A Comparative Web Browser for Browsing and Comparing Web Pages

By Akiyo Nadamoto and Katsumi Tanaka

Presented by Peng Zang
Comparative Viewing, a Hassle

- Simultaneous and independent navigation of many windows
- Manual content synchronization
- Automated comparison services incomplete
Proposed Solution: CWB!
Featuring

- Automatic content-synchronization
- Linked navigation of comparison sites
- Common word highlighting on selection
How Does it Work?

- Preprocessing to create keywords DB
- Auto content-sync on triggers:
  - Anchor clicking
  - Page scrolling
  - Back/Forward navigation
- Highlight common words on user selection.
Creation of Keywords DB

- **Motivation:** Similarity checking has to be real time

- **Basic Process:**
  - Separate content into paragraphs
  - Compute word frequency
  - Frequency $\times$ weight vector $\rightarrow$ word vector
  - Get title and subtitles
  - Fix subject keywords (from titles) and content keywords (from content) via truncation.
Auto Content Sync

- Thresholded search

- Case 1: Web page has subtitles
  - BFS for similar titles/subtitles
  - Content similarity search

- Case 2: Simple page, no subtitles
  - Content similarity search

- Content search via subject or content vector?
Concerns

- **Arbitrary Thresholds**
  - To be considered “similar”
  - For truncating length of subject/content keywords

- Where do we get weight vector?

- Preprocessing on server-side

- Hand coded webpage cleaning
More Concerns

- Compared pages must be heavily similar
- Assumes interest paragraph at the middle
- What if wrongly synchronized?
- Only two websites compared.
- Results lackluster
Questions

- Is this really comparative?
- What's the contribution?
- Is it practical or the need real?
- Who would use it?
Future Work

- Link based
- Client side
- Doesn't require heavily similar passages
- Based on anchor text alone