



Analyzing Image Searching on the Web: Learning from Web Searching Study for information Services

Youngok Choi
choiy@cua.edu

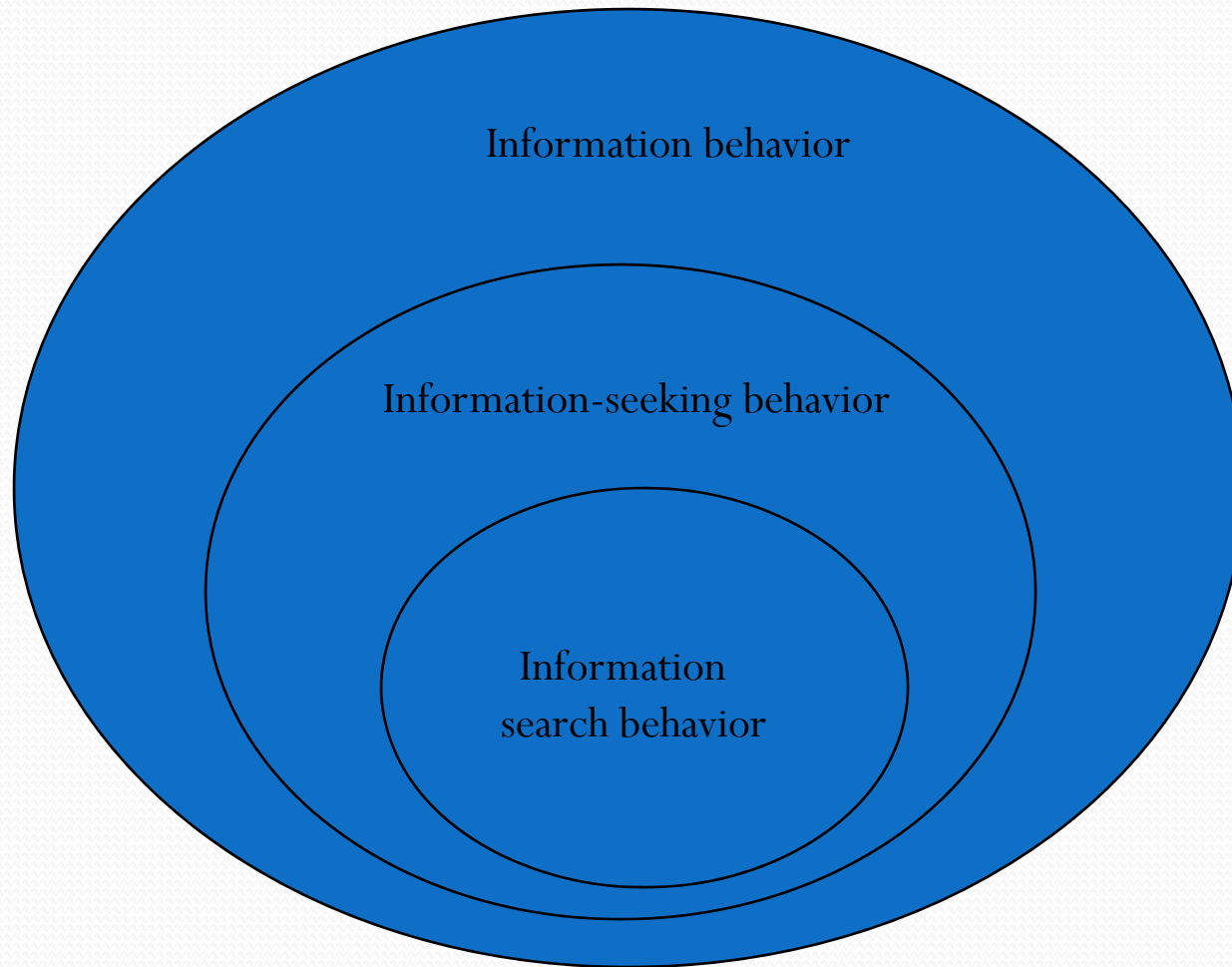
School of Library and Information Science
The Catholic University of America

January 30, 2009
Bridging the Spectrum (CUA-SLIS Symposium)

Why studies searching behavior for images on the Web?

- The Web as an information system.
- Strong preference for a rich variety of media besides text
- Digital images have been increasingly used in educational settings
 - Green, 2006, Harley et al, 2006, Pisciotta et al, 2005, Shonfeld, 2006
- Images are the most common type of digitized materials in many libraries and cultural organizations
 - (Palmer et al., 2007).

