The Future Belongs to Data

Tim O’Reilly

O’Reilly Media, Inc. www.oreilly.com

Database Summit
May 30, 2008
How many of you have O’Reilly books?
What We Do At O'Reilly

Change the world by spreading the knowledge of innovators
How we do it

• Find interesting technologies and people innovating from the edge

• Amplify their effectiveness by spreading the information needed for others to follow them.

• Books
How we do it

• Find interesting technologies and people innovating from the edge

• Amplify their effectiveness by spreading the information needed for others to follow them.

• Books, Conferences
How we do it

• Find interesting technologies and people innovating from the edge

• Amplify their effectiveness by spreading the information needed for others to follow them.

• Books, Conferences, Online
“Martha Stewart for Geeks”

-- Newsweek
O'Reilly AlphaTech Ventures (OATV) is in the business of funding disruptors, innovators, and hackers of the status quo. From open source to the early days of the web and its recent resurgence, O'Reilly Media has played a key role as trend spotter and advocate for putting world-changing technologies on the map. OATV is our effort to bring the legacy, resources, and relationships of O'Reilly to each investment we make.

Profile: Eric Wilhelm, Instructables

“When people ask what they should post to Instructables, I tell them to share something they’re passionate about.”

– Eric Wilhelm, Co-Founder and CEO of Instructables

© 2007, O'Reilly AlphaTech Ventures, LLC | webmaster@oreilly.com
All trademarks and registered trademarks appearing on oatv.com are the property of their respective owners.
Watch the Alpha Geeks

• New technologies first exploited by hackers, then entrepreneurs, then platform players

• Three examples
  – Wireless community networks predict universal Wi-Fi
  – Screen scraping predicts web services and the internet as platform
  – “The pedal powered internet” predicts new focus on energy

Rob Flickenger and his potato chip can antenna
This is not new
Nor is it limited to technology...
Build, CRAFT, hack, play, MAKE.

Maker Faire is a two-day, family-friendly event that celebrates arts, crafts, engineering, science projects and the Do-It-Yourself (DIY) mindset. It's for creative, resourceful folks who like to tinker and love to make things. We call them Makers.

Tickets for Maker Faire Austin available now!


Maker Faire Austin: Call for Makers! Are you a maker or crafter who would like to show off your project at Maker Faire Austin, October 20-21? The call for makers is open! Deadline for entries extended to Thursday, September 6, 2007. Read more.

Appearing at Maker Faire Austin
Four-Player Tabletop MAME Cabinet

QUICK-PEEL AN EGG
Save 4 days of your life
"The future is here. It's just not evenly distributed yet."

--William Gibson
Pattern Recognition
The "Killer Apps" of the New Millennium
What Makes Them Interesting To Me

- The Internet, not the PC, is their platform
- Built on top of open source, but not themselves open source
- Services, not packaged applications
- Exploring how to become platform players via web services APIs
- Data aggregators, not just software
- Network effects from user contributions key to market dominance
- The most successful are “semantic learning systems”, leveraging implicit metadata
Infoware, not software
Web 2.0
web2.0
Desktop Application Stack

Proprietary Software (Control by API)

System Assembled from Standardized Commodity Components

Hardware Lock In By a Single-Source Supplier
Free and Open Source Software

Cheap Commodity PCs

Intel Inside
### Internet Application Stack

<table>
<thead>
<tr>
<th>eBay</th>
<th>Google</th>
<th>Amazon.com</th>
<th>MapQuest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camel</td>
<td>Eclipse</td>
<td>Apache</td>
<td>MySQL</td>
</tr>
<tr>
<td>CD</td>
<td>Network Solutions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Proprietary Software As a Service**

**Integration of Commodity Components**

**Subsystem-Level Lock In**
Web 2.0
web 2.0
Collective Intelligence

Harnessing network effects to build applications that get better the more people use them
## Turning 1.0 into 2.0

<table>
<thead>
<tr>
<th>Feature</th>
<th>Web 2.0</th>
<th>Phone Co</th>
</tr>
</thead>
<tbody>
<tr>
<td>Massive Data Centers</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Software as a service</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data from customers</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data gets better all the time</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data mining of customer behavior</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Real time user-facing services based on that data</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
How Ridiculous Is This?

• Dialed calls (last 10)
• Received calls (last 10)
• Missed calls (last 10)

My phone company remembers every call. Why don’t they remember it for me?

My phone and my email already know who my friends are.
Web 2.0 is about finding *meaning* in user-generated data, and turning that meaning into real-time user-facing services.
PageRank

From Wikipedia, the free encyclopedia

PageRank is a link analysis algorithm that assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set. The algorithm may be applied to any collection of entities with reciprocal quotations and references. The numerical weight that it assigns to any given element $E$ is also called the PageRank of $E$ and denoted by $PR(E)$.

