CHI 2003 Tutorial

Information Foraging

Peter Pirolli and Stuart Card
PARC
3333 Coyote Hill Road
Palo Alto, CA 94304

Instructor Biographies

Peter Pirolli is a Principal Scientist in the User Interface Research area at PARC. He received his B.Sc. in Psychology and Anthropology from Trent University, Canada, where he conducted a published research related to eyewitness memory. He earned his M.S. and Ph.D. in Cognitive Psychology from Carnegie Mellon University where he developed computational models of students learning to program and helped develop an intelligent tutoring system for programming. He became a Professor in the School of Education at the University of California, Berkeley, where he was also Associate Director of the Cognitive Science Program. During that period he studied computational models of metacognition, intelligent tutoring systems, and design problem solving, and he became a Fellow of the National Academy of Education. He joined Xerox PARC in 1991, where he is engaged in studies of human-information interaction, information foraging theory, and the development of new user interface technologies.

Stuart Card is a Senior Research Fellow at PARC and head of the User Interface Research group. He has developed models in human-computer interaction, including GOMS and the Fitts's Law model of the mouse as well as new user interface te3chniques, such as Rooms and focus+context information visualization methods. His most recent book, *Readings in Information Visualization, Using Vision to Think*, was written and edited with J. Mackinlay and B. Shneiderman.

Objectives

Participants will learn techniques of information foraging analysis for characterizing human information-seeking behavior. The objective is that participants should be able by the end of the tutorial to perform analyses in information foraging.

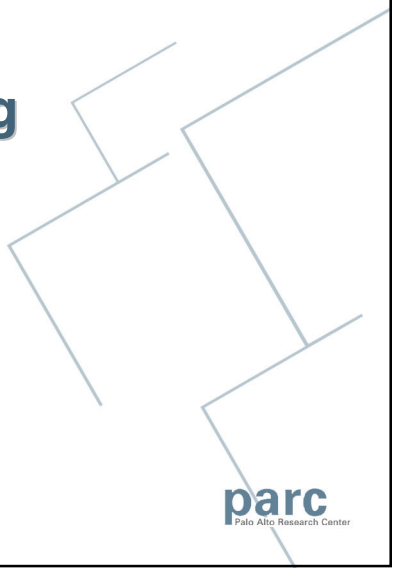
- Information foraging theory as a new method for analyzing information-intensive work.
- Models and empirical tools for analysis of adaptation to information environments cognitive mechanisms.
- Demos, hands-on exercises, and applications.
- Students take away resources to aid analysis and teaching.
- Emphasis on applications to Web and information system design.

Agenda

1. Problem and Rationale
   - Improving human-information interaction
   - Informavores; information as extended phenotype
   - The glut of information
   - Optimal foraging theory
2. Levels of Behavioral Theory
   - Time scales of cognition
   - Adaptation and proximal mechanisms
   - Search, pursuit, handling tradeoff
3. Instrumentation for information foraging analysis
   - Problems with conventional usability analysis
   - Protocol analysis and logs
   - Problem spaces and Web Behavior Graphs
   - Human-Information Interaction
4. Adaptation Level of analysis
   - Adaptationist approach (example: memory)
   - Optimal Foraging Theory examples
   - Exercise: plotting gain functions
   - Resources: Value of information; Resource costs and opportunity costs
   - Basic foraging model: Charnov's Theorem; application: Patches on the Web
   - Diet Selection model: all-or-none theorem; Application
5. External factors: Information ecology
   - Cost structure of information — Cost of Knowledge Characteristic
   - An information ecology: the World Wide Web
6. Proximal Mechanisms: Aggregate Behavior
   - Information scent model of aggregates of users
   - Example: WUFIS model using spreading activation
   - Phase shifts
7. Proximal Mechanisms: Dynamic Programming
   - Using dynamic programming to locate optimal paths
   - Exercise: Scatter/Gather state space
8. Proximal Mechanisms: Cognitive Models
   - Information scent
   - User-tracing method
   - ACT-IF and SNIF/ACT models
9. Information Farming (Enhanced Foraging)
   - Social/Cooperative Foraging
   - Sensemaking
10. Human-Information Interaction

