

Validating InfoScent™ Models with Large Scale User Studies

Ed H. Chi
(joint work with UIR, ASD)

User Interface Research Group
Palo Alto Research Center

2003-07

NPUC IBM Almaden



Outline

- Motivation and Background: InfoScent™
- Bloodhound Evaluation
- LumberJack Evaluation
- Conclusion

2003-07

NPUC IBM Almaden

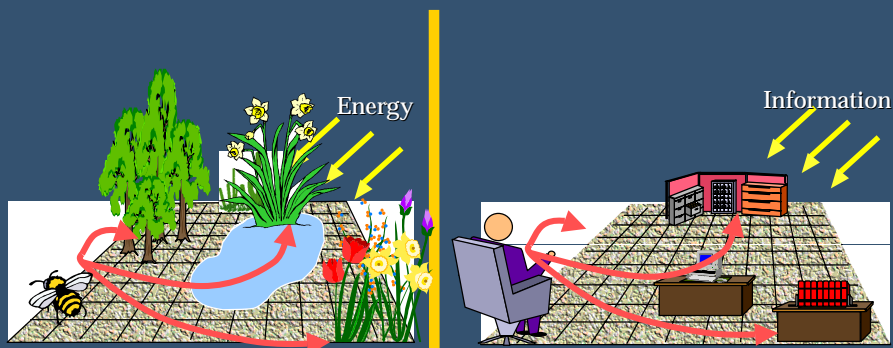
Copyright 2002 PARC



2

Theory: Information Foraging

Analogy to Optimal Foraging



2003-07

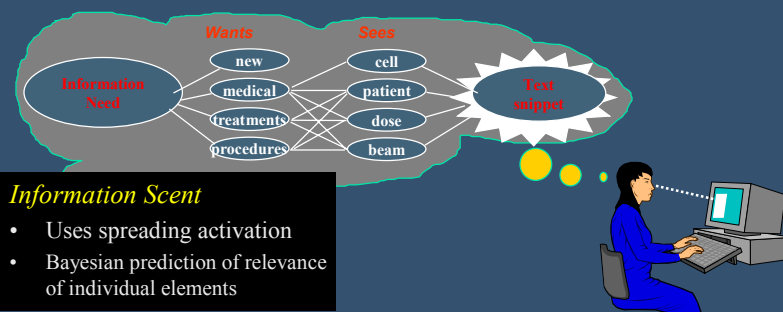
NPUC IBM Almaden

Copyright 2002 PARC

parc

3

User Modeling with Information Scent



2003-07

NPUC IBM Almaden

Copyright 2002 PARC

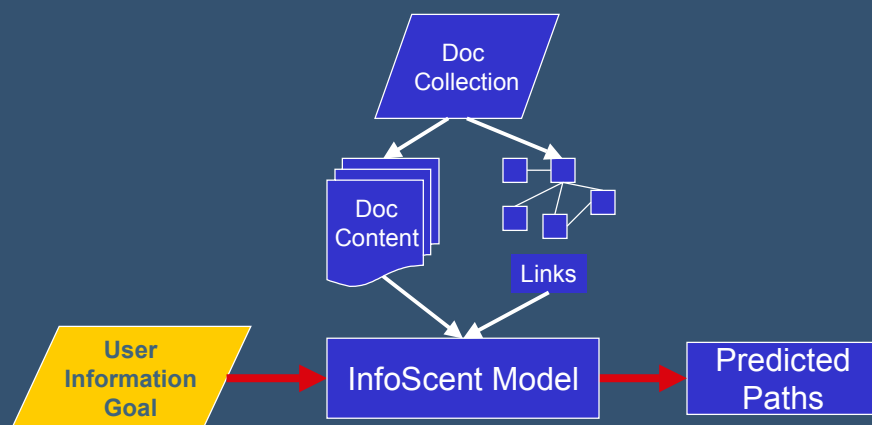
parc

4

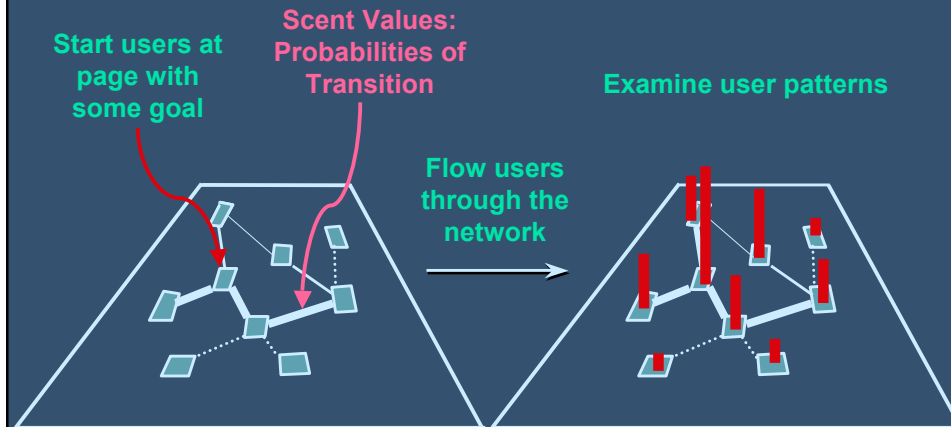
