Re-Searching the Web for Language Professionals

William H. Fletcher
United States Naval Academy
CALICO Orlando, FL
15-17 March 2001
Information-Age Mindset
The Changes in Students and Implications for Higher Education

- Doing Rather Than Knowing
  - Short half-life of information makes dealing with complex information more important than accumulated knowledge

- Nintendo over Logic
  - Learning by Trial-and-Error
  - “Don’t bother me with the instructions!”

- Multitasking Way of Life
  - Capturing individual’s attention among greatest challenges of this age
Information-Age Mindset (2)

- Zero Tolerance for Delays
  - I want my information when *I* want it – either right now, or else later, when I am ready for it (email!) or when it is relevant

- Consumer / Creator Distinction Blurred
  - Disregard for intellectual property as no boundary perceived between the creator and user of information (cut-and-paste, download MP3s instead of buying CDs)
Challenges of Information Age

- InfoGlut: “Info, info everywhere, I’ve got no time to think!”
- Students must learn to be *transformers*, not mere *transducers* of information
- Intellectual engagement with information more important than ever
- Information and Knowledge Management Skills keys to success
Size of the Web

- Now 2 billion webpages directly accessible via public links (Cyberveillance projection July 2000)
- Several times that number in “invisible web” and restricted sites
- Estimated 80%-95% content in English, but...
- Since mid 2000, non-Anglophones outnumber English speakers online
- Former group increasingly has smaller percentage fluent in English
- For many countries and languages, data are simply unavailable
Search Purposes

- General users typically seek...
  - a specific site
  - any well-stocked site meeting their needs
- Scholarly searchers must examine and evaluate a range of sites to identify the most relevant and reliable resources
- Educators want to foster similar online research behavior in their students
Typical Search Behaviors

- Marked preference for directories with pre-selected links organized by topic over full-text search engines
- Simple queries – single word or phrase – predominate (80%-90%)
- 10%-25% of attempted complex queries (Boolean operators, bracketing) are ill-formed
- Users tend to work in a single window, calling up one document at a time, then returning to search engine for another link
Typical Search Outcomes

- Users follow up first few links only, and settle on a page after browsing from these links
- Usual outcome is *a* match, not *best* match
InfoNautics

- **F**ind relevant information efficiently
- **A**ssess its reliability
- **I**ntegrate it into personal “knowledge base”
- **R**etrieve and use it when needed
Searching Stages

- Presearch
- Search
- Research, Re-Search, Critical Evaluation
- Intellectual Interaction
Presearch

- Paper and pencil brainstorming
  - Paper-based, since a computer “demands” action, not reflection
- List words, phrases and variants to seek
- Determine logical relationships and groupings (Boolean operators, brackets)
- Identify criteria for inclusion / exclusion
Search

- Quick search to refine query
- KWiCFinder / WebKWIC search to locate sources
- Follow up search results to find most useful information on each site
- My assignments require sources from 3-5 different sites
Research / Re-Search and Critical Evaluation

- Paper-based evaluation
- Evidence for reliability / possible bias
  - Nature of site, author’s credentials
  - Research methodology, information sources
- If necessary, follow up sources cited to identify primary sources and methodology for analysis
Intellectual Interaction

- Paper-based exercise to relate online content to prior knowledge and conceptions.
- What did this source...
  - confirm that you already knew?
  - add to your knowledge about this subject?
  - reveal about connections among ideas?
  - contribute to this assignment?
- Why would you (not) recommend it to someone else researching this topic?
How Do I Use the Web for Instruction and Research?

- Eloquent examples
- Authentic texts
- Verify current / possible usage
- Locate relevant online sources references for research projects
Webidence is Evidence, not Conclusive Proof
Why KWICFinder?

- Automate process of search and retrieval
- Expedite evaluation of webpages
- Provide specific enhancements for foreign language users and linguists
- Encourage students and colleagues to take full advantage of online resources
Why AltaVista?

- *All* words are indexed, including "stopwords"
- Distinguishes case and "special characters"
- Supports Boolean operators, bracketing, and wildcards
- True world-wide coverage, with search by language
- No limits to length or complexity of the query
- Literal text search, without "second-guessing"
How Does KWiCFinder Complement AltaVista?

- Intuitive input for foreign characters, bracketing, operators, dates
- Tamecards, restricted wildcards, sic
- Inclusion / exclusion criteria
- Automatic search and retrieval in the background with KWiC abstracts
How Does XML Enhance KWicFinder?

- Search results become a dynamic database which can be manipulated:
  - categorize, annotate, delete, combine searches, citations and documents
- Free off-the-shelf tools permit restyling of reports by developer or end-user
  - Layouts
  - Languages
  - Data format
Why WebKWiC?

- Original hope: cross-platform, cross-browser solution
- Minimal entry threshold: small download of HTML pages + JavaScript
- Support for non-Western European languages
Why Google?

- Link popularity ranking puts relevant sites at or near top of list
- Straightforward approach to Advanced Search ("implicit Booleans") easy to learn, thus most likely to be used by students independently
- Largest number of pages analyzed
- Matching pages always* available in cache with KWIC markup
How Does WebKWiC Complement Google?

- Focuses and enhances interface for language learners
- Provides tools to navigate among citations and documents
- Simplifies management of multiple windows
Pleas(e)

Visit [http://KWiCFinder.com](http://KWiCFinder.com)

- Download and try KWiCFinder and WebKWiC
- View bibliography as well as this and related presentations
- Use these tools with your students
- Give me critical feedback and suggestions