The Future of MediaWiki and the Wikimedia projects

Erik Möller – August 6, 2005
The Purpose of Technology Research

- Many (thousands) very active content producers
- Very few (less than 10) very active developers
- New projects with specific needs
- Research can
  - Identify useful software enhancements
  - Write specifications and make recommendations
  - Supervise and review implementation process
  - Get the community involved in technical processes

Wikimania - August 6, 2005
Wikimedia Research Network

• Attempt to bring individuals together to
  – work on specs
  – study Wikimedia content and communities
  – coordinate external contacts
  – organize community meetings

• Current activities
  – Single login specs
  – Development tasks
  – User survey
Why peer review?

• Beyond existing mechanisms
• Main criticism against Wikipedia
  – From academia
  – From search engines
  – From pundits
• Fact-checking is a collaborative process
• As much work as the encyclopedia itself
• First step: Article survey

Wikimania - August 6, 2005
Article survey

Revisions for Infinite monkey theorem

<table>
<thead>
<tr>
<th>Revisions for Infinite monkey theorem</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Revision #1044</td>
<td>View this revision</td>
<td>2005-08-02 08:18:31</td>
<td>by Admin</td>
<td>&quot;&quot;</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>No opinion</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Neutral point of view</td>
<td>No opinion</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Factual accuracy</td>
<td>No opinion</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Quality of references</td>
<td>No opinion</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Completeness</td>
<td>No opinion</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Quality of lead section</td>
<td>No opinion</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Images and illustrations</td>
<td>No opinion</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Relevance</td>
<td>No opinion</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Use my previous assessment where selected 'No opinion' | Clear my older validation data | OK |
Page protection

- Pages only editable by sysops
- Edit warring or distributed vandalism, decided by sysop
- English Wikipedia: avg. 12 protections per day
- However, some pages stay protected very long
  - Lack of processes or responsibility
  - e.g. Sexual abuse of children

Wikimania - August 6, 2005
Alternatives

• Code which exists (recent, not in use):
  - User edits invisible copy of page
  - Sysops can “verify” a revision
  - Displayed copy is last verified one during period of protection

• Ideal solution:
  - If no sysop “verifies”, page is automatically published if no activity for n minutes
My thoughts on peer review

• Must be “wiki-like”
  – Fast and easy
  – Consensus-based

• One basic concept for Wikipedia, Wikinews, etc.

• Distinction between “stable” and “unstable” revisions

• Offer to view only stable revisions
How are articles reviewed?

- Categorized review queues
- Initial review
  - Sectioned discussion page
    - Neutrality, Accuracy, Comprehensiveness etc.
    - A set number (e.g. 3) has to agree in each of these areas
      - There must be no unresolved disagreements (or high threshold)
  - After time X elapses, revision is tagged as stable in areas where stable

*Wikimania - August 6, 2005*
Subsequent changes

- Problem: Articles change often
- Subsequent review can be simplified:
  - Special queues which show diffs
  - Only continuing consensus required within timeframe
- Possibility: make use of minor edit flags
  - Not always correctly set
  - Automatic evaluation flawed
- Possibility: sysop speed-up
Synthesis

- Page protection and peer review very similar
  - Protection is another category of review: vandalism watch
  - Users can choose whether they want to see, link to or crawl
  - Revisions that may be inaccurate
  - Revisions that may not be neutral
  - Revisions that may not be comprehensive
  - Revisions that may be vandalism / nonsense
    - Of all articles / those subject to frequent editing

Wikimania - August 6, 2005
Discussion is essential

Wikipedia - August 6, 2005
Wikidata

• Instead of simple text fields, arbitrary data types
• Schemas associated with namespaces
• Each namespace is a table in a rel. DB
  – Some data displayed, other data relational
• Individual fields have histories, diffs, etc.
• Thousands of applications
  – Ultimate Wiktionary (sets language needs)
  – Infoboxes
Namespace manager

**System namespace**
- **Internal slot**: 2
- **Support subpages**: 
- **Search by default**: 
- **Default link prefix**: 
- **Parent namespace slot**: blank for none

**Delete Names**
- **N/A Media**
- **N/A Special**

**System namespace**
- **Internal slot**: 1
- **Support subpages**: 
- **Search by default**: 
- **Default link prefix**: 
- **Parent namespace slot**: blank for none

**Delete Names**
- **N/A Special**

Wikimania - August 6, 2005
Company namespace

Editing Company: Google

You've followed a link to a page that doesn't exist yet. To create the page, start typing in the box below (see the help page for more info). If you are here by mistake, just click your browser's back button.

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Public</td>
</tr>
<tr>
<td>Founded (year)</td>
<td></td>
</tr>
<tr>
<td>Current main location</td>
<td></td>
</tr>
<tr>
<td>Key people</td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td>Select one</td>
</tr>
<tr>
<td>Products</td>
<td></td>
</tr>
<tr>
<td>Revenue</td>
<td>USD (2004)</td>
</tr>
<tr>
<td>Website</td>
<td></td>
</tr>
</tbody>
</table>

Edit summary: [Save data] [Show preview]
Country namespace

**Editing Country: Germany**

You've followed a link to a page that doesn't exist yet. To create the page, start typing in the box below (see the [help page](#) for more info). If you are here by mistake, just click your browser's **back** button.

<table>
<thead>
<tr>
<th>Flag</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence - From</td>
<td></td>
</tr>
<tr>
<td>Population (2005)</td>
<td></td>
</tr>
<tr>
<td>GDP (PPP, 2005)</td>
<td></td>
</tr>
<tr>
<td>Head of State</td>
<td></td>
</tr>
</tbody>
</table>

**Edit summary:**

[Save data] [Show preview]

*Wikimania - August 6, 2005*
Tagging

• Current category system very stupid
  – Tagging = editing
  – Edit many pages to make simple changes
    • Renaming, reordering
  – No proper hierarchical view
  – No synonyms or l10n

• Needed: Tag manager with autocompletion
  – Manage hierarchy, add tags to pages
Project-specific needs

- Wikinews
  - Easy submission interface (Wikidata)
  - Real-time editing
- Wikisource
  - Translation interface
- Wikimedia Commons
  - Media manager, upload form
  - Combined tag queries
Tag manager

- Most frequently used tags visible
- New ones can be added easily
- "Google Suggest" style completion
Project needs (cont.)

• Wikibooks
  – Modularization
  – Book TOCs
• Wikiquote
  – Quotes as elements (Wikidata)
• Wikipedia
  – Smart references
  – “Unreferenced” markers
On the horizon

• Wikiversity/Wikisophia
  – E-learning
  – Research
  – Publishing

• Geodata for articles (already happening)
  – Integration with Google Earth

• Map generation (automatic and human)

• More cool extensions

Wikimania - August 6, 2005
The Road Ahead™

- Single login
- Well-structured list of development tasks
  - Familiarity with MediaWiki
  - General coding skill requirements
  - Size of task
  - Importance of task
- Test case for targeted development
- Work with universities, companies

*Wikimania - August 6, 2005*
Wiki power

• Parallels between wiki and code
• Natural synthesis of wikis and blogs
• What is a wiki?
  – Philosophy of maximizing participation
  – Toolset rather than single tool
• Wikimedia has chosen to launch projects in key areas
  – Responsibility to innovate technology