CSC 484 Lecture Notes Week 7

Data Gathering and Analysis

I. **Relevant reading**.

   A. Textbook Chapters 7 and 8

   B. Selected portions of Chs 13 and 14.

Relevant reading, cont'd

C. Weeks 7 and 8 research reading
         (one paper for two weeks)

"Integrating statistics and visualization..."

by Perer and Shneiderman

2008 SIGCHI

Relevant reading, cont'd

D. Certain teams should read ahead.

  1. `2d3d` read *Chapter* 13.

  2. `swat` read of *Section* 14.3.

II. **Intro to Ch 7 (Section 7.1).**

A. Planning, conducting data gathering.

B. The book considers for requirements and usability evaluation.

C. Focus of 484 is *evaluation*.

Intro to Ch 7, cont'd

D. Three specific techniques:

1. in-person interviews

2. questionnaires

3. (non-intrusive) observation

Intro to Ch 7, cont'd

E. Additional techniques in Chs 12, 13, 14.

III. **Four key data gathering issues (Sec 7.2).**

A. *Setting goals (Sec 7.2.1).*

1. Very important at outset.

2. Surprisingly easy to forget.

Four data gathering issues, cont'd

a. Be completely clear on user tasks.

b. Be clear on what you need to know.

Data gathering issues, cont'd

3. 484 goals defined:

    a. overall project goals in Milestone 2

    b. usability study goals in Milestone 3

Data gathering issues, cont'd

B. *Relationship with participants (Sec 7.2.2).*

   1. Establish and maintain a professional rel'p.

      a. In 484, subjects sign consent form.

Data gathering issues, cont'd

b. See

`calpoly.edu/˜sdavis/human2.htm`

for template.

Data gathering issues, cont'd

c. Subject anonymity most likely not necessary for 484.

    i. If you take photos.

    ii. If you obtain qualitative results.

Data gathering issues, cont'd

C. *Triangulation (Sec 7.2.3).*

   1. Means using > 1 technique.

   2. Doing so provides more useful and believ-
      able results.

Data gathering issues, cont'd

3. In 484 studies

   a. *questionnaires,*

   b. *subject performance data,*

   c. *possibly other forms of observation,*

   d. *possibly in-person interviews*

Data gathering issues, cont'd

D. *Pilot studies (Sec 7.2.4).*

   1. Small, separate study.

   2. Used to "debug" data gathering techniques.

Data gathering issues, cont'd

3. E.g., pilot questionnaire.

4. Can be indispensable.

5. In 484, no time for these.

IV. **Data recording (Sec 7.3).**

A. Forms are well known, i.e.,

1. *Hand-written, PDA, laptop notes.*

2. *Questionnaires.*

3. *Still photographs.*

4. *Audio recording.*

5. *Video recording.*

Data recording, cont'd

B. Noteworthy considerations:

1. Always ask permission of interviewees.

2. Avoid adding bias.

3. Explicit data recording may distract.

Data recording, cont'd

4. One team member ask, another records.

5. Transcribing can be time consuming

Data recording, cont'd

C. Table 7.1 (book page 297) has comparison.

D. In your studies, think over the pros and cons.

V. Interviews (Sec 7.4).

   A. "Conversation with a purpose".

   B. Four general types (Secs 7.4.1 - 7.4.4).

Interviews, cont'd

1. *Unstructured* -- open-ended discussion

2. *Structured* -- predetermined questions

3. *Semi-Structured* -- combination

4. *Group* -- multiple interviewees

Interviews, cont'd

C. Planning, conducting interview (Sec 7.4.5).

   1. Even unstructured should have a plan.

   2. Open-ended questions when you don't know in advance all answers

Interviews, cont'd

3. Closed questions in a structured interview

4. "Closed" means fixed set of answers.

5. Book has additional guidelines, pp. 304-307.

Interviews, cont'd

D. Other forms of interview (Sec 7.4.6).

1. Phone and online possibly useful.

2. Generally no substitute for face-to-face

Interviews, cont'd

E. "Enriched" interviews (Sec 7.4.7).

F. Table 1 (in notes) summarizes

	Unstructured	**Structured**	**Semi-Structured**
Replicatable	Not easily	Yes	Somewhat
Amenable to Statistical Analysis	No	Yes	Somewhat
Easily Transcribable	No	Reasonably	Somewhat
Type of Planning	General Agenda	Rigid Agenda	Rigid then General
Type of Questions	Open-ended	Fixed Answer Set	Combination

Interviews, cont'd

G. 484 will use questionnaires.

1. `swat` will conduct interviews

2. Other teams can employ as appropriate.

VI. **Questionnaires (Sec 7.5).**

A. Same questions as structured interview.

B. Questions must be
   very clear and unambiguous.

Questionnaires, cont'd

C. Motivation is an issue.

1. Easier to encourage responses in person.

2. Mitigated by in-person questionnaires,

   as in 484.

