

## Between the lines - the holes in the 'net

ECIS Annual Conference  
Berlin, 21-24 October 2002

**John Royce**  
Library Director : Robert College of Istanbul

Between the lines - the holes in the 'net.

---

---

---

---

---

---

---

---



## The big MYTHtakes!

- It's all there, it's up-to-date and it's all good stuff
- It's full-text and it's free
- It can easily be found
- Anyone can use a search engine
- Everyone's an expert

after Sherman and Price

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## The Internet

- telnet, ftp, gopher, lynx
- email
- Newsgroups & listservs
- Chat rooms
- World Wide Web

*and much more*

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## The WEB

- The Web – *that* big – and *that* fast!!!
- Distinguish between documents
  - posted on the Web
  - retrieved through the Web
- Much of the Web **cannot** be indexed
- Much of the Web **can** be indexed – but it isn't...

Between the lines - the holes in the 'net.

---

---

---

---

---

---

---

---



## Searching the web

- Known sites/ pages
- Intuitive guesswork
- Directories
- Search engines
- Metasearchers

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## Directories

- Work by subjects – classification scheme
- *Your* classification may not match the classifier's classification
- **Slow, labor intensive; cannot keep up with the growth of the internet**

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## On the other hand, search engines

- Searches are automated
- Many search **only** the web
- Work on *words*, not subjects
- Bots follow hyperlinks
- Different search engines give different results

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## Words, words, words

- Word frequency (care: sales and sex sites)
- Word frequency in opening lines or paragraph (location/ frequency)
- Title
- Metatags
- Top page only
- Whole site search
- ...and [the links](#)

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## Search engines: The naming of parts

- The bot
  - Spiders
  - Consenting adults
- The index
  - What gets indexed?
- The index search software
  - The algorithm
  - The hits and their display

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## Double or Quit?

- Different search engines give different results
- “The amount of overlap between the major web search engines remains surprisingly small.” (Greg Notess, 1998.)
  - Duplicate hits v Unique hits
- Tip: use more than one search engine

Between the lines - the holes in the 'net.

---

---

---

---

---

---

---

---



## Metasearchers

- Search search engines
- Some aim to eliminate duplicate hits
- Like search engines, some are more informative than others

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## What most search engines won't find

(though *some* are getting better)

- Sites requiring query entry
- Sites requiring password entry
- Sites with “**Keep Out**” signs
- Words as .gifs (eg. company logos)
- .pdf files, .doc files, .ppt files, etc
- Multimedia files
  - etc etc etc

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## The Holes in the 'Net

➤ “The Invisible Web consists of useful material found “on the web” that general search tools cannot or will not crawl / index / make searchable.”  
(Sherman and Price.)

- The opaque web
- The private web
- The proprietary web
- The truly invisible web

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## Caught in the Invisible Web

- Interactive data
- Directories
  - often compiled by experts
- Databases
  - FAQs, shopping catalogs, science and medical databases, discussion databases, phone and address databases, product reviews etc
- Archives
  - newspapers and journals, listservs
- Public and government information
- Reference tools
  - dictionaries, encyclopedias, thesauri, etc.

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## The Invisible Web

(BrightPlanet survey, as summarised by Gord Hotchkiss)

- Invisible Web 400 to 550 times bigger than the surface Web.
  - 550x information, 550x documents
- More than 100,000 Invisible Web sites.
- Growing faster than the surface Web.
- Invisible Web sites are narrower with deeper content than surface Web sites.
- More than half Invisible Web content is in topic specific databases.
- More than 95% Invisible Web content is publicly and freely available.

Between the lines - the holes in the 'net.

---

---

---

---

---

---

---

---



## Intelligent Searching

- Know WHERE to seek as well as HOW to seek.
- Know and use the Advanced Features of search engines.
- Use specialized search engines and directories.
  - topic specific / country specific
- Use invisible web directories.
- Get to know key sites **before** you need them.

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## Databases, such as...

<ul style="list-style-type: none"> <li>➤ Fee-paying           <ul style="list-style-type: none"> <li>➤ Dialog</li> <li>➤ Lexis-Nexis</li> <li>➤ Ebsco Ultra Host</li> <li>➤ Sirs Knowledge Source</li> <li>➤ Questia</li> <li>➤ Northern Light Special Collections</li> <li>➤ Ingenta</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>➤ Free           <ul style="list-style-type: none"> <li>➤ ERIC</li> <li>➤ NASA</li> <li>➤ FindArticles</li> <li>➤ MagPortal</li> <li>➤ FirstGov</li> <li>➤ OpenGov</li> <li>➤ Library of Congress</li> <li>➤ xrefer.com</li> </ul> </li> </ul>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Between the lines - the holes in the 'net.

---

---

---

---

---

---

---

---



## The tools *are* getting better

**BUT there are still:**  
**no quick, easy answers,**  
**no magic solutions,**  
**no fast fixes.**

Between the lines - the holes in the 'net.

---

---

---

---

---

---

---

---

## Analogy (1)

### The Library <<<>>> The Internet

- To find information in a library, use the catalog/ OPAC to find one or more books to look at.
- **The catalog does not hold the answer.**
- Go to the book/s and seek the information, perhaps using one book as a springboard to another.

- To find information on the internet, use a directory or a search engine to find one or more pages or sites to look at.
- **The search tool does not hold the answer.**
- Go to the sites and seek the information, perhaps using one site as a springboard to another.

Between the lines - the holes in the 'net.

---

---

---

---

---

---

---

---

---

---

## Analogy (2)

### The Library <<<>>> The Internet

- Sometimes the information is in a journal or other resource, and is not indexed in the library catalog.
- A different kind of index, and sometimes intuition or educated guesses, might get you there.
- **Finding information is rarely a one-step process.**

- Sometimes the information is in a database or other resource, and is not indexed by a search engine.
- A different kind of index, and sometimes intuition or educated guesses, might get you there.
- **Finding information is rarely a one-step process.**

Between the lines - the holes in the 'net.

---

---

---

---

---

---

---

---

---

---

## Strategic Imperatives

- Know what you're looking for, the better to know when you have found it!
- Consider: **where** are you most likely to find the information you need?
- Stop searching once you have found what you're looking for (but check for accuracy, currency etc...)
- Know when to give up on a site, a search strategy, a search

Between the lines - the holes in the 'net.

---

---

---

---

---


---

---

---

---

---



## Reminder!

### What's NOT on the web

- Books
  - Main exceptions: some public domain texts, e-books (pay per view?)
- Journals
  - Main exceptions: parts of current edition; pay-for databases – only recent & may not keep text forever.....
- Other media, other multimedia
- *Selection criteria / organization / history / stability / depth*

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## Don't stop thinking!

- The need for traditional tools remains
- Material on the net often leads to a lead to a lead to the answer, sometimes that lead is a person
- **Health warning!** Visible OR invisible: the material may still be unreliable! Don't abandon those critical Higher Order Thinking Skills!!

Between the lines - the holes in the 'net.

---

---

---


---

---

---

---

---



## Between the lines - the holes in the 'net

ECIS Annual Conference  
Berlin, 21-24 October 2002

**John Royce**  
Library Director : Robert College of Istanbul  
<http://vm.robcol.k12.tr/~jroyce/workshops>  
email: [jroyce@robcol.k12.tr](mailto:jroyce@robcol.k12.tr)

Between the lines - the holes in the 'net.

---

---

---

---

---

---

---

---