Wilson's nested Model, 1999





Information-seeking (In general)

- Uncertainty
- Question answering stages
- Gradual, exploratory, iterative
- Compound with tasks in a context



Research Objectives

- To examine natural image information searching behaviors of college students on the Web in terms of querying and browsing strategies
- To examine which of the searcher's activities are affected by contextual factors



Methodology-Participants

- 29 participants
 - 22 females and 7 males
 - average age : 21 years
 - 27 native English speakers, 2 non-English speakers
 - Major
 - 22 – Media Studies
 - 7 – others (English, Political Science, Sociology, Business Management)
 - Year
 - Freshman (5)
 - Sophomore (5)
 - Junior (9)
 - Senior (8)
 - graduate (2)




Data Collection

- Background Survey Questionnaire
 - Demographic data, computer experience, searching experience, hours per day using the Web, self-rated searching expertise, and self-rated digital image searching experience.
- Three search sessions per participant
 - Pre-search questionnaire for each session
 - Task, intended use of an image, topic familiarity
 - Screen capturing via Camtasia V. 5 for each session
 - Post-search questionnaire for each session
 - Relevance judgment and feedback
 - Think-aloud, Interview, Observation



Tasks

- Academic task
 - For class assignment or projects, for study guide
- Work related tasks
 - Creating a slideshow on the production of ethanol and other biofuels (internship), preparing a publication for incoming freshmen students, to prepare a college newspaper article or a yearbook, to use in Career Services Website, to create a brochure for a college newspaper, for an attorney general campaign
- Personal interest
 - For a mission trip, making a documentary on a personal interest, for career, for participating in a marathon race



	Session1	Session2	Session3	Total	%
Academic task	20	15	18	53	60.92
Work-related task	4	7	5	16	18.39
Personal interest	5	7	6	18	20.69
Total	29	29	29	87	100



Browser

	IE	Firefox
Session1	15	14
Session2	16	13
Session3	14	15



Search Duration (in Seconds)

Total: 94,770 seconds (26 hours 19 minutes 30 seconds)

	Minimum	Maximum	Sum	Mean
session1	152	2,639	31,357	1,081.28
Session2	48	2,763	33,462	1,153.86
Session3	116	2,243	29,951	1,032.79

Initiating approach for search

	Search Engine Web	Search Engine Images	Search engine Web -> Images	Direct access to site/page
Session1	7 (24.14%)	12 (41.38%)	5 (17.24%)	5 (17.24%)
Session2	11 (37.93%)	7 (24.14%)	9 (31.03%)	2 (6.90%)
Session3	10 (34.48%)	9 (31.03%)	6 (20.69%)	4 (13.79%)



Search Queries

Total 979 Queries

	Web_Query	Image_Query	Local_Query	Total_Query	Mean_	SD
Session1	110	156	54	320	11.03	8.29
Session2	121	158	45	324	11.17	9.59
Session3	136	169	30	335	11.55	8.58



Words per query

	Maximum	Mean	Std. Deviation
Session1	6	3.07	1.43
Session2	5.25	2.95	0.97
Session3	4.4	2.89	1.10



Query Formulation

- Only 3 participants used Boolean operators in 22 queries.
- 11 participants used a quotation mark in 89 queries (0.1 % of 979 queries)
- Few advanced search option use
- Participants frequently copied and pasted texts on a web page to modify their queries
- Many participants (N=11, N=7, N=10) used a query suggestion by Google and YouTube site (“Did you mean.....”)



Moves and Tactics on the Web

- Querying
 - Typing a query or URL for an active or direct interaction
- Navigating
 - Back, Forward, Home, Image_tab, Web_tab, Menu, Button
- Scanning
 - SEG_Next, SEI_Next, Local_Next : Moving around in search results pages (i.e. Previous or Next to move to search results)
- Extracting
 - Clicking on Image, Enlarging an image, SE_result_click, PageLinking, Saving

		Session1		Session2		Session3	
Tactic	Action	Frequency	%	Frequency	%	Frequency	%
Querying	Web_query	110	3.19	121	3.32	136	4.50
	Image_query	156	4.52	158	4.34	169	5.60
	Site_query	54	1.57	45	1.24	30	0.99
	URL	72	2.09	73	2.00	56	1.85
Navigating	Back	927	26.89	930	25.54	866	28.68
	Forward	10	0.29	11	0.30	15	0.50
	Home	2	0.06	2	0.05	0	0.00
	Image_tab	73	2.12	73	2.00	86	2.85
	Web_tab	34	0.99	34	0.93	38	1.26
	Menu	34	0.99	47	1.29	8	0.26
	Button	44	1.28	21	0.58	31	1.03
Scanning	SEG_Next	8	0.23	26	0.71	12	0.40
	SEI_Next	300	8.70	376	10.33	354	11.72
	Local_Next	146	4.23	339	9.31	40	1.32
Extracting	Image_clicking_SE	301	8.73	337	9.26	333	11.03
	Image_clicking_local	442	12.82	385	10.57	131	4.34
	SE_result_click	138	4.00	141	3.87	139	4.60
	PageLinking	192	5.57	150	4.12	165	5.46
	Enlarging	230	6.67	287	7.88	281	9.30
	Saving	175	5.08	133	3.65	131	4.34
Total		3448	100.00	3689	100.00	3021	100.00