Modeling using InfoScent™

- Users have information goals, their surfing patterns are guided by information scent.
- Two questions
 - Given an information goal and a starting point
 - **Where do users go? (Behavior)**
 - Given some surfing pattern
 - **What is the user's goal? (Need)**

WUFIS: Web User Flow by Information Scent



How does it work?



2003-07

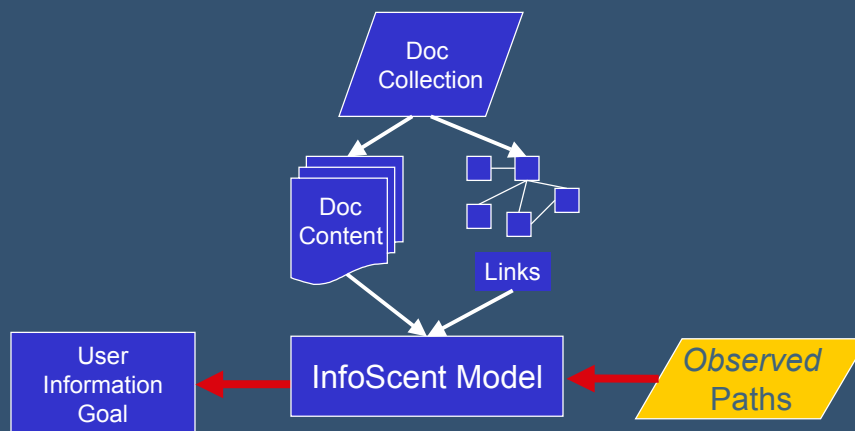
NPUC IBM Almaden

Copyright 2002 PARC

parc

7

IUNIS: Inferring User Need by Info Scent



2003-07

NPUC IBM Almaden

Copyright 2002 PARC

parc

8

Outline

- Motivation and Background: InfoScent™
- LumberJack Evaluation
- Bloodhound Evaluation
- Conclusion

LumberJack Motivation

What are users' information goals?

