

# Building a Test Collection for Investigating Contextual IR

---

**Diane Kelly**

School of Information & Library Science  
University of North Carolina  
[kelly@ils.unc.edu](mailto:kelly@ils.unc.edu)

IRiX Workshop / SIGIR '04 / Sheffield, UK / 29 July 2004



1

## Why Build a Collection?

---

- Building a test collection is a time-consuming and laborious process
- Standard test collections allow for:
  - More rapid progress
  - Comparison of techniques
  - Examination of alternative hypotheses
  - Replication of previous findings
- Examples of highly used test collections: TREC



## Minimum Criteria

- The collection should contain more than just documents, topics and relevance judgments
- The collection should contain information about users, such as their:
  - information needs and goals
  - information-seeking context
  - behaviors within this context
- The collection should be constructed within a natural use environment & over an extended period of time

## Challenges

- Discussion about:
  - what context is
  - what elements of context matter in IR
  - how to measure these elements has to occur
- Construction & evaluation of new data collection tools
- Recruitment, retention and compensation of subjects
- Funding

## An Example Collection

**Users:** Seven Ph.D. students from unique departments

**Time Period:** 14 weeks = 1 university semester

### Instruments & Variables

- Laptops & printers (compensation)
- Client- and proxy-side loggers (**behaviors**: display time, URLs requested, print & save, keystrokes, local copies of all documents requested)
- Task & Topic Questionnaires (**context**)
  - Task [Endurance, Frequency & Stage]
  - Topic [Persistence & Familiarity]



## An Example Collection

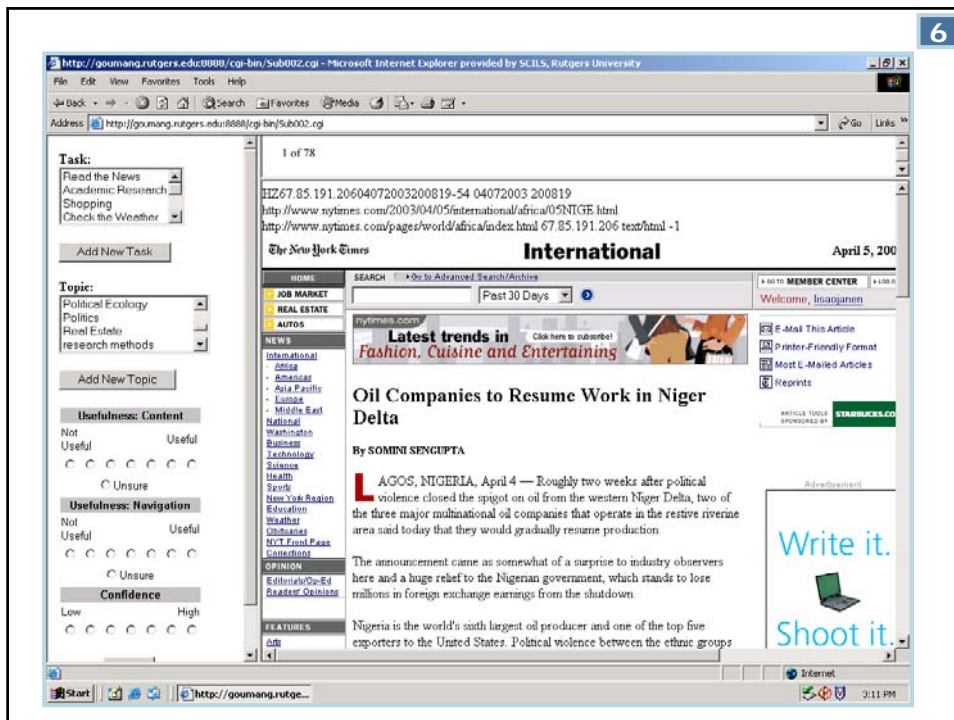
### Instruments (continued)

- Evaluation Interface (**document usefulness** and confidence ratings; 7-point scales)

### Procedures

- Initial Meeting
- Weekly one-hour meetings
- Exit Interview





**7**

```

<document>
  <id>0000001</id>
  <date>03052004</date>
  <referrer>http://www...</referrer>
  <url>http://www...</url>
  <title>Oil Companies to Resume ...</title>
  <episode 1>
    <time>01:24:30</time>
    <user userid=2>
      <context>
        <task>Read the news
          <endurance>7</endurance>
          <frequency>3</frequency>
        </task>
        <topic>Political Ecology
          <persistence>7</persistence>
          <familiarity>6</familiarity>
        </topic>
      </context>
    </user>
  </episode 1>
  </document>

```

*Continued on next slide...*

Continued from previous slide...

```

<behaviors>
  <display_time>00:00:40</display_time>
  <print>1</print>
  <save>0</print>
  <keystrokes>0</print>
  <scrolling>0</scrolling>
</behaviors>
<usefulness_content>6</usefulness_content>
<usefulness_navigation>1</usefulness_navigation>
<confidence>7</confidence>
</user>
</episode>
CONTENT OF DOCUMENT
</document>

```

## Results: QUANTITY OF DATA

		Client Log	Proxy Log	# Req	# Eval	# of Tasks	# of Topics
S U B J E C T	1	2.6 MB	1.7 GB	15,499	843	6	9
	2	6.8 MB	83 MB	5,319	771	11	80
	3	3.9 MB	39 MB	3,157	381	19	17
	4	2.0 MB	42 MB	3,205	351	25	35
	5	1.5 MB	48 MB	3,404	193	12	25
	6	21.7 MB	2.9 GB	14,586	1,181	21	40
	7	4.9 MB	2.1 GB	11,657	1,125	33	26
TOTAL		43.4 MB	6.9 GB	56,827	4,845	127	232

## Costs

	Current	Current+	Projected
<b>Users</b>	7	7	21
<b>Time Frame (weeks)</b>	14	14	14
<b>Hours spent per user/per week</b>	6	3	3
<b>Total Hours</b>	588	294	882
<b>Hardware, software, compensation</b>	\$10,000	\$10,000	\$210,000*
<b>Page Evaluations</b>	4,845	4,845	14,535

## Getting Started ...

### In-workshop exercise

- What does context mean to you?
- If you were building a test collection for IRiX, what are the one or two “contextual variables” that you would definitely collect?

## Future Directions

- Conceptualizing and operationalizing context and interactions
- Assembling and evaluating a set of data collection tools
- Establishing an appropriate data format
- Seeking funding