Tactics across three sessions

	Session1		Session2		Session3	
	Frequency	%	Frequency	%	Frequency	%
Querying	392	11.37	397	10.76	391	12.94
Navigating	1,124	32.60	1,118	30.31	1,044	34.56
Scanning	454	13.17	741	20.09	406	13.44
Extracting	1,478	42.87	1,433	38.85	1,180	39.06
Total	3,448	100.00	3,689	100.00	3,021	100.00

Relevance across three sessions

	Usefulness			Satisfaction			Confidence		
	Session 1	Session2	Session 3	Session1	Session 2	Session 3	Session 1	Session 2	Session 3
Mean	6.07	5.9	5.93	6	5.5	5.59	5.31	5.76	5.62
Median	6	6	6	6	6	6	6	6	6
Mode	7	7	7	7	7	6	6	7	7
SD	0.961	1.205	1.387	1.035	1.689	1.402	1.312	1.504	1.347

Effects of contextual factor (1) – ANOVA test

- Task type on session duration ($F=3.55$, $p=0.04$) and querying ($F=4.16$, $p=0.03$) in session 2
 - More time was spent to conduct search on academic task
 - More querying was used for academic task
- Searching expertise on querying ($F=4.07$, $p=0.03$) in session2 and navigating ($F=5.71$, $p=0.01$) in session3
 - Those with a higher level of searching expertise used more navigation
 - Those with a lower level of searching expertise used more querying

Effects of contextual factor (2) – ANOVA test

- Differences of Topic familiarity on Relevance (satisfaction, confidence, usefulness) were observed in session 2 and session 3
 - Satisfaction ($F=3.16$, $p=0.03$) and Confidence ($F=3.91$, $p=0.01$) in session 2; Usefulness ($F=2.81$, $p=0.048$) in session 3
 - When the participants felt very familiar with the search topic, their level of satisfaction, usefulness, and confidence tended to be higher.
- Differences of digital image searching experience on the level of satisfaction were observed in session 2
 - The more digital image searching experience, the more the participants were satisfied with the search results.



Effects of time factor on tactics

- No difference among search tactics in three sessions of 8 participants who conducted image searches for a same topic (Repeated Measures ANOVA test & Friedman Test)
- Effect of time on searching tactics in two sessions of 8 participants who conducted image search for a same topic (Paired samples t-test)
 - Only the level of confidence changed during two sessions for the same topic ($t=-2.55$, $p=0.04$)
 - Their confidence level went up at the second session
 - Mean= 5.12 (the first session), Mean=6.25 (the second session)

Some concerns on image searching on the Web

- Difficulty finding archival images or images from the turn of the century on-line.
- Copyright issues
- “Image titles don't accurately show what their content is and don't come up in a search for a certain topic.”
- “Same images over and over again”
- Narrowing down the number of images to sort through and eliminating images that are not of use.
- Finding many unrelated images
- Identifying good search terms
- Poor quality or size
- Not enough knowledge of different sites and types of sites for images.



Conclusions (1)

- Heavy reliance on Google Image Search and Web Search
- Rare use of special image search engines or image search sites
- Short queries; High degree of modification; Low use of Boolean operators
- Browsing through a general image search result pages was frequent
- Frequently clicked on an image or enlarged it to view/select relevant images



Conclusions (2)

- Effects of contextual factors (task, searching expertise, topic familiarity, digital image searching experience, time) on searching behaviors and relevance were present.
- There was more continuity than change in searching behavior.
- Source credibility, quality, size, and details of an image mattered in relevance judgment.
- The participants looked for a browsing clue for images, i.e. a link to photo gallery, photo album, slide show, images, or photos on a web page.



Implications for LIS Services

- Information literacy programs
 - For various target groups
 - How to identify search concept and search terms
- Identifying appropriate information channel for visual information
 - Portals
- Web site design
 - Search function/option
 - Categories , Term selection support
 - Browsing option for photos or images
 - For example, “Photo album,” “Slide Show”
 - Zooming option
 - Add contextual information for an image



Many Thanks to
OCLC/ALISE LISR Grant
&
Ryan McPherson!