GrassRoots:
Socially-Driven Web Sites for the Masses

Frank Uyeda
Diwaker Gupta, Amin Vahdat, George Varghese

University of California, San Diego
This Talk in a Nutshell

• Grass-roots communities wish to have websites that allow them to submit and flexibly search for data.

• These communities require tools that are simpler than those currently available (e.g. Apache, PHP, MySQL).

• To build such tools we need models for:
  – data objects, page layout, and, most importantly, search.

• We have instantiated these model in an easy to use language called GrassRoots, and compiler called GR.
Web Application Development

Pre-packaged
- Configuration
  - Wikis, Blogs, CMS (WordPress, Joomla)
- Database Configuration (Pre-specified)

Custom
- Presentation (Dreamweaver)
- Application Logic (PHP, Ruby, Python)
  - Web Frameworks (Zend, Rails, Django)
- Database Configuration

GrassRoots
- Presentation
- Application Model
- Application Logic (Auto Generated)
- Database Configuration (Auto Generated)

Database (MySQL)
Outline

- Motivation
- Modeling
- Results
The Spread of the Social Web

• Big social networks -- international phenomenon.
  – North America: Facebook (250M)
  – South America / India: Orkut (67M users)

• Not just “social networks”
  – YouTube (video sharing), Digg (social bookmarking)

• Growing interest in smaller sites specialized by industry, enterprise and communities.
  – Wellpoint (insurance)
  – Cisco (company specific)
Who Wants Social Sites?

• Lots of people!
  • **Example:** Physics Researchers
    – Want to share data sets, tag interesting features
  • **Example:** Digital Artists
    – Want to share data visualization programs & collaborate
  • **Example:** Local Parents & Baby-sitters
    – Want job postings, referral network

• Require database & application logic.

These communities **lack resources and expertise.**
Need cheap, easy-to-use tools!
Limitations of Existing Tools

• **Difficult to prototype new ideas**
  – Not clear which web development framework will work best
  – Require knowledge of database schemas, web programming languages, design techniques

• **Significant time and expertise needed to develop an operational site**
  – Large, complex code base
  – Integration with user management, access control and web API’s
  – Engineer for security and privacy

• **General techniques to scale are unknown**
  – Performance tuning is “black magic”
  – Hire a consultant

*Addressed in this talk*
## Opportunity: Different But Similar

<table>
<thead>
<tr>
<th>Site</th>
<th>Objects</th>
<th>Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flickr</td>
<td>Images</td>
<td>Keyword, Tags, Comparison (geo-tags)</td>
</tr>
<tr>
<td>YouTube</td>
<td>Video</td>
<td>Keyword, Tags</td>
</tr>
<tr>
<td>Last.fm</td>
<td>Audio</td>
<td>Tags, Structural</td>
</tr>
<tr>
<td>Del.icio.us</td>
<td>URLs</td>
<td>Tags</td>
</tr>
<tr>
<td>Digg</td>
<td>URLs</td>
<td>Taxonomy, Keyword</td>
</tr>
<tr>
<td>Craigslist</td>
<td>Listings (Image + Text)</td>
<td>Taxonomy, Keyword, Comparison</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>Articles (Text)</td>
<td>Keyword, Structural</td>
</tr>
<tr>
<td>Facebook</td>
<td>User Profile (Image + Text)</td>
<td>Structural, Tags</td>
</tr>
</tbody>
</table>
The Essence: Search & Submit

- Users create and upload content
- Content organized/ranked based on user input
- Search based on:
  - Associated keywords (e.g., Del.icio.us tags)
  - Structural relationships (e.g., friends in Facebook)
  - Taxonomy / Hierarchy (e.g., categories in Digg)
  - Comparison / Proximity (e.g., geo-tagging in Flickr)
Approach: Add Abstraction

- **Model:** Create an abstract model for community driven web sites.

- **Specify:** Allow developers to express an abstract site model in the GrassRoots language.

- **Compile:** The GrassRoots compiler generates web code and configures storage.
Outline

• Motivation
• Modeling
• Results
New Insights Behind Our Model

• Insight 1 (Layout): Pages are composed of panes that are populated by search results.

• Insight 2 (Navigation): All navigation is search.

• Insight 3 (Search): Graph-based search with attribute filtering covers existing social search mechanisms.
Data Modeling: like UML

• GrassRoots objects:
  – High-level types (e.g., video, image, text),
  – Composite types like C-structs
  – Built in attributes: taggable, commentable

• Relationships as Graphs
  – General Graph (e.g. friends in Facebook)
  – Directed Graph (e.g. YouTube Subscribers)
  – Hierarchy / Tree (e.g. Craigslist categories)
What is flickr™?

