The Stone Soup of Data

Jamie Taylor - Metaweb

WWW2007
eschew mentalistic explanations

you tend not to get caught up in them when working with preverbal animals!
Let the data speak for itself!

This is a very useful
Description
of a community
2 users making their world better!
Very different than what you hear people saying when describing on-line communities.
Story of Stone Soup is told here.
What is the story about?

Multiplying a scarce resource through community

This is an economic tale
community was altered
the rules of exchange were manipulated
Changed the Economy
community = economy
--Everyone better off--
So far...

• Jamie is an economist in disguise
• Stone Soup is an economic tale
• Community is another word for economy
To understand a community....
...follow the flow of value.
Social computing

User generated content, data, metadata

Harnessing the collective intelligence...

Web 2.0 is about community
Not out of Date!

Technologies aren’t hip, but message is about community.

Perfectly accurate today!

Read it again!
Dan Bricklin: Cornucopia of the Commons

- Three ways to build a database
  - Organize manually
  - Organize mechanically
  - Get volunteers to build it

Interesting examples
- CDDB....
  - Take community output and make it private.
THE WEALTH OF NATIONS

ADAM SMITH
"I have never known much good done by those who affected to trade for the public good."

-Adam Smith, 1776
It only took me half a decade to find people to do this with!
Architecture of Participation

Open Source communities are “free markets for ideas”

-Tim O’Reilly
Architecture is Politics

-Mitch Kapor
One of Stuart’s many interesting works showing how abstract models can shed light on basic phenomena.

Taylor/Jordan 2000 work: Multiple NK Models, operating at different scales, describes why Linux didn’t suffer the Mythical Man Month.

One of Stuart’s many interesting works showing how abstract models can shed light on basic phenomena.
NK Boolean Networks

As you increase the number of connections...

(next slide)
Creates tunable landscapes for optimization problems

- Single-peaked
- Moderately rugged, correlated
- Randomly multi-peaked, uncorrelated
"I couldn't do what I did with Linux for Windows, even if I had the source code. The architecture just wouldn't support it."

- Linus Torvalds as quoted by Tim O’Reilly
“Too much of the windows source code consists of interdependent, tightly coupled layers for a single developer to drop in a replacement module.”

-Tim O’Reilly
When designing systems for community participation you must choose a good political architectural model.
Participatory Communities

<table>
<thead>
<tr>
<th>Platform</th>
<th>Visitors Contributing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flickr</td>
<td>0.16%</td>
</tr>
<tr>
<td>YouTube</td>
<td>0.2%</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>4.56%</td>
</tr>
</tbody>
</table>

Source: Hitwise, reported by Bill Tancer 4/25/2007 in Time Magazine

The WP figure is not without controversy
Web 2.0 has been successful in significantly broadening the amount material available to us, but reviewing the latest data reveals that we're still in the very early stages.


Suggesting.....

He thinks the relative % of visitors “participating” will grow!

Popular Press....this is the impression people will walk away with.

(People will think we failed...since its unlikely to change!)
Bradley Horowitz: Scalable Communities

creators
synthesizers
consumers
Bradley Horowitz: Scalable Communities

creators

synthesizers

consumers
Aaron Swartz

- Measure volume of content contributed
- Large number of users contributing large amounts of data, a low number of times

Both are correct! Contributions at many different scales.
Community at All Scales
Econ 101 Refresher

- Agents have diminishing marginal returns
- Economic actions produce negative feedback
- Nice textbook equilibria
- We all live happily ever after
Contemplate your typical clock

We see them everywhere.

Didn’t have to be this way.
Many early clocks were 24 hour clocks.
Pity the clock maker
who builds these today
except as novelty items,
which for some reason
our operations team
purchases
10,000 year clock
Conceived by
Danny Hillis, co-founder of Metaweb
Brian Arthur

• Positive feedback amplifies economic forces
• Multiple equilibria
• Initial conditions can determine outcome
• The “optimal” outcome isn’t always found
• Homo Economicus might be dead
Duncan Watts

Cumulative Advantage: The Experiment

Network effects as an explanation for VHS/Betamax
Largely Untestable...
experiment published in Science
Stone Soup Review

- Cumulative advantage
- Architecture for participation
- Community at all scales

Cumulative Advantage:
- once people started to contribute, it was easier for others to contribute.
- Positive feedback.