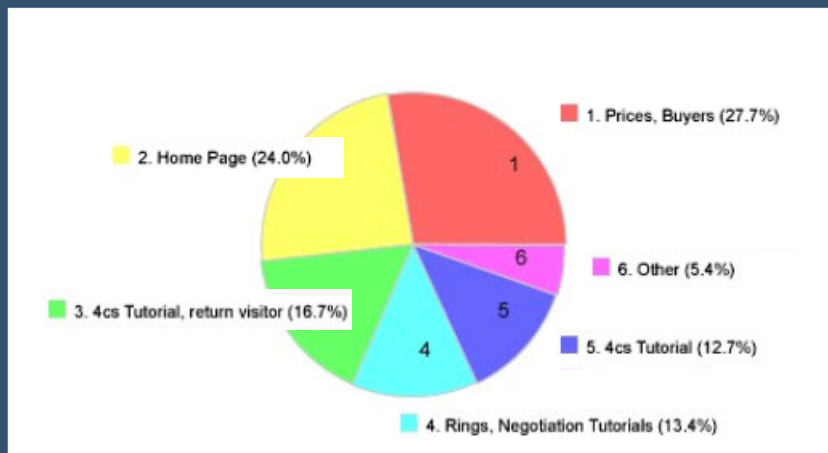


Understanding the composition of web user traffic.

Strategy: Use all available data to discover user goals. (Content, Usage, Topology)

Clustering Results

<http://www.diamondreview.com>



Users reached end of tutorial, had nowhere to go.

2003-07

NPUC IBM Almaden

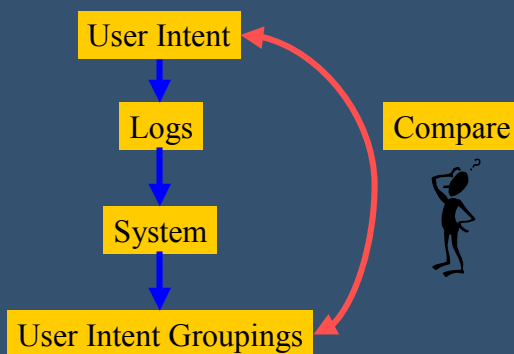
Copyright 2002 PARC

parc

11

LumberJack Evaluation

Does the system correctly infer user intentions?



Lesson: Easy to get user trace data, but hard to get user trace data with associated user goals!

2003-07

NPUC IBM Almaden

Copyright 2002 PARC

parc

12

User Study



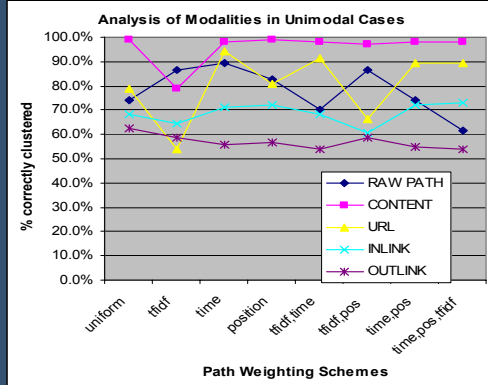
- Asked users to surf specific tasks on www.xerox.com
 - captured actions using the WebQuilt proxy logger [Hong01]
 - done at their leisure.
- 15 unique tasks:
 - Tasks developed after exploring xerox.com and reading user e-mail feedback
 - 5 task groups with 3 tasks per group.
 - Products, TechSupport, Supplies, Company Info, and Jobs
- Participation:
 - 21 users signed up, 18 went through, **104** usable sessions.

