

Relevance

Jaime Arguello

INLS 509: Information Retrieval

jarguell@email.unc.edu

How Do Users Assess Web-Pages

a case-study

- What document attributes do users focus on when they judge a webpage as being useful or not useful to a specific type of information seeking task?
- Does attribute importance depend on the type of task?
- Does attribute importance depend on the level of time pressure?
- Does attribute importance depend on the stage in the task?

(Tombros *et al.*, '04)

How Do Users Assess Web-Pages

a case-study

- Implications:
 - ▶ Predicting relevance (query-document score + prior)
 - ▶ Surrogate representation
 - ▶ Feature generation for a machine-learned model

Method

- 24 subjects and 3 information-seeking tasks
 - ▶ **Task 1 (background search):** find demographic information about who uses the internet
 - ▶ **Task 2 (decision task):** find the best hi-fi speakers given a budget
 - ▶ **Task 3 (list task):** find interesting things to do over a weekend in Kyoto, Japan
- Every subject did every task
- 12 subjects were given 15 minutes per task
- 12 subjects were given 30 minutes per task

Method

- Subjects were instructed to “think aloud” as they judged visited pages as being useful or not useful
- Search sessions were recorded
- Subjects’ reasons for judging documents useful or not useful were annotated, organized into categories, and analyzed
- A pre-determined set of criteria was not given to the participants

Number of Documents Judged

	Useful		Not useful		
	15'	30'	15'	30'	Totals
Task 1	37	74	90	115	316
Task 2	40	65	76	76	257
Task 3	56	74	91	68	289
Totals	133	213	257	259	862

- Not surprisingly, 'not-useful' judgements were more frequent than 'useful' judgements
- See anything else interesting?

Results

- Reasons for judging a document 'useful' or 'not useful' were divided into 5 feature categories
 1. **text/content features:** based on the text
 2. **structure features:** based on the way information is presented and links to other webpages
 3. **quality features:** all things related to webpage quality (recency, authority, depth/scope)
 4. **non-textual features:** presence of non-textual items
 5. **physical/accessibility features:** language, file-size, load time, etc.

Results

text/content features

	Useful		Not useful		Combined	
	#	%	#	%	#	%
Text	367	46.69	349	42.77	716	44.69
Content	185	23.53	204	25	389	24.28
Numbers	109	13.87	49	6	158	9.86
Titles/headings	37	4.71	34	4.17	71	4.43
Query terms	34	4.33	29	3.55	63	3.93
Too much	2	0.25	33	4.04	35	2.18

- Percentages are based on the total number of feature-mentions when judging a document useful, not-useful or both combined
- Interesting trends?

Results

structure and quality features

	Useful		Not useful		Combined	
	#	%	#	%	#	%
Structure	176	22.39	170	20.83	346	21.60
Layout	60	7.63	95	11.64	155	9.68
Links	80	10.18	28	3.43	108	6.74
Links quality	5	0.64	37	4.53	42	2.62
Table data/table layout	31	3.94	10	1.23	41	2.56
Quality	133	16.92	150	18.38	283	17.67
Scope/depth	28	3.56	59	7.23	87	5.43
Authority/source	61	7.76	23	2.82	84	5.24
Recency	31	3.94	35	4.29	66	4.12
General quality	8	1.02	25	3.06	33	2.06
Content novelty	5	0.64	4	0.49	9	0.56
Error on the page	0	0	4	0.49	4	0.25

Results

non-textual and 'other' features

	Useful		Not useful		Combined	
	#	%	#	%	#	%
Non-textual items	99	12.60	44	5.39	143	8.93
Pictures	99	12.60	44	5.39	143	8.93
Physical properties	11	1.40	103	12.62	114	7.12
Page not found	0	0	36	4.41	36	2.25
Page location	6	0.76	16	1.96	22	1.37
Page already seen	1	0.13	16	1.96	17	1.06
Language	1	0.13	4	0.49	5	0.31
File type	0	0	4	0.49	4	0.25
File size	2	0.25	1	0.12	3	0.19
Connection speed	1	0.13	13	1.59	14	0.87
Subscription/registration	0	0	13	1.59	13	0.81