Architecture for Participation:
- Loosely coupled. If one contribution failed, others could succeed.

Community at all scales:
- Different type of contribution
- Individual contributions, family
Stone Soup Effect

- Cumulative advantage
- Architecture for participation
- Community at all scales

Multiply a scarce resource through community
Freebase is a database

- Not a set of documents
- Has fine-grained structure
- Lets you ask unexpected questions
- It’s machine readable
Freebase is practical

- Solves problems people currently have
- Not an ontology or inferencing engine
- Optimized for real-time execution
Freebase is universal

- Works across all languages
- Thousands of databases combined
- Everything is reconciled
### Bodybuilding

<table>
<thead>
<tr>
<th>Bodybuilding awards</th>
<th>Mr Olympia 1970</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mr Olympia 1971</td>
</tr>
<tr>
<td></td>
<td>Mr Olympia 1972</td>
</tr>
<tr>
<td></td>
<td>Mr Olympia 1973</td>
</tr>
<tr>
<td></td>
<td>Mr Olympia 1974</td>
</tr>
<tr>
<td>Admitted steroid use</td>
<td>Jawohl!</td>
</tr>
</tbody>
</table>

### Movie database

<table>
<thead>
<tr>
<th>Film performances</th>
<th>Hercules in New York</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conan the Barbarian</td>
</tr>
<tr>
<td></td>
<td>Predator</td>
</tr>
<tr>
<td></td>
<td>Kindergarten Cop</td>
</tr>
<tr>
<td></td>
<td>Terminator</td>
</tr>
<tr>
<td></td>
<td>True Lies</td>
</tr>
<tr>
<td></td>
<td>Jingle All the Way</td>
</tr>
<tr>
<td></td>
<td>The Last Action Hero</td>
</tr>
<tr>
<td></td>
<td>Total Recall</td>
</tr>
</tbody>
</table>

### Politics database

<table>
<thead>
<tr>
<th>Political party</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offices held</td>
<td>Governor - California</td>
</tr>
<tr>
<td>Political Mentor</td>
<td>Sargent Shriver</td>
</tr>
</tbody>
</table>
Arnold Schwarzenegger

- **Types:** Person (People), Film Actor (Film), Film Producer (Film), US Politician (Government), Athlete (danny's types), Interesting People (rubylabs's types), Herkunft (carlo_velten's types)
- **Also known as:** Arnold Alois Schwarzenegger, The Governor
- **Gender:** Male
- **Date of Birth:** Jul 30, 1947
- **Place of Birth:** Thal, Austria
- **Country Of Nationality:** United States
- **Profession:** Actor, Politician, Bodybuilder
- **Spouse(s):** Maria Shriver - Apr 26, 1986
- **Parents:** Aurelia Jadmy Schwarzenegger, Gustav Schwarzenegger
- **Children:** Christopher Schwarzenegger, Patrick Schwarzenegger, Christina Schwarzenegger, Katherine Schwarzenegger
- **Sibling(s):** Meinhard Schwarzenegger
- **Height (meters):** 1.88
- **Weight (kg):**
- **IMDB Entry:** http://www.imdb.com/
- **Party:** Republican Party
- **Sport:** Powerlifting, Bodybuilding

### Web Link(s)
- http://www.schwarzenegger.com

### Employment history
- Flex Magazine - executive editor
- Planet Hollywood - Co-founder

### Education
double-click to add "Education"

### Film performances
- **Twins** - Julius
- **True Lies** - Harry Trasker
- **Total Recall** - Quaid
- **The Terminator** - Cyborg t800
- **The Running Man**
- **The 6th Day**
- **Terminator 2: Judgment Day**
It is all about UI

...or almost all about UI

When launching we discovered...not surprisingly

UI reveals the platform
- graph store isn’t visible
- Service/Query tier isn’t visible (to most)

UI makes or breaks
All about Utility
It is all about Client UI Features

...or almost all about Client UI Features

People *really* like our UI
People are making big contributions through it.
but we can/will do more to improve it....think of it like an experiment. (Like a garden!)
If community/data products are to succeed, we all need to focus on usability.
Not particularly scholarly
.....but useful and interesting.
Talks about serving your customers (aka community)
What have loyalty programs shown us?