Results: 340 combinations of clustering schemes

	uniform	tfidf	time	pos	Tfidf, time	tfidf, pos	time, pos	tfidf, time, pos	Ave. rage	Std. Dev.
RAW PATH	74%	87%	89%	83%	70%	87%	74%	62%	78%	10%
CONTENT	99%	79%	98%	99%	98%	87%	98%	98%	96%	7%
URL	79%	54%	84%	81%	91%	66%	89%	89%	81%	14%
INLINK	68%	64%	71%	72%	68%	61%	72%	73%	69%	4%
OUTLINK	63%	59%	56%	57%	54%	59%	55%	54%	57%	3%
CU [0.75,0.25]	88%	78%	98%	99%	97%	81%	98%	93%	92%	8%
CU [0.5,0.5]	81%	78%	97%	87%	96%	80%	97%	92%	88%	8%
CU [0.25,0.75]	79%	79%	94%	87%	93%	78%	97%	92%	87%	8%
CI [0.75,0.25]	80%	94%	96%	97%	96%	95%	97%	98%	94%	6%
CI [0.5,0.5]	87%	94%	94%	94%	96%	95%	96%	98%	94%	3%
CI [0.25,0.75]	87%	88%	94%	93%	92%	95%	90%	93%	92%	3%
CUI [0.5,0.25,0.25]	83%	94%	94%	95%	96%	85%	97%	98%	93%	6%
CUI [0.25,0.5,0.25]	92%	81%	95%	95%	95%	79%	97%	94%	91%	7%
CUI [0.25,0.25,0.5]	87%	94%	94%	94%	96%	69%	94%	95%	91%	9%
CUI [0.33,0.33,0.33]	92%	93%	94%	95%	95%	85%	97%	95%	93%	4%
Average over all	83%	81%	91%	89%	89%	81%	90%	88%	86%	
Std. Dev over all	9%	13%	12%	12%	13%	13%	13%	14%	13%	

Outlink-based schemes performed poorly (omitted).

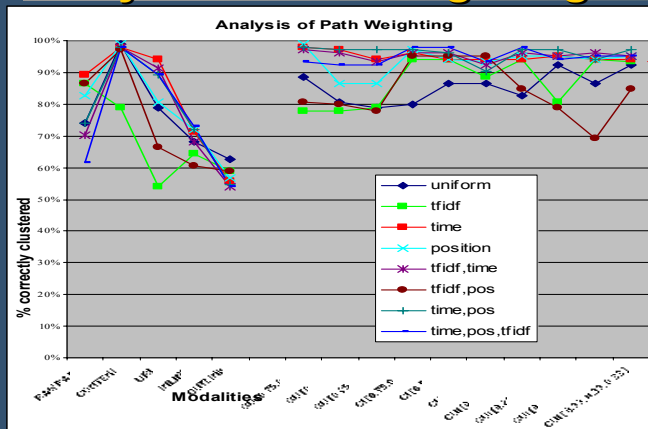
Analysis: Modalities



Content is King!
Mean=0.96,
StdDev=0.07

Linear Contrast shows Content sig. different:
(unimodal) $F(1,105)=32.51$, $MSE=.005361$, $p<0.0001$
(multimodal) $F(1,35)=33.36$, $MSE=.007332$, $p<0.0001$

Analysis: Path Weighting



View Time
is best!

Paired t-Test between Time-based and non-Time based weightings: $n=60$,
 $t(59)=4.85$, $p=4.68e-6$
V.T.mean=89.5%, s.d.=12.7%, non-V.T.mean=83.2%, s.d.=12.0%

Outline

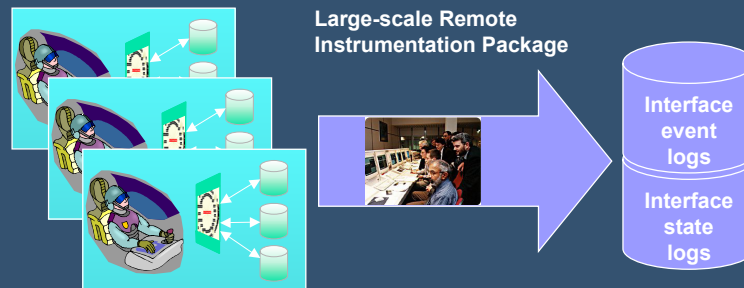
- Motivation and Background: InfoScent™
- LumberJack Evaluation
- Bloodhound Evaluation (how we got smarter)
- Conclusion

Bloodhound Validation Experiment

- Subjects: 244 users recruited from variety of sources
 - Mailing lists and Craig's List, \$20 per subject
 - For intranets, we recruited the employees of the company.

Lesson: Getting large amount of user data requires special infrastructure.

