INFSCI 2140
Information Storage and Retrieval
Lecture 8: Alternative Retrieval Techniques

Peter Brusilovsky
http://www2.sis.pitt.edu/~peterb/2140-51

Ad-hoc IR in text-oriented DS

- The context (L1)
- Querying and matching (L2,L3)
- How to evaluate results (L4)
- How it all works internally (L5,L7)
- Better search and presentation taking user unto account: RF, QE, UM (L6)
- Better organization and visualization of search results (L10)
- What else?

A Broader view of the course

- Classic Information Retrieval
- Improving Classic Information Retrieval
- What else beyond Classic IR?
  - Alternative IR (L8)
  - Multimedia IR (L8)
  - Visualization (L10)
- Newest trends
  - User Modeling for IR (L6, L8, L10)
  - Web IR (L9)

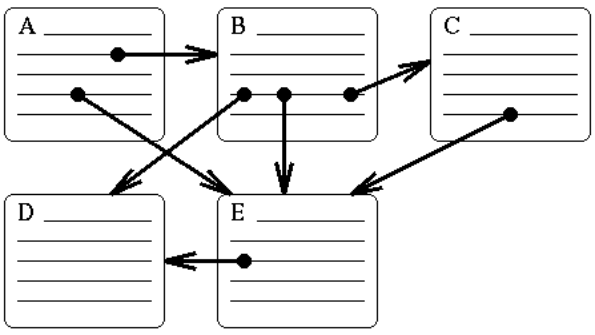
Alternative techniques?

- ~~Information Filtering~~
- ~~Natural Language Processing~~
- Hypertext: Links and browsing
- Citation Processing
- Adaptive hypermedia / navigation support
- Implicit queries (recommenders)
- Multimedia IR
- Dynamic queries (L10)

Hypertext: Definition

- Shneiderman [SK89]:
  - "a database that has active cross-references and allows the reader to "jump" to other parts of the database as desired".
- A hypertext is a database.
- The typical user action is a jump between parts of the database.
- Apart from pieces of information, called *nodes*, the database contains links between related nodes

Hypertext

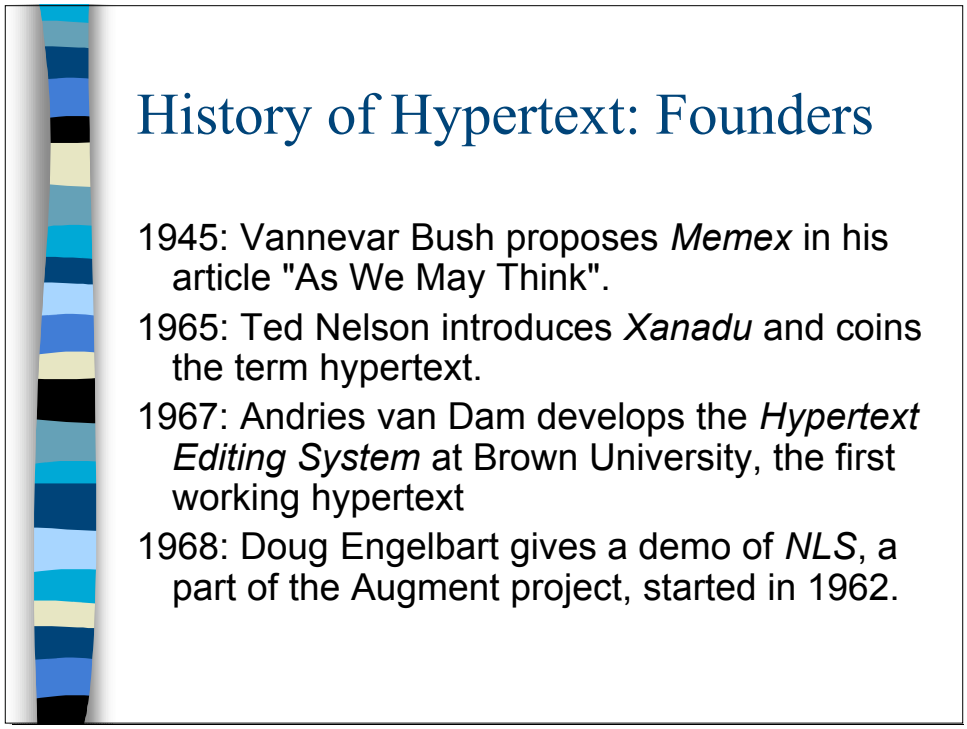


Hypertext: Anatomy

- Nodes
  - scrolling (WWW “pages”)
  - fixed size (HyperCard)
  - several pages
- Anchors (hot world, hot spot)
- Links
  - Anchor to node
  - Source anchor to destination anchor (Intermedia)

Hypermedia

- J. McDaid [McDaid-91]:
  - Hypermedia is ... an extension of Nelson's earlier coinage, "hypertext" (for non-sequential writing), hypermedia implies linking and navigation through material stored in many media: text, graphics, sound, music, video, etc.