- Typical supermarket
  - Top quartile of customers spend 50x more per year than bottom quartile customers
- Why then do we have express lanes?
“Stones for Creators”

- Documentation
- Targets high value members of the community
- Invest appropriately
High Value Creators:
A community we are still working on servicing.

Topics
Dudes
Vägen ut
What's Cooking?
Twin Peaks
The Simpsons
24
The Best of Everything
The Friend Who Walked the West
The Sun Also Rises
Barefoot in the Park
The Langoliers
Laurel Canyon
The French Connection
Responsiveness is critical

**About Me**

I did time in Cambridge the rest of Francisco reached me and I ent...

I blame this turn of events on a Fourth Grade teacher who taught me BASIC on a CDC Cyber 70 (I am a Minnesotan through and through.) In college, my affliction became worse when I learned assembly on a VAX 11/750. And my destiny sealed, when, in graduate school, I had my own PDP-8 (it was the '90s you know) and I was compelled to write lots of SKED.
Small changes have a huge impact on behavior.

About Me

I did time in Cambridge the rest of Francisco reached me and I entered

I blame this turn of events on a Fourth Grade teacher who taught me BASIC on a CDC Cyber 70 (I am a Minnesotan through and through.)

In college, my affliction became worse when I learned assembly on a VAX 11/750. And my destiny sealed, when, in graduate school, I had my own PDP-8 (it was the '90s you know) and I was compelled to write lots of SKED.
“Data Stones”

- Seed Data
- Wikipedia, Music Brainz, Government data
- Books, Restaurants, Companies
- Early users: Gene, Proteins, Taxa
- Schema
- 700 Types

Listening to the users
Letting them guide us
Every assertion missed is data lost

Low cost ways to participate.

Do you back up your data?

Data missed, is like your site being down!
Freebase Will Prove Addictive

Danny Hillis' latest venture, Metaweb, is about to unveil its first product, the aptly named freebase, tomorrow. While freebase is still VERY alpha, with much of the basic functionality barely working, the idea is HUGE. In many ways, freebase is the bridge between the bottom up vision of Web 2.0 collective intelligence and the more structured world of the semantic web.

Early visitors to the site might be inclined to dismiss it as an early candidate to join the ranks of deadpool. After all, is there really room for what at first glance appears to be a new Wikipedia and the Open Directory Project, another site that purports to collect the world's information in one place?

THE Alpha Geek

Great review.

Really understands what we are trying to do.
In playing with the application since, I've found that some of these work in theory, but not yet in fact. For example, as I added myself as the CEO of O'Reilly Media, that should have also updated the entry for Tim O'Reilly (person), but it didn't.

....In theory, the entry for Sebastopol, California should have changed to reflect a new company located there, but it didn't.

But that's the way it is supposed to work...
In addition, two of the companies I added as O'Reilly subsidiaries aren't really subsidiaries, which suggests wholly-owned companies, but affiliates.

...I don't have the privileges to create a field in a company record, "Affiliated companies," but presumably, if it takes off, Metaweb will promote community members to the level where they can make these kinds of edits.

Doesn't he know you can prototype the changes...
Freebase as A Stone Soup of Data

- Cumulative advantage
- Architecture for participation
- Community at all scales

Cumulative Advantage:
- once people started to contribute....easier for others to contribute....positive feedback.

Architecture for Participation:
- loosely coupled. If one contribution failed, others could succeed.

Community at all scales:
- Schema Designers, Developers, Domain Administrators, Individual Contributors.