- Community photo-sharing website
- Users associate with each other
- Images organized using:
  - Keyword tags
  - User “photo sets”
  - Group “pools”
Insight 1: Page Layout Model

Static Pane

Search Pane

Picture Summary Pane

Set Summary Pane
Formalizing Insight 1

• A page is composed of one or more panes
  – Panes are populated by embedded searches

• Pane
  – A region within the Page
  – Handles the input and output of a particular data collection, or displays static content
  – Defined once, and reused across many pages
  – Pane aesthetics customized with CSS
Insight 2: Navigation is Search

**User-specified Search**

**Pre-specified Search**
• Navigation associates clicking on a data object to a page and search parameters.
  – Pages embed searches.
  – the “linkto” keyword provides parameters to searches.

• Syntax:
  [object in pane] -> linkto [page]( [params] );

• Example: Clicking on a user’s name displays all pictures owned by that user.
  user -> linkto all_users_pictures( user );
Insight 3: Graph-based Search Model

Hierarchy

Attribute

Filter

Ordering

san diego craigslist > north san diego county > for sale / wanted > antiques

search for: table  
price: 500 - 1000

Sort by: most recent or best match

Aug 11 - Antique SOLID American Oak Pedestal Table Early 1900â€™s - $549 - (Mission Valley) pic

Aug 11 - Antique dining room furniture - $1000 - (Vista) pic

Aug 11 - SPARKLING SILVER ELEGANCE FOR CELEBRATING – REGALE YOUR TABLE & DECOR - $695 - (South Orange County) pic

Aug 11 - Standard Antique Wood Bed circa 1900’s - $500 - (San Marcos, CA.) pic

Aug 10 - Dining Room Table and Chairs 6 Antique - $500 - (Del Mar)

Aug 10 - Antique oak Dining table 6 chairs and Hutch. - $500 - (Fallbrook)

Aug 8 - Antique Table - $500 - (Encinitas) pic

Aug 6 - 60" Round Early 20th c. Trestle Table - $600 - (Carlsbad-La Costa-Encinitas) pic

Sort by: most recent or best match

Found: 8 Displaying: 1 - 8

Copyright © 2009 craigslist, inc.

Back to top of page
Formalizing Insight 3

SELECT <collection>
[FROM <structural relation>]
[WHERE <filter condition> ...]
[ORDER BY <ranking function> ...]

• Structural relations:
  – Graphs: neighbor,
  – Tree: subtree, parent, children
• Filter conditions:
  – matches, contains, greater than, between, within distance, tagged by
• Ranking:
  – Combination of attributes or graph properties (e.g. node degree)
Outline

• Motivation
• Modeling
• Results
Abstraction Allow Powerful Modifications

• Claim: Small changes to the specification provide important features at low cost.
• Example:
  – Flickr tags photos with keywords
  – Facebook tags photos with Users

• How do we change our Flickr specification to incorporate User tagging?
Facebook User Tagging

Photos of Kevin Kuo

In this photo: Christopher Lau, Lix, Kevin Kuo (photos), Sharon Prude Williams, Rigobert Vindola, Vanessa Arriola

Added July 28

From the album: "Old school pics- for posterity sake" by Kevin Kuo
Adding User Tags to Flickr

COMPOSITE Picture {
    IMAGE   pic;
    TEXT    pic_title;
    TEXT    pic_description;
} (taggable, taggable by USER, commentable);

PAGE pic_detail( Picture p ) {
    Detail(Picture) main : LOOKUP Picture p;
}

Detail Pane : Picture{
    _owner -> linkto user_profile( _owner );
    _tag -> linkto tag_result( _tag );
    _tag.USER -> linkto user_profile( _tag.USER );
    "add tag" -> linkto add_tag( this );
    "add to set" -> linkto add_to_set( this );
    "add to group" -> linkto add_to_group( this );
}
name: Bob
Tags: value: log value: ducks
Tags: name: Joe name: Bob
Comments:

pic_description: I like ducks
add to group
add to set
add tag
GR Compiler Implementation

- Implemented in Java (~15K lines of code)
- Page generation in various languages.
  - Currently supports PHP
GR vs. Ruby on Rails: Abstraction and Performance

Implemented Flickr-like site using Rails plug-ins & GR

- **Code Complexity**
  - 50 lines of code across 19 files (vs. 180 lines in 1 file)

- **Picture retrieval throughput:**
  - Grassroots gives 2x max throughput.
    - Grassroots only generates necessary code
    - Ruby has large call tree.

- **Tag search throughput** – 20 most recent
  - 500:1 performance difference, favoring GrassRoots.
  - Suspect poor SQL queries & failure to parallelize.
Summary

• Need better tools for constructing social sites.
• Leverage the commonalities among sites.
• We provide abstractions to ease development.
  – Pages are composed of panes
  – All navigation is search
  – Graph-based search with attribute filtering
• Abstractions provide flexibility & opportunity for optimization.