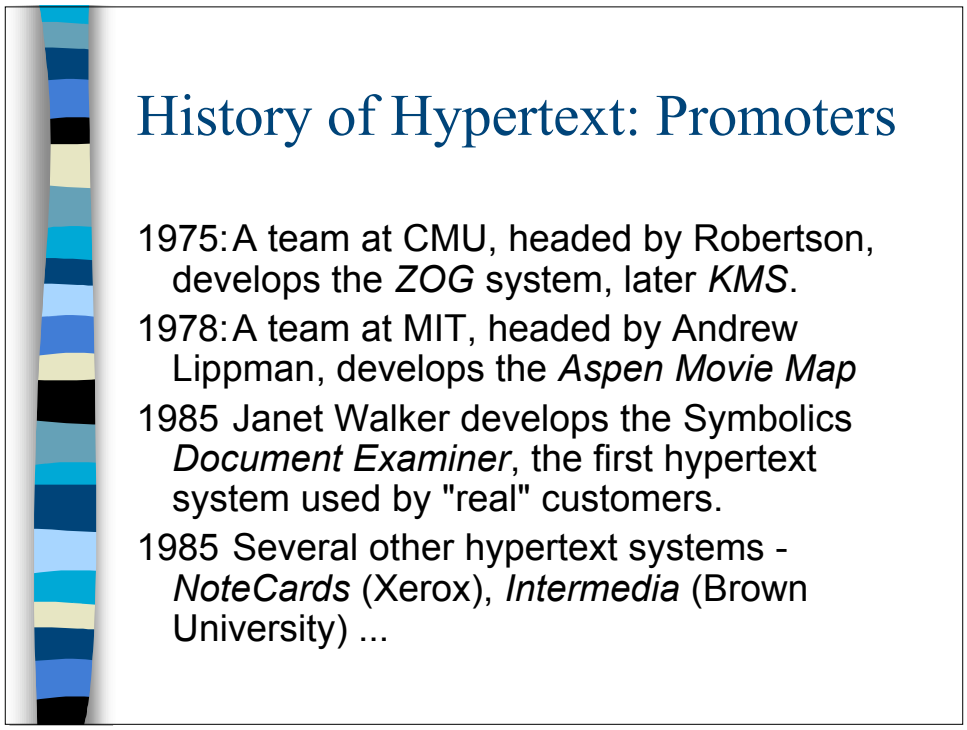
History of Hypertext: Founders

1945: Vannevar Bush proposes *Memex* in his article "As We May Think".

1965: Ted Nelson introduces *Xanadu* and coins the term hypertext.

1967: Andries van Dam develops the *Hypertext Editing System* at Brown University, the first working hypertext

1968: Doug Engelbart gives a demo of *NLS*, a part of the Augment project, started in 1962.



