Budget
Arguments

Contributional Implementation

Magic!
10s - $0
Michael Layzell
michael@thelayzells.com

Contributional Implementation

Magic!
(Not really)
Opinions

All Sources Within Budget

10s - $0
Michael Layzell
michael@thelayzells.com

Contributional Implementation

Google
False
0.4

Github
True
0.8

Email

Unk
-

Voting

Result
True
0.667

Too Long
Source Selection Strategy

Inputs
- Budget
- Arguments

Opinions
- Value
- Trust

Source
- Value
- Trust

Contributional Implementation

Aggregation Strategy

Result
- Value Quality
What am I doing?

Generic Source Selection & Aggregation Strategies
   Dempster-Schafer Theory

Designing CI API
   Transparency, Simplicity

Improving CI Algorithm
   Parallelism, Expanded Budget

Implementing all of the above
Demo
Questions?