WikiTruthiness

Displaying the level of contentiousness of Wikipedia pages via a coloring scheme.

http://www.wikitruthiness.com

Katherine Baker
David Koenig

Aaron Miller
Cullen Walsh
Aspirations v. Reality

Goals: Total article content contention determined from reversions, edit wars, other indicators at a paragraph/sentence level.

Final Results: Determine recent article contention on a sentence → word level by assigning scores based on content insertion, deletion, and modification.
Technical Overview

Choose Result

Have Cached Result?

Yes

Display

No

Search

Search Results

Compute Version Diffs

Fetch Wikipedia Content

Analyze Diff Graph

Mark Up Content w/ Analysis Results

Cache Result

Refresh

http://www.wikitruthiness.com/
Front End Details

- **Search Utilizing Google Search API**: initiate processing (Ruby on Rails)
- **Wikipedia Scraper**: fetch data for processing (RoR)
- **Render Output w/ Mediawiki API**: display the results (Ruby on Rails)

Work by Cullen Walsh

http://www.wikitruthiness.com/
• **Difference Analysis:**
  version differences graph (Python)

• **Contention Identification:**
  linear scaling (KDE approx.) (Python)

Work by David Koenig

http://www.wikitruthiness.com/
Middleware Details

- **AWS**
  - S3 – Caching Results & Wikipedia Data
  - EC2 – small instance for front end; high CPU instance for analysis
- **MySQL**
  - Queuing requests, storing Wikipedia article versions (30 most recent)

Work by David Koenig and Cullen Walsh
Demonstration

WikiTruthiness

Search: Basic Linux "Search"

BasicLinux - Wikipedia, the free encyclopedia
BasicLinux also known as BasLin, BL3, or BasLinux, is a mini Linux distribution optimized for old (386 - 486) PCs, and laptops which uses the commandline or ...
View original Google result

fuser (Unix) - Wikipedia, the free encyclopedia
[edit] External links: fuser(1) : Identify processes using file descriptor icon, This Unix-related article is a stub.
View original Google result

List of BASIC dialects by platform - Wikipedia, the free encyclopedia
Chipmunk Basic (Apple Macintosh, CLI ports for Win32, MacOS, Linux, Unix) — A direct descendant of the original ...
View original Google result

Rocky Mountain BASIC - Wikipedia, the free encyclopedia
Agilent now refers customers to either TransEra for BASIC for Linux, or ...
View original Google result

List of BASIC dialects - Wikipedia, the free encyclopedia
Blunt Axe Basic (aka BXBASC) (Win32, Linux) (B)Bxbasic (Mac OS, Linux, Unix) — Direct descendant of the ...
View original Google result

BasicLinux also known as BasLin \ BL3 , or BasLinux , is a mini Linux distribution optimized for old \ 386 - 486 \ PCs \ and laptops which uses the \ commandline or a simple GUI .

It's based on stripped down version of Slackware 3.5 and uses an obsolete, but small, v2.2.26 kernel with busybox to provide a low-RAM Linux \ 11 \, capable of browsing the web, doing email , and functioning as an X terminal .

It's offered in 2 versions, a hard drive one that boots from DOS , and a floppy drive one that uses 2 disks.

Both versions can be used with an 386 CPU . The Hard Drive installable version requires 3 MB of RAM and 20 MB of disk space, plus a bootable DOS. The floppy version needs 12 MB of RAM.

Installed applications include: IcaWin MagicPoint Mutt mpg321

References

External links
- BasLin official site
- BasLin Mailing List Archives
- BasLin at Freshmeat
- BasLin on Softpedia

<table>
<thead>
<tr>
<th>Linux distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arch Linux</td>
</tr>
</tbody>
</table>

Project for CSE 454, University of Washington, Fall 2010
Contact us at (ajmiller)walshjdouble0keenig)@cs.washington.edu
Search powered by Google
Experimental Methodology

• Compare against related work: WikiTrust

• WikiTrust highlights untrustworthy words in a Wikipedia article based on many parameters

• Compute precision, recall against WikiTrust

• **True Positives** = # blocks which contain > 0 WikiTrust highlighted words

• **False Positives** = # blocks which do not contain any WikiTrust highlighted words

• **False Negatives** = # WikiTrust highlighted words which are not within our blocks

http://www.wikitruthiness.com/
## Experimental Results

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst</td>
<td>10.84%</td>
<td>52.43%</td>
</tr>
<tr>
<td>Average</td>
<td>20.25%</td>
<td>68.93%</td>
</tr>
<tr>
<td>Best</td>
<td>38.82%</td>
<td>79.37%</td>
</tr>
</tbody>
</table>

Results of evaluating 33 articles

Work by Katherine Baker and Aaron Miller

http://www.wikitruthiness.com/
Challenges

- Getting the algorithm and coloring to work
- Obtaining cache coherency across memcached, S3, and MySQL
- Comparing data formats of WikiTrust and WikiTruthiness outputs
- Retrieving articles from Wikipedia in a timely fashion

http://www.wikitruhiness.com/
What We Learned

• Mixing technologies and having them interface is difficult
• Choosing your development language is important (e.g. Python not always best)
• Limited version history to 30 most current for speed; in production, would use more revisions
• Good evaluation requires significant time and effort, esp. when crawling and processing-intensive algorithms are involved

http://www.wikitruthiness.com/
Questions

Email:

{ajmiller, kbaker4, koenig,ckwalsh}@cs.washington.edu