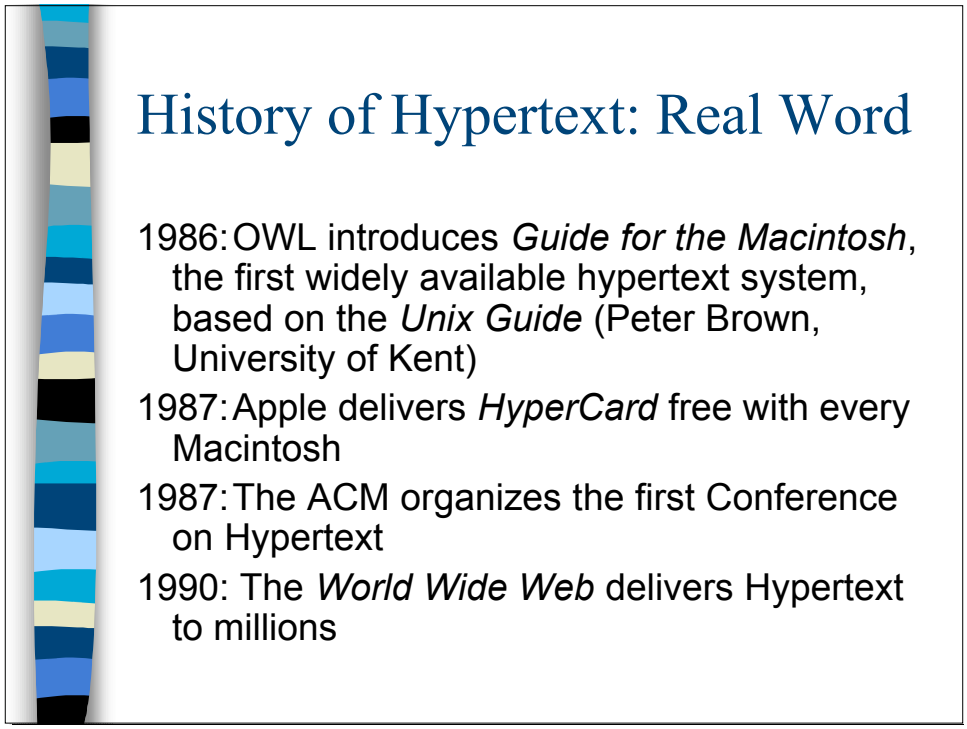
History of Hypertext: Promoters

1975: A team at CMU, headed by Robertson, develops the *ZOG* system, later *KMS*.

1978: A team at MIT, headed by Andrew Lippman, develops the *Aspen Movie Map*

1985 Janet Walker develops the Symbolics *Document Examiner*, the first hypertext system used by "real" customers.

1985 Several other hypertext systems - *NoteCards* (Xerox), *Intermedia* (Brown University) ...



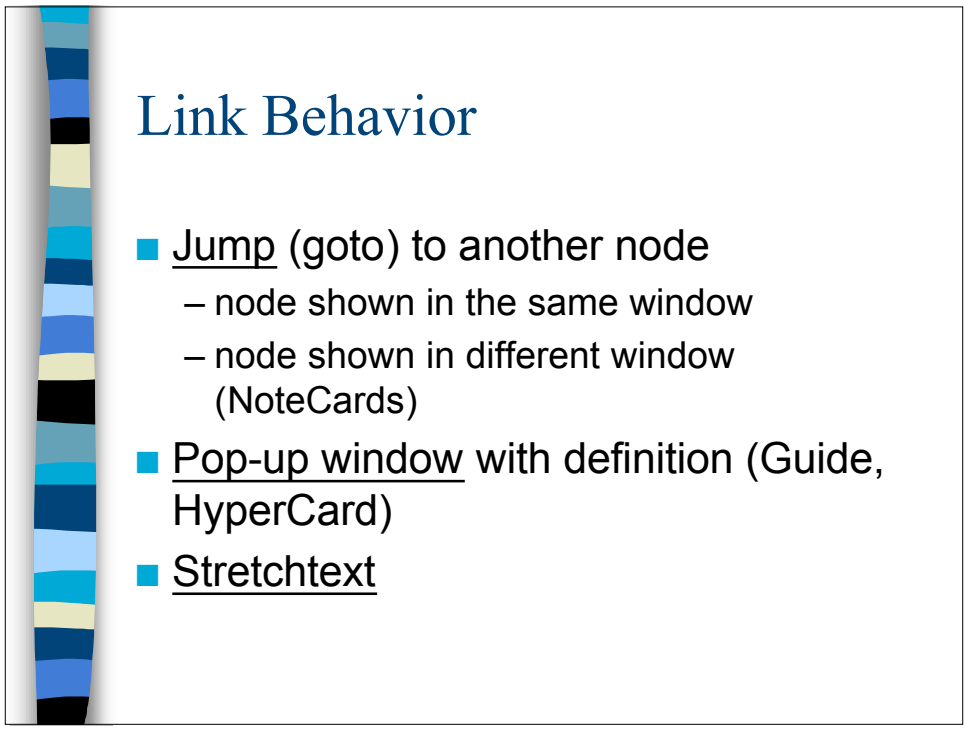
History of Hypertext: Real Word

1986: OWL introduces *Guide for the Macintosh*, the first widely available hypertext system, based on the *Unix Guide* (Peter Brown, University of Kent)

1987: Apple delivers *HyperCard* free with every Macintosh

1987: The ACM organizes the first Conference on Hypertext

1990: The *World Wide Web* delivers Hypertext to millions



Link Behavior

- Jump (goto) to another node
  - node shown in the same window
  - node shown in different window (NoteCards)
- Pop-up window with definition (Guide, HyperCard)
- Stretchtext

Another Node

Now jump back!

