

# 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!

The screenshot displays the Barnes & Noble website interface. At the top, the URL is <http://www.amazon.com/exec/obidos/subst/home/home.html/102-6661912-5275313>. The main header features the Barnes & Noble logo, a "FREE SHIPPING" offer (When you spend \$25 or more), and a navigation menu with categories like HOME, BOOKS, USED & OUT OF PRINT, BUSINESS & TECHNOLOGY, MEDICINE & SCIENCE, NEW & USED TEXTBOOKS, LIBROS EN ESPAÑOL, CHILDREN, and DVD & VIDEO. A secondary navigation bar includes INTERNATIONAL, NEW RELEASES, TOP SELLERS, and TODAY. A search bar is present with a dropdown menu set to "All Products" and a search query "the art of computer scienc". A promotional banner for "FALL FAVORITES" features books by Sidney Sheldon and Maeve Binchy. A sidebar on the left contains a "SEARCH" section with a "GO" button and a "BROWSE" section with a "Book Browser" link. A "FASTER DELIVERY" section highlights "Next Day" and "Same Day" services. A "WEB SEARCH" section is powered by A9.com. A "BROWSE" section lists featured stores like Apparel & Accessories, Electronics, and Jewelry & Watches. An "Amazon.com Exclusive" section promotes the book "AMERICA" with a 32% off discount. A "New and Future Releases" section features "Logitech MX1000 Laser Cordless Mouse". The footer shows "Page loaded."

Location: <http://www.amazon.com/exec/obidos/subst/home/home.html/102-6661912-5275313>

**BARNES & NOBLE.com**  
www.bn.com

**FREE SHIPPING**  
When you spend \$25 or more  
SEE DETAILS

Shop in **Jewelry & Watches** | **amazon.com** | **VIEW CART**

WELCOME | YOUR STORE | BOOKS | APPAREL & ACCESSORIES | ELECTRONICS | TOY & GAME

INTERNATIONAL | NEW RELEASES | TOP SELLERS | TODAY'S DEALS

Hello, Sign in to get personalized recommendations. New

**Save Up to 30% - Top 100 CDs.**

**SEARCH**  
Know what you're looking for?  
Search books, music, and video  
Bookstore  
the art of computer scienc **GO**

**BROWSE**  
Need help finding the right book?  
Browse All Subjects with Book Browser

**FALL FAVORITES**  
Order the Season's Biggest Titles Now

BLOCKBUSTERS, NONFICTION, KIDS' BOOKS, DVD AND VIDEO

**FASTER DELIVERY**  
Next Day service to many East Coast locations  
Same Day to Manhattan, NY  
DELIVERY IS FREE When you spend \$25 or more  
SEE DETAILS

**WEB SEARCH**  
Powered by A9.com

**BROWSE**  
Add favorites  
Featured Stores  
\* Apparel & Accessories  
\* Electronics  
\* Jewelry & Watches  
New Stores  
In Beta (What's this?)  
\* Musical Instruments  
\* Gourmet Food  
\* Health & Personal Care  
\* Beauty  
\* Sports & Outdoors  
Books, Music, DVD  
\* Books

**An Amazon.com Exclusive**  
**AMERICA**  
Sure, we could write a pithy 50-word blurb about *American Idol*, but it's much easier--and fun--to let the show's host, Simon Cowell, tell you all about his irreverent new book, *American Idol: The Book*.  
Amazon.com Exclusive: *Join the Show* with *American Idol: The Book* 32% off

**It's a Wrap: HBO's Mr. Show on DVD**  
Pre-order the complete series of *Mr. Show* three-pack collection, or buy *Mr. Show: The Movie* à la carte. Either way, you save 30% off.