Remote In-Situ Instrumentation (WebLogger-Remote)



- 391 Users
- 8 WWW Sites
- 8 Tasks/Site

2003-07

NPUC IBM Almaden

Copyright 2002 PARC

parc

19

Tasks

- Material: Four different types of sites were studied.
 - Help.yahoo.com (help system of yahoo)
 - www.rei.com (eCommerce camping store)
 - HIVInsite.ucsf.edu (Medical HIV info)
 - parcweb.parc.com (Intranet)
- Each site had 8 tasks
 - Total $8 \times 4 \text{ sites} = 32$ tasks
 - Each task given a goal in the form of a question.
 - Tasks were chosen:
 - after we are familiar with the site
 - 50-50 mixture of hard/easy.
 - At least 50 users were assigned to each task.

2003-07

NPUC IBM Almaden

Copyright 2002 PARC

parc

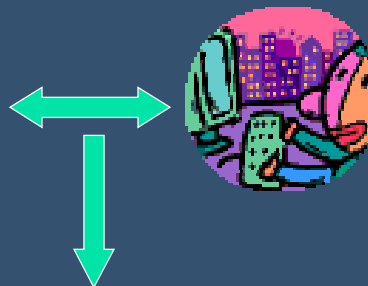
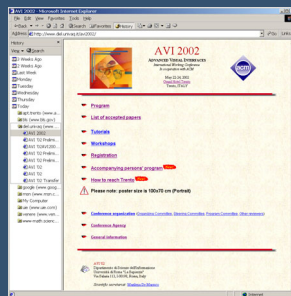
20

Procedure

- Used remote version of WebLogger [Reeder et. al. 2001]
- Received a URL and instructions.
 - Performed the test at their leisure (and in situ).
 - Could abandon a task, or continue at a later time.
 - No search engines allowed (want to study navigation).
- Each subject received 8 tasks across different sites (counter-balanced for difficulty).
- A button signifies the end of the task, record feedback.
- Records: Time to handle each task, pages accessed, keystrokes.

Lesson: Have users perform tasks as natural as possible.

Instrumentation: WebLogger



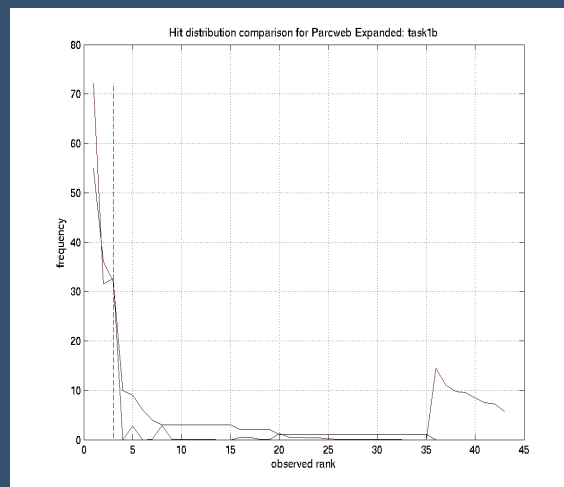
(BEFORE-NAVIGATE	(http://altavista.com/)	105.331s	0.100s	951763010	10:36:50)	
(DOC-MOUSEMOVE	(881 122)	105.431s	0.100s	951763010	10:36:50)	
(NAVIGATE-COMplete	(http://www.altavista.com/)		105.632s	0.201s	951763011	10:36:51)	
(EYETRACKER-SYNC	(103)	106.242s	0.610s	951763011	10:36:51)	
(DOCUMENT-COMplete	(http://www.altavista.com/)		106.773s	0.531s	951763012	10:36:52)	
(SCROLL-POSITION	(0 0 759 1181)	106.853s	0.080s	951763012	10:36:52)	
(DOC-MOUSEMOVE	(874 123)	107.024s	0.171s	951763012	10:36:52)	
(DOC-MOUSEMOVE	(874 123)	107.044s	0.020s	951763012	10:36:52)	
(DOC-MOUSEMOVE	(874 123)	107.214s	0.170s	951763012	10:36:52)	
(EYETRACKER-SYNC	(104)	107.244s	0.030s	951763012	10:36:52)	
(CHAR	(a	874 123)	108.125s	2.904s	951763013	10:36:53)
(EYETRACKER-SYNC	(105)	108.245s	1.001s	951763013	10:36:53)	
(DOC-KEYPRESS	(a INPUT)	108.446s	0.201s	951763013	10:36:53)	

Tell the true story of the data

Lesson: Now that you have some idea of what the data tells you, see if it's right, and tell the true story of the data (even if it's not what you expected).