Link Behavior

- Jump (goto) to another node
  - node shown in the same window
  - node shown in different window (NoteCards)
- Pop-up w (Guide, HyperCa
- Stretchte

Pop-up Window activates when you click the mouse button and disappears when you release the button

Link Behavior

- Jump (goto) to another node
  - node shown in the same window
  - node shown in different window (NoteCards)
- Pop-up window with definition (Guide, HyperCard)
- Stretchtext

Link Behavior

- Jump (goto) to another node
  - node shown in the same window
  - node shown in different window (NoteCards)
- Pop-up window with definition (Guide, HyperCard)
- Stretchtext - expansion of text (Guide)
  - additional text is inserted when you click!
  - you may also collapse the expansion back

Hypertext Navigation

- Goals
  - Understanding “what’s there”
  - Find some information
  - Learn something
- Problems ⇨ Navigation aids
  - Where am I?
  - Where I can go from here?
  - Where I should go?

Navigation aids

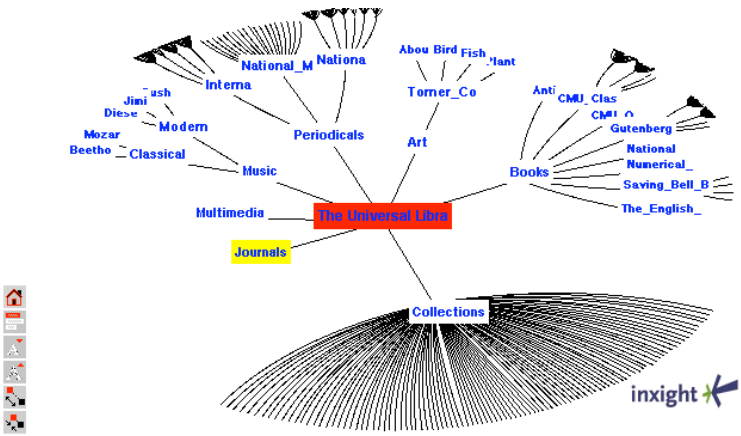
- Metaphors (museum, travel)
- Getting the whole picture
- Going back
- Avoiding loops
- Giving more information to decide
- Using structure (sequence, tree)
- Adaptive navigation support

Getting the whole picture...

... without getting lost

- Guided tours
- Landmarks
- Index pages
- Maps
  - Local map
  - Global map
- Fisheye views and hyperbolic trees

Hyperbolic Trees



More at: http://www.inxight.com/products/sdks/st/

Going back

- "Back" button
- History list
  - partial: path only (Netscape)
  - full: all visited nodes (IE)
- History tree
- Bookmarks

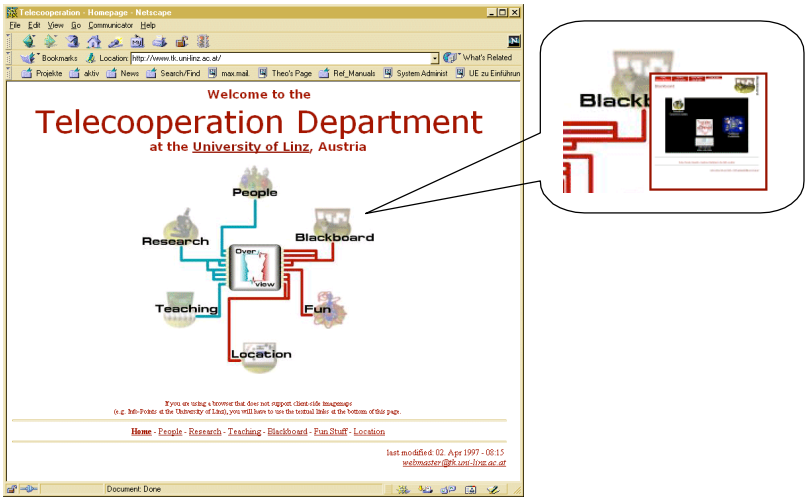
Where I should (should not) go?

- Loops and inefficient navigation
  - Unique anchors
  - Bread crumbs
  - Highlighting visited links
- More information to decide
  - Typed links
  - Sneak preview
- Adaptive navigation support

Typed Links

- Verbal typing:
  - subheading, prefix
- Spatial typing:
  - location, grouping
- Typed by visual cues
  - Font: size, style, type
  - Icon/bullet
  - Color of font, background, icon/bullet

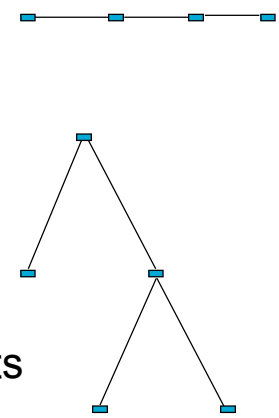
Sneak Previews



More at http://www8.org/w8-papers/4b-links/visual/visual.html

Using the structure

- Sequence (HyperCard)
  - forward
  - backward
  - home
- Tree
  - top
  - ancestor
  - descendants
  - siblings
- Table of contents