PageRank was developed at Stanford University by Larry Page (hence the name PageRank) and later Sergey Brin as part of a research project about a new kind of search engine. The project started in 1995 and led to a functional prototype, named Google, in 1998. Shortly after, Page and Brin founded Google Inc., the company behind the Google search engine. While just one of many factors which determine the ranking of Google search results, PageRank continues to provide the basis for all of Google's web search tools.

The name PageRank is a trademark of Google. The PageRank process has been patented (U.S. Patent 6,285,999). The patent is not assigned to Google but to Stanford University.

Contents

1 General description
2 PageRank algorithm
   2.1 Simplified PageRank algorithm
   2.2 PageRank algorithm including damping factor
3 PageRank variations
   3.1 Google Toolbar
   3.2 Google directory PageRank
   3.3 False or spoofed PageRank
   3.4 Manipulating PageRank
   3.5 Other uses of PageRank
4 Google's "rel=nofollow" proposal
5 See also
6 References
7 Further reading
8 External links

General description

Google describes PageRank[2]

"PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence,
### Wells - Personal

**Tuesday, April 15th**

**ATM Withdrawal**  
Tags: cash  
7 tips available, including Using cash for discretionary spending.

**T Mobile**  
Tags: communications

**Comcast Cable**  
Rating: Captive  
Tags: cable_tv  
1 tip available: Students: Take advantage of Comcast's sharply discounted student rate.

**Thursday, March 27th**

**Weatherford BMW**  
Tags: auto  
7 tips available, including vs. Bavarian Pro

**Monday, March 24th**

**Crate & Barrel**  
Tags: household  
10 tips available, including Recent recalls prove: Generic foods/items often the same as Name Brands.

**The Mon**

**Bi Rite Market**  
Tags: birthday  
1 tip available: Personalize everything.

**Friday, March 21st**

**Stem Florist**  
Tags: birthday  
1 tip available: Personalize everything.
### Bavarian Pro vs. Weatherford BMW

<table>
<thead>
<tr>
<th></th>
<th>Bavarian Pro</th>
<th>Weatherford BMW</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spending</strong></td>
<td>$611.65</td>
<td>$1327.29</td>
</tr>
<tr>
<td><strong>Visits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Popularity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Recommend it</strong></td>
<td>96%</td>
<td>17%</td>
</tr>
<tr>
<td><strong>Often visit</strong></td>
<td>33%</td>
<td>15%</td>
</tr>
<tr>
<td><strong>Never return</strong></td>
<td>9%</td>
<td>65%</td>
</tr>
<tr>
<td><strong>Commonly</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>repairs $593.70</td>
<td>auto $1244.29</td>
</tr>
<tr>
<td></td>
<td>auto $603.98</td>
<td>car $924.01</td>
</tr>
<tr>
<td></td>
<td>car $300.43</td>
<td>car repair $1318.18</td>
</tr>
</tbody>
</table>

**Tips**
- Bavarian Pro (1)
- Weatherford BMW (1)

*debbie says:*
The Programmable Web
Mashup Dashboard
Thousands of web mashups with new updates daily

TweetWheel
Find out which of your Twitter friends know each other.
APIs: Twitter
Tags: fun, messaging, microblogging

All Mashups | Popular Mashups | Matrix | Add Yours | How-To Guide

Top Mashup Tags
- mapping (39%)
- photo (10%)
- shopping (9%)
- search (8%)
- video (7%)
- travel (6%)
- news (4%)
- sports (4%)
- messaging (4%)
- real estate (4%)

Click on a slice or label to see those mashups

ProgrammableWeb.com 05/29/08
Mashup Dashboard

Thousands of web mashups with new updates daily


TweetWheel
Find out which of your Twitter friends know each other.
APIs: Twitter
Tags: fun, messaging, microblogging

Our Sponsors

Popular Directory Searches

Celebrity Mashups
Video Mashups
Popular New
Subsystems

- Location
- Identity
- Products
- Advertising relevance
- Music
- Events
What do we need to do differently?

• To support this future?
• To make sure that the future internet OS continues to work like today’s (i.e. mostly open and interoperable)
• To make data mashups less of a hack and more of a truly composable set of data services?
Web 2.0 is Sharecropping

I gave a 5 minute talk at Ignite Boston this evening. The basic premise, for those of you who don’t do flash or don’t feel like clicking through is this: If you don’t own your tools, you’re going to be in a whole mess of trouble. Maybe not today. Maybe not tomorrow. But one day, you’re going to be very, very unhappy with the fact that you’ve given up your right to software self-determination.

Web 2.0 is Sharecropping

Jesse Vincent
jesse@fsck.com
For more information

http://www.oreilly.com
http://tim.oreilly.com
http://radar.oreilly.com
http://www.oreillynnet.com/go/web2