WikiTruthiness

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

http://www.wikitruthiness.com
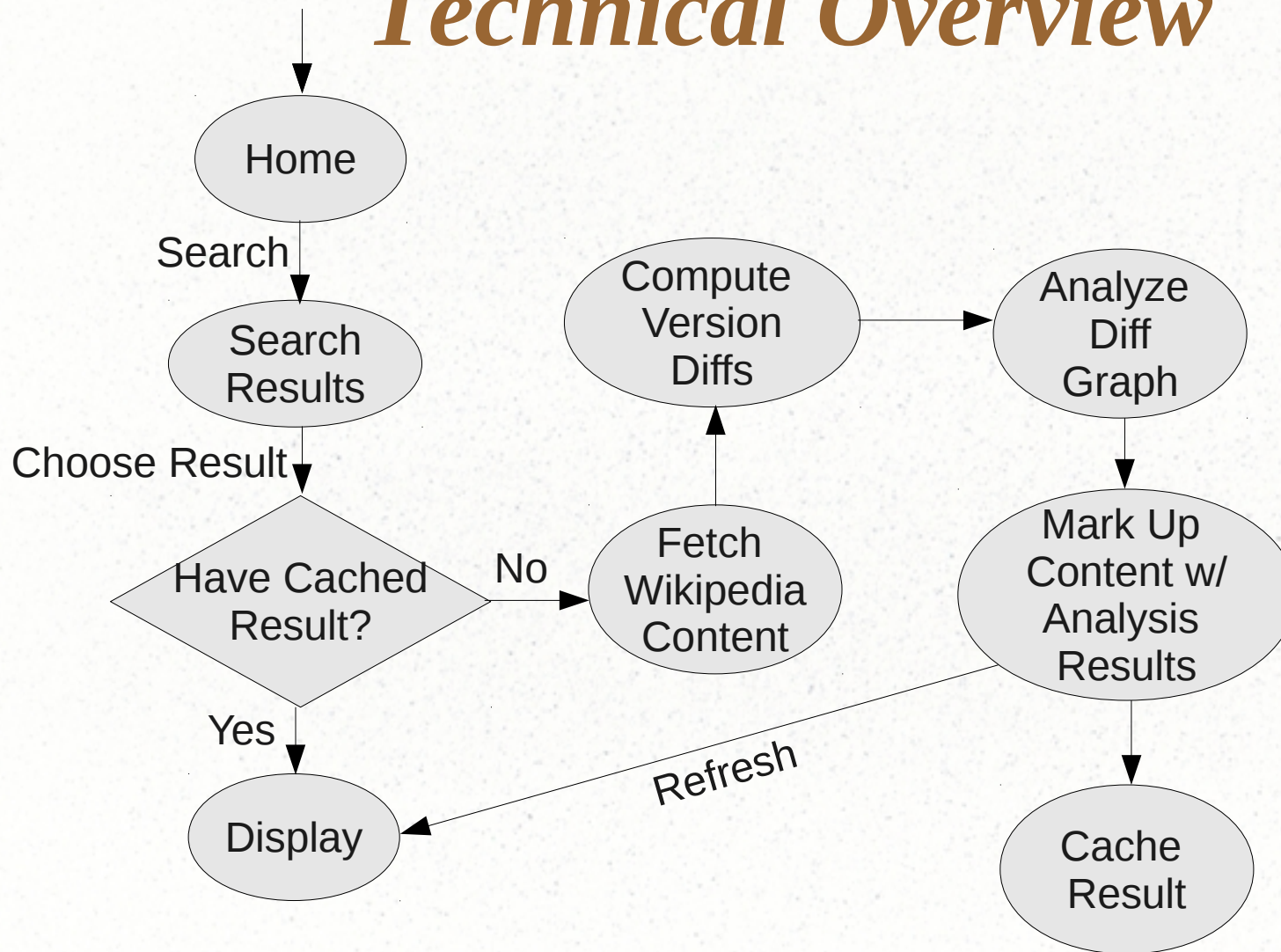
Katherine Baker