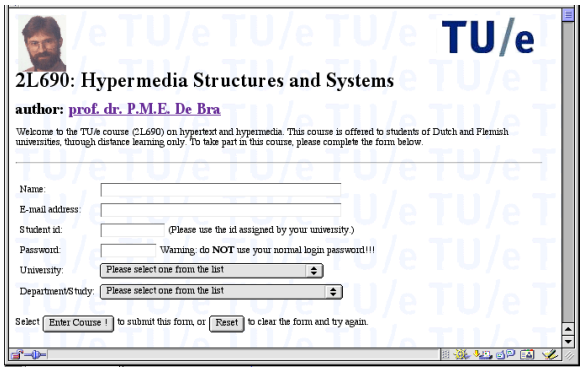


Navigation aids: summary

- backtracking
- bread crumbs
- highlighting links
- unique anchors
- sneak preview
- typed links
- history list
- history tree
- landmarks
- bookmarks
- maps
- indexes
- fish-eye views
- guided tours

Hypertext Course

- http://wwwis.win.tue.nl/2L690/
- Your Student id is your pitt e-mail



From IR to Hypertext

- Browsing offers an alternative and powerful way of information access
- What can we do if there are no links between documents and browsing is not possible?
- Similarity-based navigation
  - Document similarity
  - Access similarity

Citation processing

- Citations as extra information about a document
  - Co-citation as a similarity measure
  - Link between documents citing same source
- Citations as a measure of importance
- Citation-based navigation
- Examples: WebOfScience, CiteSeer
- Problem: no citation standard

What's in a citation?

- Authors
- Title
- Source data
  - Journal: Title, publisher
  - Book: Title, editor, publisher
  - Conference: Title, date, location
- Location in the source

Citation processing

- Is it only relevant to science papers?
- Where else similar technologies can be used?
-
-

Browsing the Web

- Can we support the user who is browsing?
- No query: need to know information need or user model
  - How to deduce that? ⇨ User modeling
- No “search results”: need to guide the user to good pages
  - How to guide? ⇨ Adaptive navigation support

Adaptive hypermedia: Why?

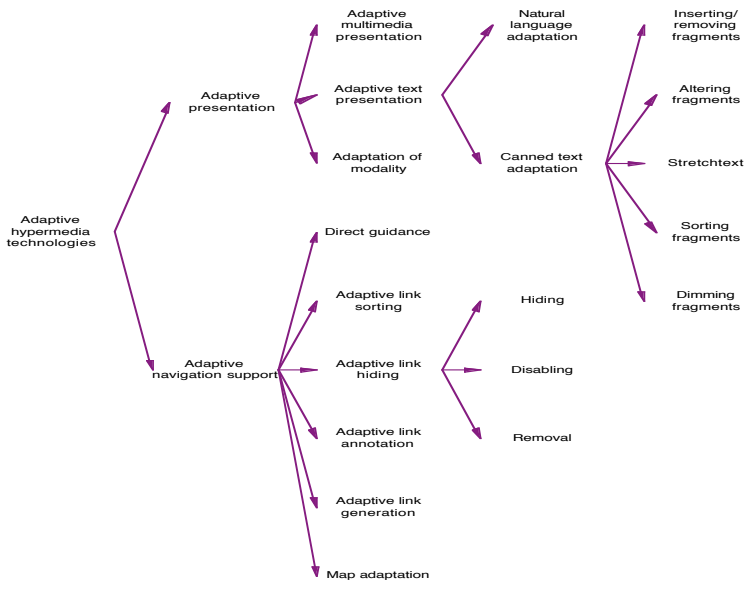
- ☑ Different people are different
- ☑ Individuals are different at different times
- ☑ "Lost in hyperspace"
- ⇨ Large variety of users
- ⇨ Variable characteristics of the users
- ⇨ Large hyperspace

Where it can be useful?

- Web-based education
  - ELM-ART, AHA!, KBS-Hyperbook, MANIC
- On-line information systems
  - PEBA-II, AHA!, AVANTI, SWAN, ELFI, ADAPTS
- E-commerce
  - Tellim, SETA, Adaptive Catalogs
- Virtual and real museums
  - ILEX, HYPERAUDIO, HIPS, Power, Marble Museum
- Information retrieval, filtering, recommendation
  - SmartGuide, Syskill & Webert, IfWeb, SiteIF, FAB, AIS

What can be adapted?