Discussion

single indicators of relevance

Content	111 (28.7%)	Links quality	16 (38.1%)
Query terms	24 (38.1%)	Links	15 (13.9%)
Scope/depth	21 (24.1%)	Pictures	12 (8.4%)
Layout	17 (11%)	Authority/source	10 (12%)
Recency	17 (25.8%)		

- Oftentimes, only a single feature was mentioned when making a judgement
- Percentages based on the number of times that attribute was the only attribute mentioned
- **Implications:** these are features that should be displayed in surrogates so that people can decide whether to look at the page more closely

Discussion

features used to judge relevance vs. non-relevance

Useful	Not useful
Content (185)	Content (204)
Numbers (109)	Layout (95)
Pictures (99)	Scope/depth (59)
Links (80)	Numbers (49)
Authority/source (61)	Pictures (44)
Layout (60)	Links quality (37)
Titles/headings (37)	Page not found (36)
Query terms (34)	Recency (35)
Recency (31)	Titles/headings (34)
Table data (31)	Too much text (33)

- Different features were used when judging a document 'useful' vs 'not useful'
- Single indicators were more common in 'not useful' judgements
- Is that surprising?

How Do Users Assess Web-Pages

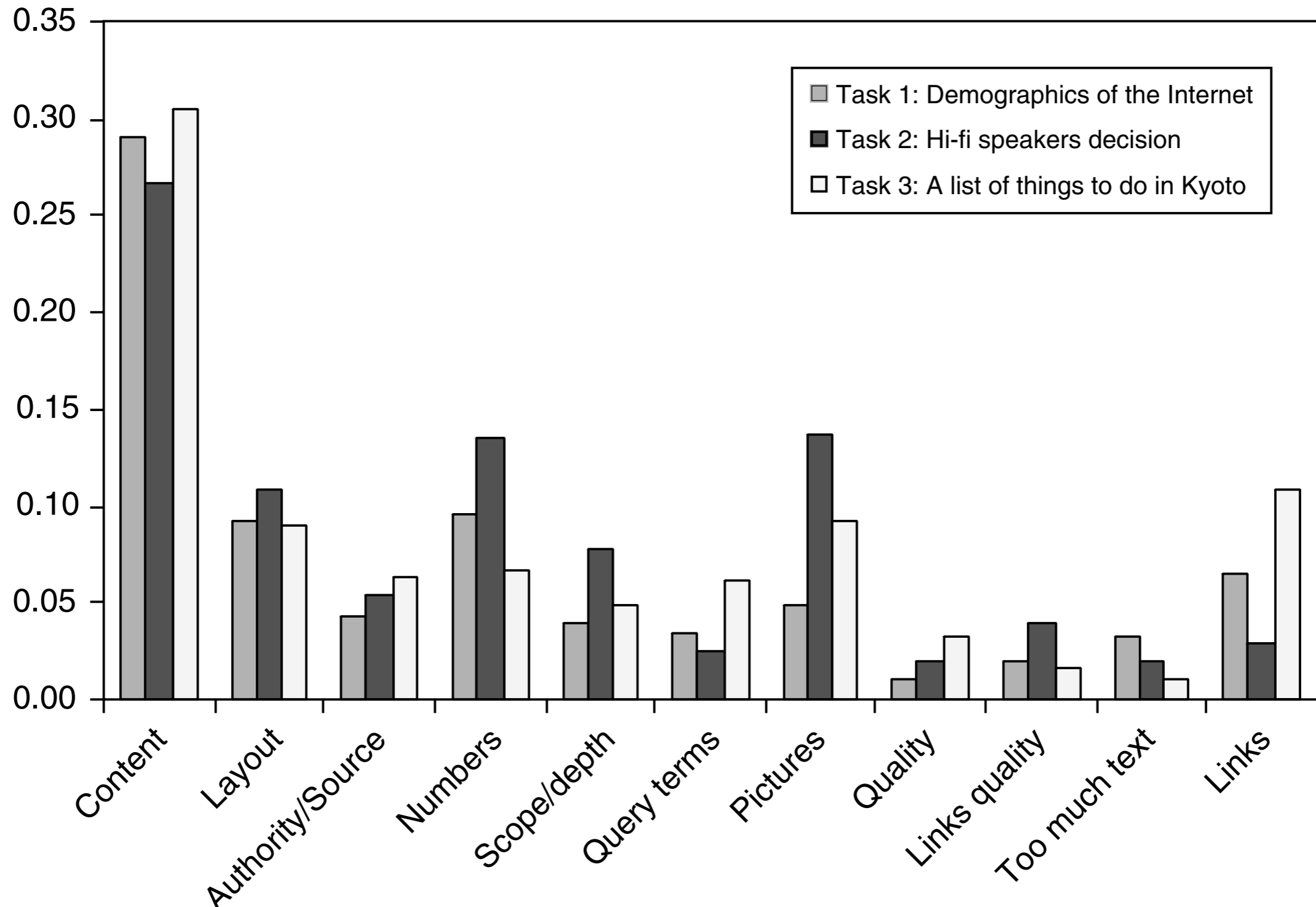
a case-study

- What document attributes do users focus on when they judge a webpage as being useful or not useful to an the information seeking task?
- Does attribute importance depend on the type of task?
- Does attribute importance depend on the level of time pressure?
- Does attribute importance depend on the stage in the task?

(Tombros *et al.*, '04)

Discussion

feature importance across tasks



How Do Users Assess Web-Pages

a case-study

- What document attributes do users focus on when they judge a webpage as being useful or not useful to an the information seeking task?
- Does attribute importance depend on the type of task?
- Does attribute importance depend on the level of time pressure?
- Does attribute importance depend on the stage in the task?

(Tombros *et al.*, '04)

Discussion

feature importance for 15-minute- vs. 30-minute-group

	Content	Layout	Numbers	Pictures	Links