- Compare Visitation Distributions
 - For each task, produce a user summary vector that describes the frequency distribution of page visit over the document space.
 - For each task, ran Bloodhound and create bloodhound predicted frequency distribution.

Freq. Distribution Comparison



Results

<i>Corr. Coeff.</i>	Yahoo	REI	HivInSite	Pareweb
task 1a	0.7528	0.4701	0.6811	0.7394
task 1b	0.7218	0.4763	0.7885	0.8756
task 2a	0.7489	0.9892	0.6671	0.8930
task 2b	0.8840	0.7073	0.6880	0.8573
task 3a	0.7768	0.7321	0.8835	0.7197
task 3b	0.6973	0.6979	0.5660	0.7123
task 4a	0.9022	0.9415	0.8407	0.8340
task 4b	0.9052	0.7600	0.4634	0.9344

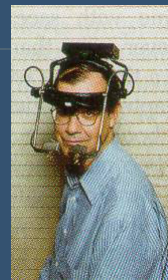
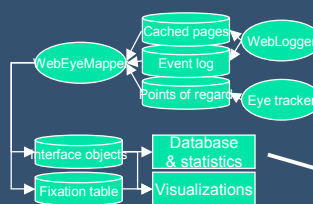
Analysis

- Produced click streams that:
- Correlated strongly 1/3 of the time
- Moderately slightly less than 2/3 of the time
 - Problem: we do not know *a priori* which third.

Lessons Condensed

- Hard to get data in the form that you want without **hard work**.
- Getting large amount of user data requires **special infrastructure**.
- Have users perform tasks **as natural as possible**.
- **Redundancy** is key to guard against problems.
 - Be careful in throwing away data.
- Tell the **true story** of the data (even if it's not what you expected).
- Get the **rawest data** you can to verify your scientific thinking.
 - Eyes don't lie!

Eyetracking



USER ID	URL	TIME	FIXATION	MOVEMENT
0001	http://www.ibm.com	100.2010	0.1000	0.1000
0002	http://www.ibm.com	100.2010	0.1000	0.1000
0003	http://www.ibm.com	100.2010	0.1000	0.1000
0004	http://www.ibm.com	100.2010	0.1000	0.1000
0005	http://www.ibm.com	100.2010	0.1000	0.1000
0006	http://www.ibm.com	100.2010	0.1000	0.1000
0007	http://www.ibm.com	100.2010	0.1000	0.1000
0008	http://www.ibm.com	100.2010	0.1000	0.1000
0009	http://www.ibm.com	100.2010	0.1000	0.1000
0010	http://www.ibm.com	100.2010	0.1000	0.1000

Eyetracking

Lesson: Get the rawest data you can. (Eyes don't lie!)