- Hypermedia = Pages + Links
- Adaptive presentation
  - content adaptation
- Adaptive navigation support
  - link adaptation



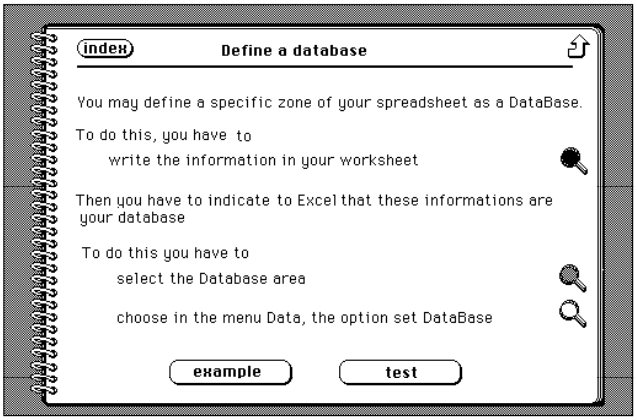
Adaptive navigation support: goals

- Guidance: Where I can go?
  - Local guidance ("next best")
  - Global guidance ("ultimate goal")
- Orientation: Where am I?
  - Local orientation support (local area)
  - Global orientation support (whole hyperspace)

Adaptive navigation support

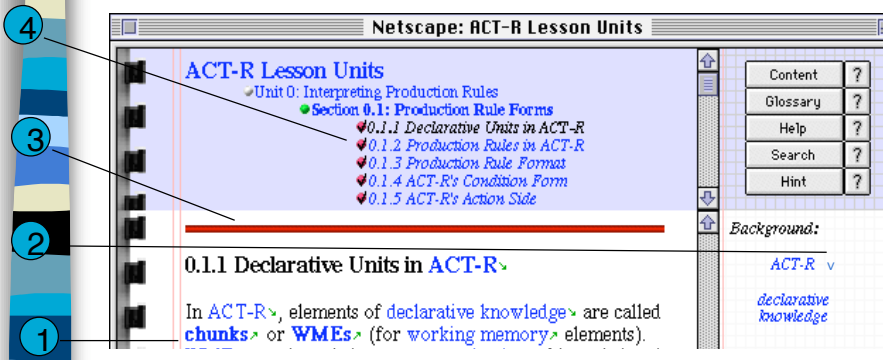
- Direct guidance (WebWatcher)
- Restricting access
  - Removing, disabling, hiding
- Sorting
- Annotation
- Generation
  - Similarity-based, interest-based
- Map adaptation techniques

Example: Adaptive annotation



Annotations for topic states in *Manuel Excell*: not seen (white lens) ; partially seen (grey lens) ; and completed (black lens)

Example: Adaptive annotation



1. Concept role
2. Concept state
3. Current section state
4. Linked sections state

What can be adapted: links

- Contextual links (“real hypertext”)
- Local non-contextual links
- Index pages
- Table of contents
- Links on local map
- Links on global map

Link types and technologies

	Direct guidance	Sorting	Hiding	Annotation	Map adaptation
Contextual links	OK		(disabling)	OK	
Non-contextual links	OK	OK	?	OK	
Table of contents	OK		?	OK	
Index	OK		?	OK	
Local map	OK		OK	OK	OK
Global map	OK		OK	OK	OK

Adaptive navigation support: evaluation

- Sorting
  - ➡HYPERFLEX, 1993
- Annotation (colors) and hiding
  - ➡ISIS-Tutor, 1995
- Annotation (icons)
  - ➡InterBook, 1997
- Hiding
  - ➡De Bra's course, 1997

Evaluation of sorting

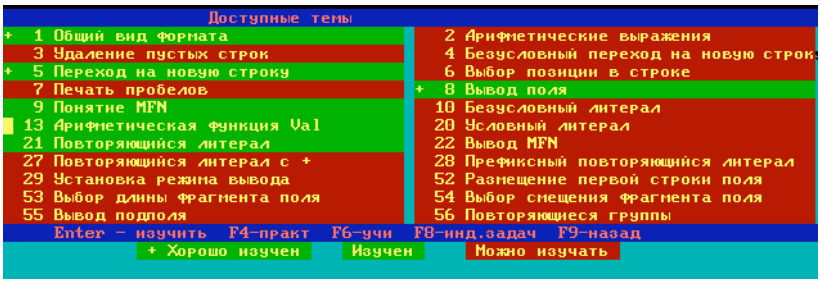
- HYPERFLEX: IR System
  - – adaptation to user search goal
  - – adaptation to "personal cognitive map"
- Number of visited nodes decreased (significant)
- Correctness increased (not significant)
- Goal adaptation is more effective
- No significant difference for time/topic

Annotation and hiding: ISIS-Tutor

- An adaptive tutorial for CDS/ISIS/M users
- Domain knowledge: concepts and constructs
- Hyperspace of learning material:
  - Description of concepts and constructs
  - Examples and problems indexed with concepts (could be used in an *exploratory environment*)
- Link annotation with colors and marks
- Removing links to “not relevant” pages