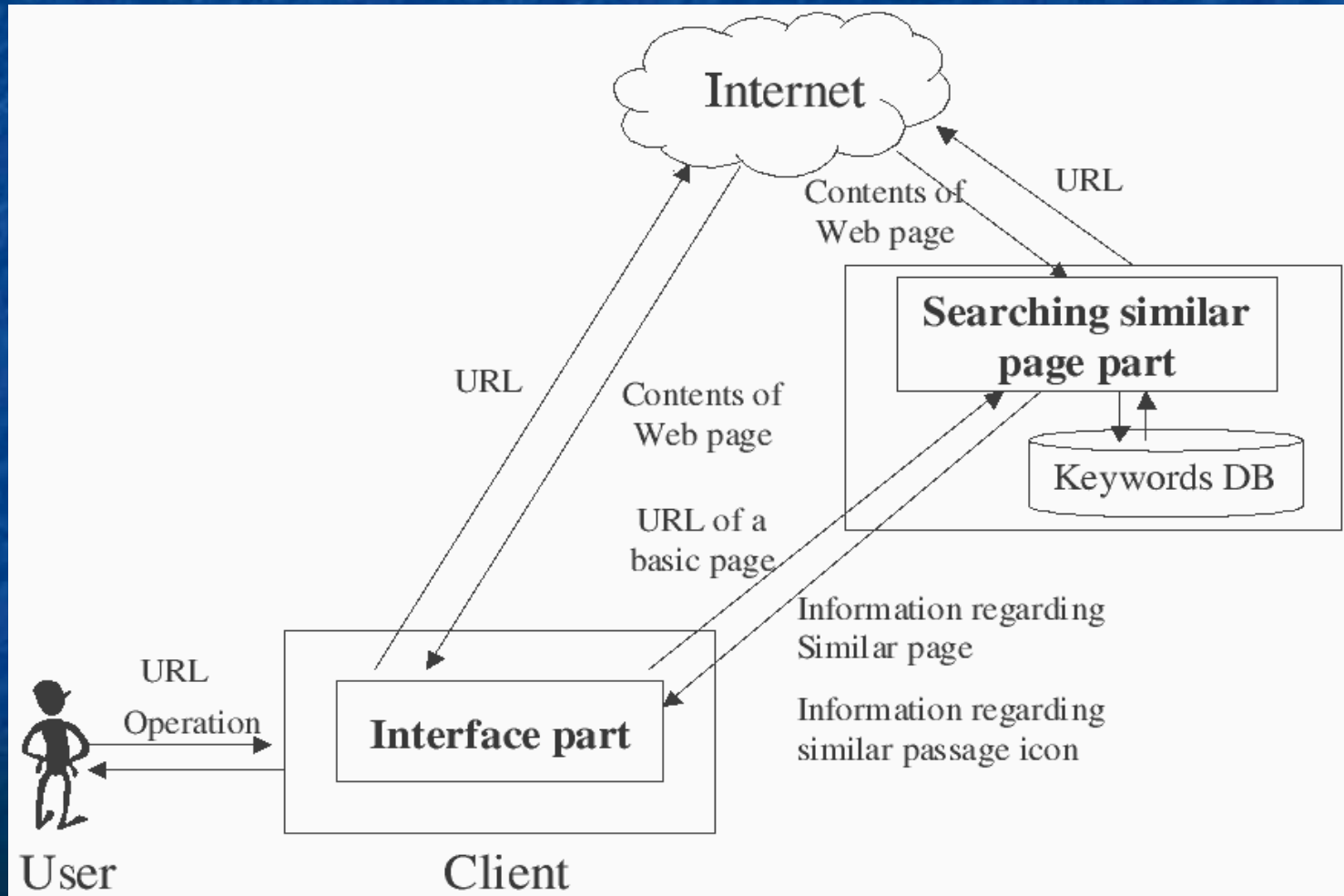
**New and Future Releases**  
**Logitech MX1000 Laser Cordless Mouse**  
From the Manufacturer

Page loaded.

# Featuring

- Automatic content-synchronization
- Linked navigation of comparison sites
- Common word highlighting on selection

# System Architecture



# 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  $\longrightarrow$  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