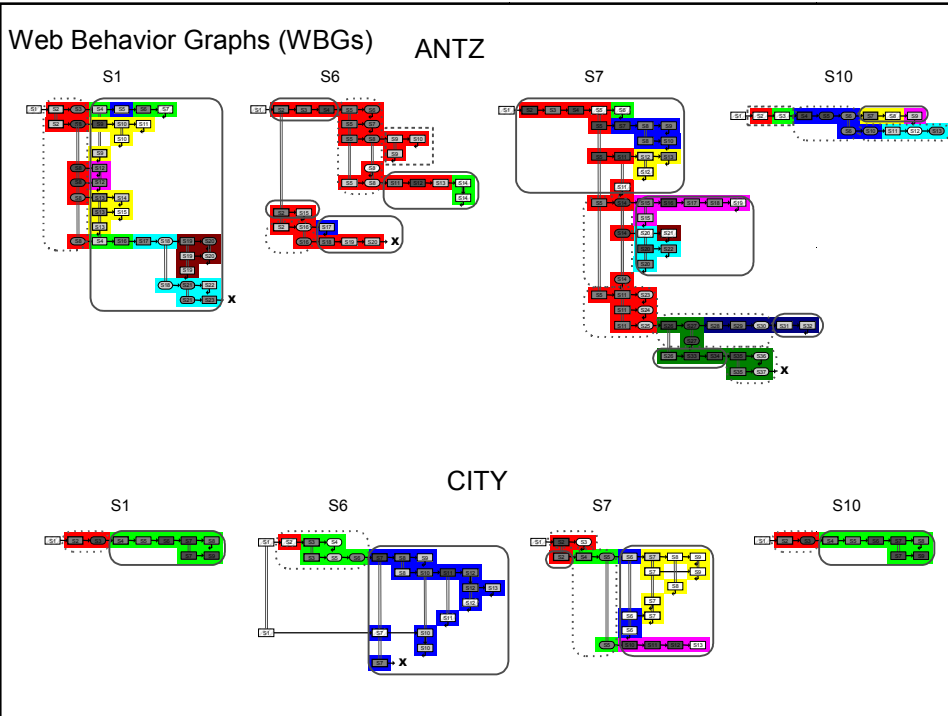
2003-07

NPUC IBM Almaden

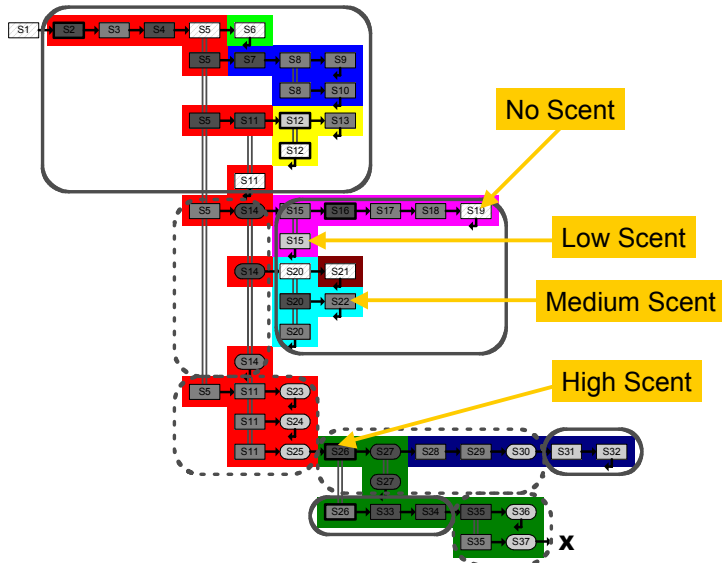
Copyright 2002 PARC

parc

31



Information Scent



Eyetracking Jakob

The screenshot shows the Yahoo! Weather website for Palo Alto, CA. Red lines indicate an eyetracking path:

- Starts at the top left (Yahoo! logo).
- Moves to the search bar.
- Then to the current weather display showing 61°F and 'Mostly Cloudy'.
- Finally to the 5-day forecast table.

Today	Tomorrow	Wed	Thu	Fri	6-10 Day
Heavy Rain/Wind High: 57 Low: 41	Rain High: 54 Low: 38	AM Showers High: 55 Low: 36	Showers High: 57 Low: 42	Showers High: 58 Low: 40	Extended forecast at weather.com

Acknowledgements

- Peter Pirolli, Stu Card, and the the UIR and Bloodhound Team at PARC.
- Participants in our user study
- Office of Naval Research
- George Karypis for CLUTO software

Contact:

Ed H. Chi (echi@parc.com)