Sample index page (annotation)

Sample index page (hiding)



Results: performance

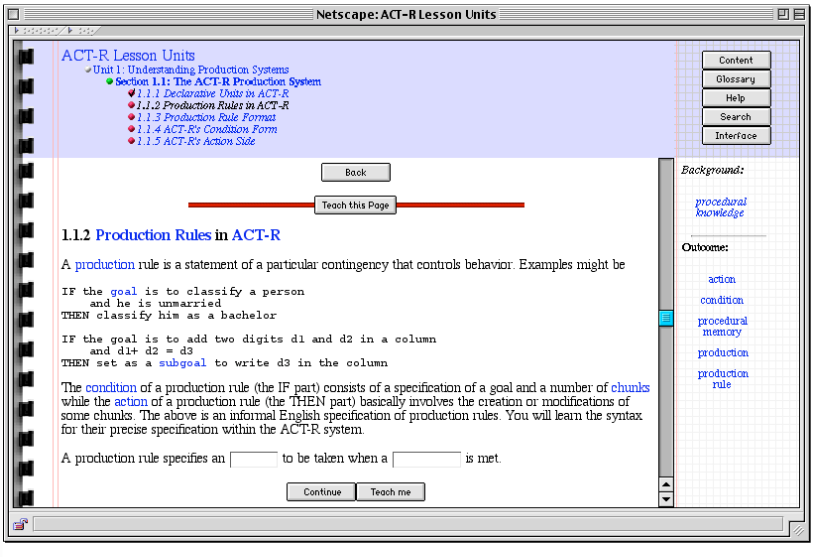
Group	Number of steps	Time (sec)	Concept repetitions	"Unforced" concept repetitions	Task repetitions
Non-adaptive	81.3	2196	17.3	11.2	6.2
Adaptive	65.2	1418	9.0	5.0	0.8
Restrictive	58.2	1785	8.9	4.8	0.4

Adaptive annotation makes navigation more efficient

InterBook: concept-indexed ET

- "Knowledge behind pages"
- Structured electronic textbook (a tree of "sections")
- Sections indexed by domain concepts
  - Outcome concepts
  - Background concepts
- Concepts are externalized as glossary entries
- Shows educational status of concepts and pages

Book view



Adaptive annotation can:

- Reduce navigation efforts
  - Results are not significant (variety of styles?)
- Reduce repetitive visits to learning pages
  - Significant - if applied properly
- Encourage non-sequential navigation
- Increase learning outcome
  - For those who is ready to follow and advice
- Make system more attractive for students

Implicit Query - Recommenders

- A user identifies 1 or several objects as being of interest
- The recommender system suggests matching objects from the DS
- Inverse paradigm - push vs. pull: the user is passive (not really querying), the system is active - recommending
- Shares a lot in common with IF

Parameters for recommenders

- Short term of long term “interest”
  - 1-2 “interesting” items in one session
  - many items over time
- Direct or indirect “rating”
  - Direct rating
  - “Looking at”: click, browse
  - Other ways to show interest:
  -
  -

Parameters for recommenders

- Content-based vs collaborative “filtering”
- Content based:
  - find items similar to the set of *interesting* by content
- Collaborative “filtering”
  - Find users who have similar opinion with you
  - What else these user consider *interesting*?

Case Study: Paper Recommender

- Suggest publication for several reasons
  - Similar to just downloaded paper
  - Cites just downloaded paper
  - Cited by just downloaded paper
  - Most frequently downloaded together
  - Added since last visit and cited by earlier downloaded paper
  - Most popular but not yet considered
- Doubled # downloads per visit!

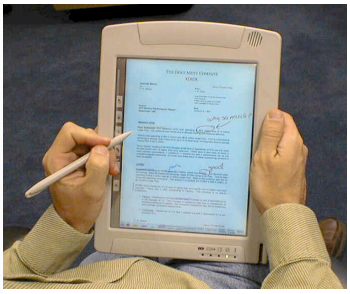
Adaptation in Paper Recommender

- What?
  - Learns relative importance of the reasons by watching agreements to suggestions
- When?
  - Clicked “suggest”, donwloaded paper, idle time
- Adaptability (user can specify)
  - Which reasons not to consider
  - Topics of interests
- URL:
  http://www.ics.uci.edu/~pazzani/Publications/Publications.html

Case Study: Paper Recommender

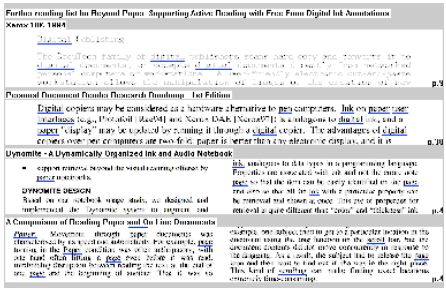
- How the users shows their interests?
- Long term or short term interests?
- Which technology is better - content-based or collaborative filtering?