Information Foraging

Peter Pirolli
Stuart K. Card

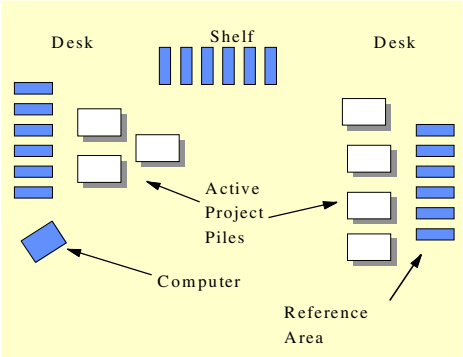


parc
Palo Alto Research Center

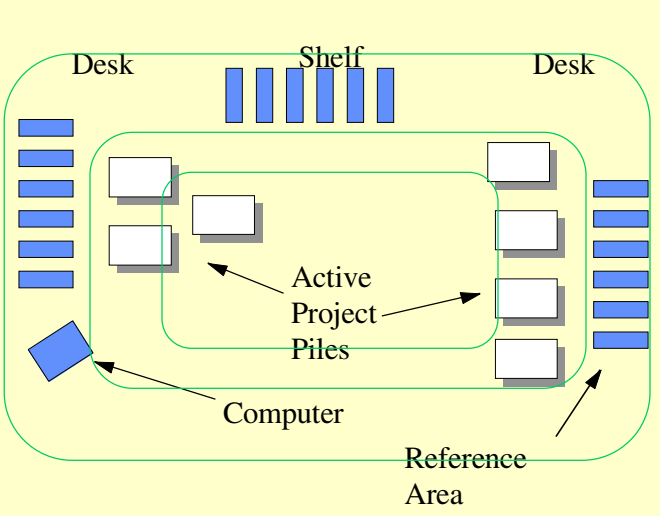
Exercise: Information Patches

- Patchy structure of the information environment
- Hypothetical analysis
- Patch Model & Charnov's Marginal Value Theorem