David Koenig

Aaron Miller

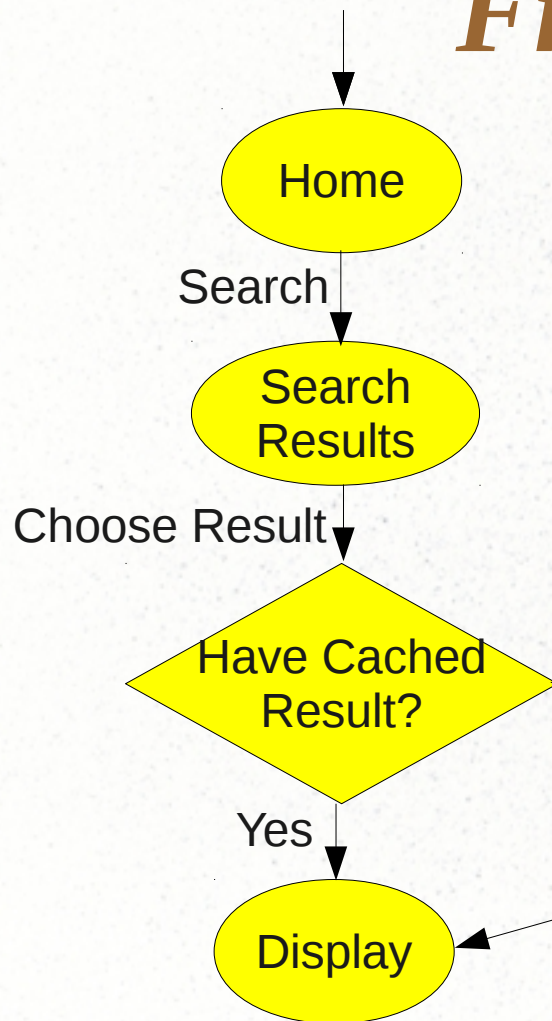Cullen Walsh

Aspirations v. Reality

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

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

http://www.wikitruthiness.com/
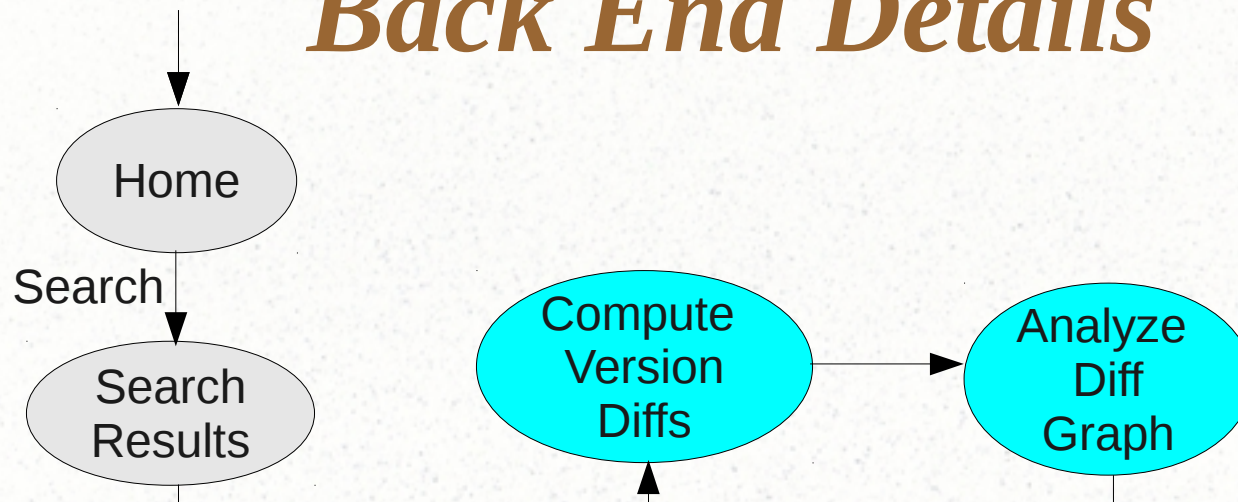
Technical Overview

```
        Home
          │
          │ Search
          ▼
       Search
       Results
          │
  Choose Result
          │
          ▼
    Have Cached          No        Fetch            Compute          Analyze
      Result?      ──────────►   Wikipedia   ───►   Version   ───►     Diff
          │                       Content            Diffs             Graph
          │ Yes                                                          │
          ▼                                                              ▼
       Display   ◄──────────────────────────  Mark Up Content w/
                        Refresh                 Analysis Results
                                                      │
                                                      ▼
                                                    Cache
                                                    Result
```

Front End Details

Home

Search

Search Results

Choose Result

Have Cached Result?