Case Study: XLibris



http://www.fxpal.xerox.com/xlibris/

- The user reads the text and annotates it using a pen
- XLibris can generate *marginal links* and *further reading list*



Case Study: XLibris

- How the users shows their interests?
- Long term or short term interests?
- Which technology is better - content-based or collaborative filtering?

Case Study: WATSON

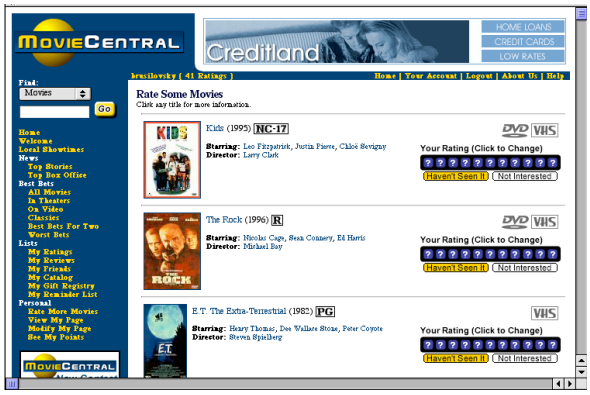
- WATSON system (Northwestern U)
  - The user types or work in GUI
  - The system observes his/her work
  - The recommendation window shows relevant resources
- Instant Queries

Case Study: WATSON

- How the users shows their interests?
- Long term or short term interests?
- Which technology is better - content-based or collaborative filtering?

Case Study: MovieCentral

- The users rate movies
- The system can suggest best bets
- Users keep rating movies while checking best bets



http://www.moviecentral.com

Alternative URL: http://www.movielens.umn.edu/

Case Study: MovieCentral

- How the users shows their interests?
- Long term or short term interests?
- Which technology is better - content-based or collaborative filtering?

Case Study: Amazon.com

- How the users shows their interests?
- Long term or short term interests?
- Which technology is better - content-based or collaborative filtering?

Web Recommenders vs. ANS

- Recommenders originate from filtering systems and use an old “search” approach, ANS originates from hypertext and focuses on navigation support
- Even advanced recommenders use simple 1-D “list of links” presentation, ANS use 1.5-D presentation
  - Power of a recommendation engine could be enhanced by power of a proper interface
- Modern AH systems require content knowledge, modern recommender systems can create/extract it.

A broader picture - adaptive IS

- Users work with an adaptive IS, the system attempts to develop a user model (knowledge, interests, goals) and assist each user adaptively
- There are different ways to model the user (knowledge, interests, collaborative, content-based…)
- There are different ways to use the knowledge to assist the user

Case Study: Knowledge Sea

- What the users are doing
- What kind of information the system is able to extract from watching the users and how
- How the system assist the users in the process of information access?

Multimedia IR

- Text as Image (Digital Libraries)
- Images
- Video
- Spoken documents
- Music
- Arbitrary sound

Text as Image (Digital Page Image)

- Text image vs. text
- Problems
  - From image to text - OCR vs manual
  - Page segmentation
  - Graphic extraction
- Examples
  - CORE, JSTOR, TULIP
  - Early Canadiana `http://www.canadiana.org/`

Image retrieval

- Application areas
  - Retrospective
    - Stock photos
  - Filtering
    - Satelite imaging
    - Law enforcement and immigration
- Main approaches
  - Concept-based retrieval
  - Content-based retrieval

Concept-based image retrieval

- Key: Concept-based indexing of images
  - Based on attributes extracted manually
  - Based on logical, high level features
- Systems for image indexing
  - ICONCLASS, A&AT, …
- What?
  - Time, location, content

Content-based image retrieval

- Key: Automatic indexing of images based on low-level features
  - Color
  - Texture
  - Shape
  - Spatial orientation and layout
  - Sketch

Examples - content based IR

- QBIC - IBM's Query By Image Content:
  `http://wwwqbic.almaden.ibm.org`
- MIT PhotoBook:
  `http://www.white.media.mit.edu/vismod/demos/photobook/`
- Virage: `http://www.virage.com`
- VisualSeek:
  `http://www.ctr.columbia.edu/~jrsmith/VisualSEEk`

Video and spoken language

- Problems:
  - Segmentation
  - Indexing
- Examples:
  - CMU Informedia:
    - http://www.informedia.cs.cmu.edu
  - Virage:
    - http://www.virage.com/services/ivq.html