VII. Questionnaire design (Sec 7.5.1).

A. Ask for demographic data;
    likely not relevant in 484.

B. Points to consider:

Questionnaire design, cont'd

1. ***Clear instructions*** -- provide them up front, including any necessary definitions.

2. ***Question ordering*** -- ask most important questions first.

Questionnaire design, cont'd

3. ***Different versions of the questionnaire*** -- consider if you need them.

4. ***Keep it short and sweet*** -- even in monitored studies, users quickly grow weary.

Questionnaire design, cont'd

C. Can have bifurcation points.

    1. E.g., *"If X is true ... "*

    2. Less likely useful in 484

Questionnaire design, cont'd

D. See book pages 313 - 314 for a general example questionnaire.

VIII. Question response formats (Secs 7.5.2).

A. *Check boxes and ranges*

1. Select appropriately

2. Be careful to avoid overlaps.

3. Avoid annoyingly long lists.

Question response formats, cont'd

B. *Rating scales*

1. Common are Likert, semantic differential.

2. Book goes over details, pp. 313-317.

IX. Administering questionnaires (Sec 7.5.3).

A. Return rates vary widely.

B. 484 is somewhat specialized case --

subjects complete questionnaires in person.

X. **Online questionnaires (Sec 7.5.4).**

A. Tools and templates available.

B. Book has details pp. 317-321.

C. Each team consider if online appropriate.

XI. **Questionnaire use in 484.**

A. Per M3 writeup, all 484 teams
      use one or more questionnaires.

B. Use in two modes:

Questionnaire use in 484., cont'd

1. integral part of prototype-based[1] study

2. qualitative adjunct to prototype-based study

_____

[1] For 2d3d team, substitute "game-based"

Questionnaire use in 484., cont'd

C. Multiple questionnaires for different user groups, e.g.,

1. `gatekeeper` -- 484 students, Byron

2. `2d3d` -- outside subjects, 484 students

3. `menupad` -- restaurant owner, 484 students

XII. **Observation (Sec 7.6).**

A. For 484, qualitative observation is secondary.

B. Interaction logs may be useful, e.g.,
   `2d3d` and `mobility`.

Observation, cont'd

C. Consider what you need to do.

D. Most important -- *be unobtrusive.*

Observation, cont'd

E. Book has details, p. 321-342.

1. *Field observation (Sec 7.6.1).*

2. *Observation in controlled environment (Sec 7.6.2).*

3. *Indirect observation via tracking user (Sec 7.6.3).*

XIII. **Choosing, combining (Sec 7.7).**

A. Questionnaire required for 484.

B. Carefully and thoughtfully consider other data gathering techniques.

C. Summary on book pp. 342-346.

XIV. Introduction to Chapter 8 (Sec 8.1).

A. Data analysis can be *quantitative*, *qualitative*, or both.

B. Ch 8 presents ways to analyze data gathered with techniques described in Ch 7.

Intro to Ch 8, cont'd

C. *Interpretation* of analysis results.

  1. Simple interp'n identifies patterns, trends.

  2. Deeper interp'n draws conclusions from
     statistical analysis.

Intro to Ch 8, cont'd

D. Interpretation must be done carefully, supported fully by data.

1. E.g., suppose stats say one group of study subjects is slower than another.

Intro to Ch 8, cont'd

2. Could be interpreted in a number of ways

   a. skill differences between groups

   b. differences in how groups trained

   c. differences in study administration

Intro to Ch 8, cont'd

3. Eliminating the effects of such factors is part of designing a good study.

Intro to Ch 8, cont'd

E. Avoid over-claiming.

   1. Be maximally conservative in conclusions.

   2. Don't use "all", "most" unfoundedly.

   3. Back up claims with hard numbers.

XV. Defs of "Quantitative", "Qualitative" (Sec 8.2).

A. Quantitative data are numeric.

B. *Qualitative data* are not numeric, in a meaningful way.

Defs of "Quantitative", "Qualitative", cont'd

1. "Meaningful way" is important.

2. See *"How to Lie with Statistics"*,
      by Darrel Huff.

XVI. First steps in data analysis (Sec 8.2.1).

A. Most of these steps are common sense.

B. If you have interview notes, transcribe them as soon as possible.

First steps in data analysis, cont'd

C. Questionnaire data may need "grooming".

1. E.g., remove unanswered questions.

2. Electronic questionnaire tools can assist.

3. But you need to read the documentation.

First steps in data analysis, cont'd

D. Initially analyze other data

   1. Photos get dated caption.

   2. File things in appropriate places.

E. Table 8.1 (book page 359) summarizes.

XVII. Simple quantitative analysis (Sec 8.3).

A. A hard-to-analyze question:

   What do you think of feature X?

   with typical responses

   - *"It's stupid."*

   - *"I liked it a lot."*

   - *"It's hard to use, because ... "*

First steps in data analysis, cont'd

B. Easy-to-analyze questions:

Feature X is useful.