Yes

Display

- <u>Search Utilizing Google Search API</u>: initiate processing (Ruby on Rails)

- <u>Wikipedia Scraper:</u> fetch data for processing (RoR)

- <u>Render Output w/ Mediawiki API</u>: display the results (Ruby on Rails)

Work by Cullen Walsh

http://www.wikitruthiness.com/

Back End Details

Home

Search

Search
Results

Compute
Version
Diffs

Analyze
Diff
Graph

- <u>Difference Analysis:</u>
  version differences graph (Python)

- <u>Contention Identification:</u>
  linear scaling (KDE approx.) (Python)

Work by David Koenig
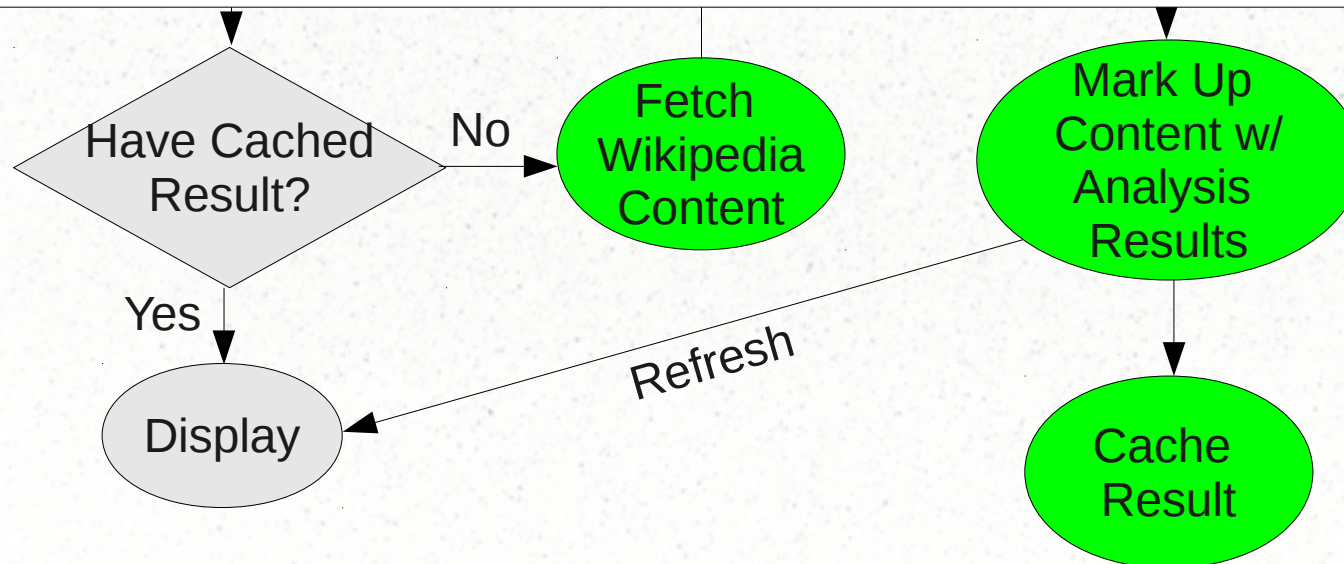
http://www.wikitruthiness.com/

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

Have Cached Result? — No → Fetch Wikipedia Content → Mark Up Content w/ Analysis Results

Yes → Display

Mark Up Content w/ Analysis Results → Refresh → Display

Mark Up Content w/ Analysis Results → Cache Result

http://www.wikitruthiness.com/

Demonstration

WikiTruthiness

Search: [Basic Linux] [Search]

Previous Page Next Page

1. 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

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

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

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

5. List of BASIC dialects - Wikipedia, the free encyclopedia
   Blunt Axe Basic (aka BXBASM) (Win32, Linux) [9] Bxbasic i... Mac OS, Linux, Unix) — Direct descendant of the ...
   View original Google result

WikiTruthiness

Search: [_____] [Search]

BasicLinux

View unprocessed article • View this article on Wikipedia

Please help improve this article by expanding it. Further information might be found on the talk page . *(April 2008)*

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 [1] , 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: IceWM MagicPoint Mutt mp321

References

1. ^ http://distro. ibiblio. org/pub/linux/distributions/baslinux/

External links

- BasicLinux official site
- BasicLinux Mailing List Archives
- BasicLinux at Freshmeat
- BasicLinux on Softpedia

v ◆ d ◆ e	**Linux distributions**
Arch Linux ◆ CentOS ◆ Debian ◆ Fedora ◆ Gentoo ◆ Knoppix ◆ Linux Mint ◆ Mandriva ◆ PCLinuxOS ◆ Red Hat Enterprise Linux ◆ Slackware ◆ SUSE ◆ Ubuntu ◆ *more*◆	

Project for CSE 454, University of Washington, Fall 2010
Contact us at (ajmiller|ckwalsh|kbaker4|koenig)@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*

- <u>True Positives</u> = # blocks which contain > 0 WikiTrust highlighted words

- <u>False Positives</u> = # blocks which do not contain any WikiTrust highlighted words

- <u>False Negatives</u> = # WikiTrust highlighted words which are not within our blocks

http://www.wikitruthiness.com/

Experimental Results

	Precision	Recall
Worst	10.84%	52.43%
Average	20.25%	68.93%
Best	38.82%	79.37%

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.wikitruthiness.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