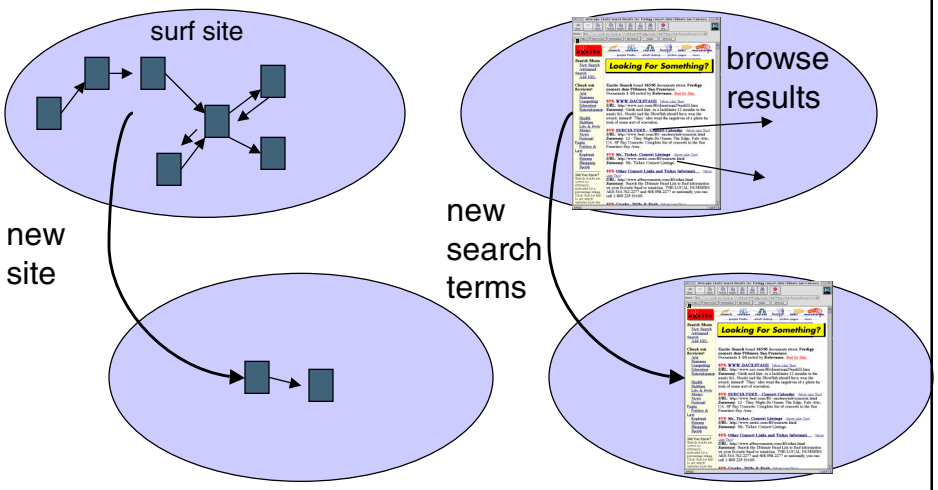
information patches



workspace cost structure

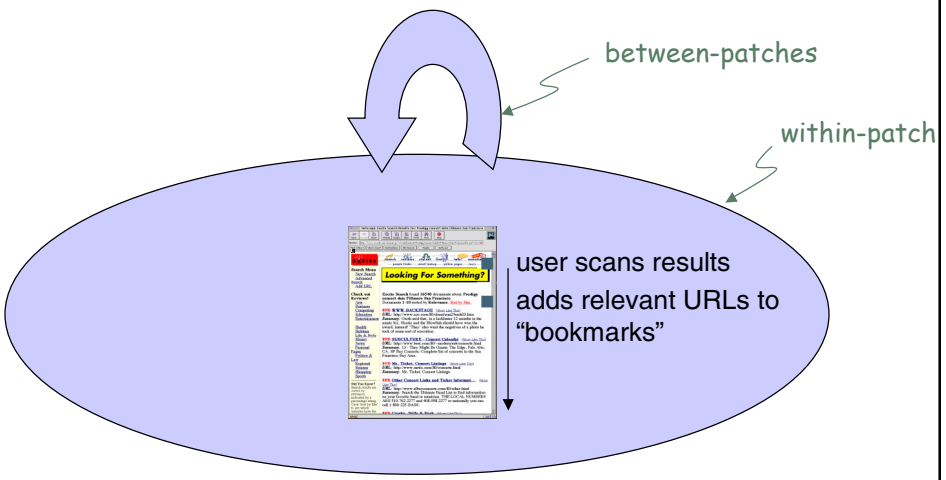


information patches

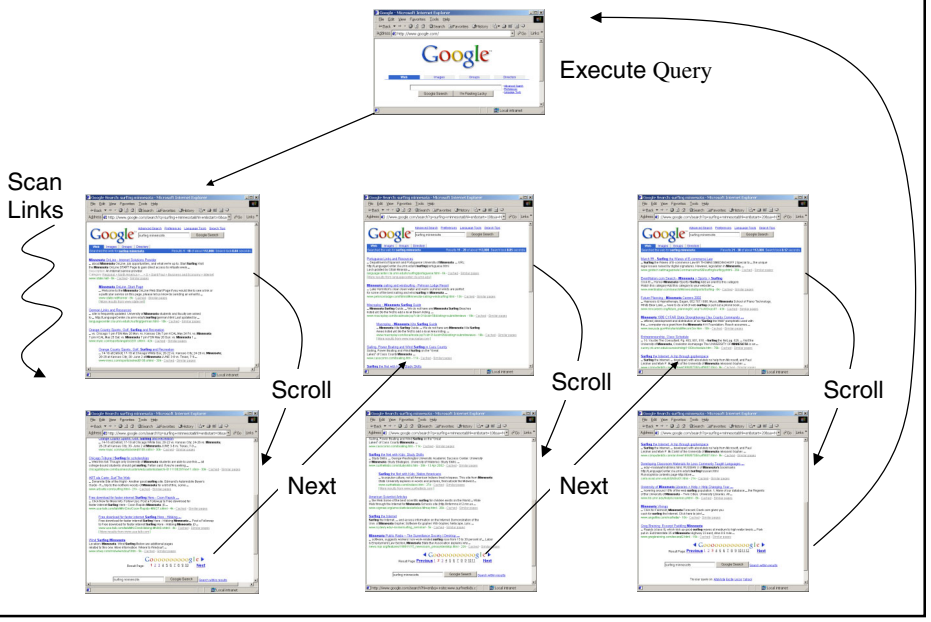


hypothetical example

- user submits query to a search engine

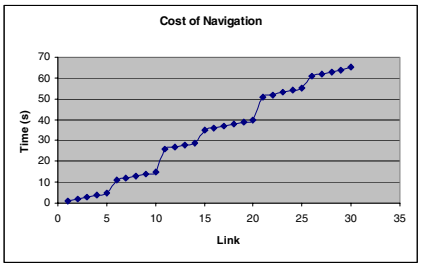


Cost Structure



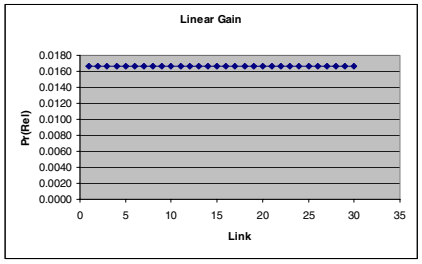
Hypothetical Costs & Assumptions

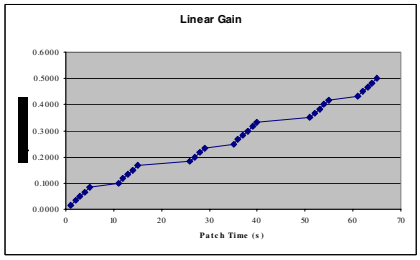
- Execute search = 10 s
- Scan link = 1 s
- Scroll = 5 s
- Next Page = 10 s
- 10 links/page
- 3 pages per search



Linear Gain Function

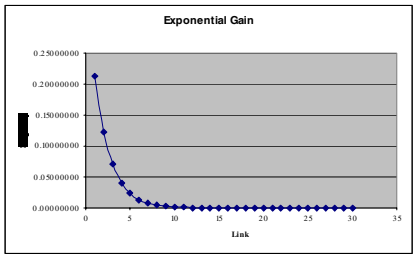
- .5 probability that any executed search will produce a relevant link
- "Poor" retrieval system presents links in uniformly random order

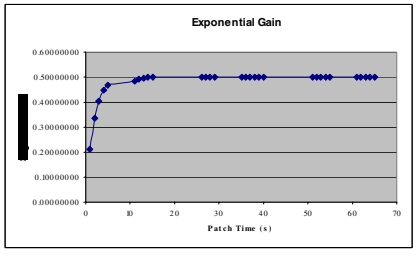




Exponential Gain Function

- .5 probability that any executed search will produce a relevant link
- "Good" retrieval system presents links in ranked manner





Questions
(for both linear & exp gain plots)

- Note: *Travel time* is the time it takes to get to the search engine home page and execute a query. It is assumed to be the cost of getting to a new page (10 s) plus the cost of executing a query (10 s).

1. Suppose our hypothetical forager adopts a policy of never spending more than 30 seconds of patch time + travel time on any given query.
   a. Find and mark the point on the curve that corresponds to this "30 second investment policy". What is the expected gain?
   b. The *overall rate of gain* over many queries will be the (expect gain)/(Patch time + Travel time). Plot a line from the intercept (0,0) to the point marked in 1.a. What is the slope of this line?
2. Repeat steps (1.a) and (1.b) above for the following "time investment" policies. In each case, what is the expected gain and the overall rate of gain at:
   a. 22 seconds
   b. 50 seconds
   c. 80 seconds
3. What "time investment policy" (in terms of number of seconds invested in browsing links each query) will maximize the overall rate of gain?