☐ *Strongly Disagree   ...   ☐ Strongly Agree*

Feature X is easy to use.

☐ *Strongly Disagree   ...   ☐ Strongly Agree*

First steps in data analysis, cont'd

C. Basic analysis examples pp. 362-373.

   1. Small-scale analyses relevant to 484.

   2. Large-scale (Box 8.3) far less relevant

       100 MB of data, 26 days, 21 hours a day.

XVIII. Simple Qualitative Analysis (Sec 8.4)

A. Book provides guidelines, but more relevant to requirements than evaluation.

B. Nevertheless, some useful info.

XIX. Identifying recurring patterns, themes (Sec 8.4.1).

A. Staring point of data analysis.

1. Sometimes the primary basis of analysis.

2. More complicated analysis may follow.

3. Patterns often apparent in graphical views.

Identifying recurring patterns, themes, cont'd

B. Unexpected patterns, themes can emerge.

   1. Book discusses emerging themes in ethnographic data.

   2. Domain not directly relevant to 484, but observations are instructive.

XX. **Categorizing data (Sec 8.4.2).**

A. Necessity depends on open-endedness of study.

B. E.g., "think-aloud" techniques require significant post-gathering categorization

Categorizing data, cont'd

1. Process same as SEs do.

2. I.e., determine emergent categories of functionality from user interviews.

Categorizing data, cont'd

3. Called "domain analysis" by SEs.

4. Top of page 383:

   "In this approach, nouns and verbs are identified and scrutinized to see if they represent significant classes."

Categorizing data, cont'd

5. Same of functional analysis used determine categories relevant to user studies.

Categorizing data, cont'd

C. Closed-form, requires (far) less post-gathering categorization.

   1. Categorization done up front.

   2. Inherent in determining meaningful answers to closed questions.

Categorizing data, cont'd

D. In 484, there's pre-gathering categorization.

1. Due largely to prototyping-based process.

2. New categorizations may emerge.

3. Part of the process.

4. Can result in major benefits.

XXI. **Looking for critical incidents (Sec 8.4.3).**

A. Identify particularly significant events.

1. E.g., users get stuck.

2. Or user has "ah hah" moment.

B. Probably not sufficient for a full analysis, but can help focus on significant problems.

XXII. Tools to support data analysis (Sec 8.5)

A. Surveys/questionnaire tools include

1. *phpESP*

2. *SurveyMonkey*

3. *InstantSurvey*

Tools, cont'd

B. *Many* stats tools;

1. `freestatistics.`
   `altervista.org/en/stat.php`

2. Microsoft Excel for ANOVA.

3. 2007 CHI paper on "Touchstone".

XXIII. Theoretical Frameworks (Sec 8.6).

A. Not general socio-cognitive frameworks.

B. Rather, they're *domain-specific*.

C. Based on empirical data.

Theoretical Frameworks, cont'd

D. Very much like SE*domain models*
        AI *ontologies*.

    1. Analysis of artifacts, activities, relationships.

    2. To help analysts understand domain.

Theoretical Frameworks, cont'd

E. Theoretical psychs should do some reading.

XXIV. Presenting the findings (Sec 8.7).

A. Presented throughout Ch 8.

B. Last section outlines three additional ways

Presenting findings, cont'd

1. ***Rigorous notations (Sec 8.7.1)***

    -- UML and other modeling notations.

2. ***User stories (Sec 8.7.2)***

    -- a childish form of scenarios

3. ***Summaries (Sec 8.7.3)***

    -- necessary part of any analysis activity

Presenting findings, cont'd

C. First two pertain to requirements.

D. *Summaries* pertain to data analysis.

XXV. **Data analysis and presentation in 484.**

A. Some techniques discussed in book are directly applicable.

B. Use what works for your project.

Data analysis and presentation in 484, cont'd

1. Except for the `2d3d` project, 484 usability studies are very small scale.

2. Big-gun stats (e.g., ANOVA) most likely not appropriate.

Data analysis and presentation in 484, cont'd

C. Techniques that *are* appropriate:

1. various forms of tables and graphs

2. at least some basic statistical analysis

3. a clearly written summary of the findings

Data analysis and presentation in 484, cont'd

D. Posted W07 examples

1. Located under `484/examples`.

2. Like storyboard examples, they're *"as is"*.

3. W07 deliverables details vary.

4. If you have specific questions,
   come by office hours any time.