15 min.	27.37%	10.26%	8.24%	8.24%	8.09%
30 min.	29.77%	9.33%	11.01%	9.43%	5.87%

Scope depth	Query terms	Recency	Authority source	Links quality	Too much text
6.53%	5.60%	4.98%	4.35%	4.04%	2.80%
4.72%	2.83%	3.56%	5.87%	1.36%	1.78%

- Interesting trends?

Discussion

feature importance for 15-minute- vs. 30-minute-group

	Content	Layout	Numbers	Pictures	Links
15 min.	27.37%	10.26%	8.24%	8.24%	8.09%
30 min.	29.77%	9.33%	11.01%	9.43%	5.87%

Scope depth	Query terms	Recency	Authority source	Links quality	Too much text
6.53%	5.60%	4.98%	4.35%	4.04%	2.80%
4.72%	2.83%	3.56%	5.87%	1.36%	1.78%

- The 15-minute-group relied more on superficial features
(Tombros *et al.*, '04)

How Do Users Assess Web-Pages

a case-study

- What document attributes do users focus on when they judge a webpage as being useful or not useful to an the information seeking task?
- Does attribute importance depend on the type of task?
- Does attribute importance depend on the level of time pressure?
- Does attribute importance depend on the stage in the task?

(Tombros *et al.*, '04)

Discussion

differences in mentions between the first and last query

		Layout	Numbers	Links quality	Content	Query terms	Pictures	Scope/ depth	Authority	Links	Quality
Task 1	First	9.4%	8.6%	2.3%	31.2%	3%	4.5%	4.5%	2.6%	5.6%	1.1%
	Last	8.1%	9.3%	3%	26.7%	4.2%	5.5%	3%	4.2%	7.2%	1.3%
Task 2	First	9.3%	16%	4%	29.3%	2%	14%	4%	5.3%	0.4%	0.3%
	Last	13.8%	12.2%	2.7%	26.4%	3.4%	14.9%	8.1%	6.8%	2.7%	3.4%
Task 3	First	11.3%	8.4%	2.1%	29.9%	6.9%	7.4%	3.4%	5.4%	11.8%	2.9%
	Last	4.5%	6.1%	0%	36.5%	5.1%	13.2%	5.9%	4.5%	10.3%	1.9%

- Interesting trends?

Conclusions and Implications

- Users judge the usefulness of web-pages using different criteria
- The most important features seem to depend on the task
- The most important features seem to depend on urgency (15-minutes vs. 30-minutes)
- The most important features seem to depend on the stage of task completion
- Caveats: (1) 'relevance' in the context of this study seems to mean 'perceived relevance' and (2) participants did not necessarily mention everything that was considered.