CSE 291 I: Usability of Programming Languages ("Programmers Are People Too")

Michael Coblenz

Grounded Theory and Thematic Analysis

Or: How to Analyze Qualitative Data

Source today: Virginia Braun & Victoria Clarke (2006) Using Thematic Analysis in Psychology, Qualitative Research in Psychology, 3:2, 77-101.

Goal: Identify Patterns in Data

- Suppose you have a bunch of free-form, qualitative data

  - Interview transcripts

  - Web forum posts

  - Survey responses

- You want to identify and report on patterns.

  - "70% of the participants wanted a Mexican restaurant at Price Center."

Problem 1: Data Are Messy

- "I'd love a place that served empanadas."

  - Are empanadas Mexican cuisine? Or Argentinian? Or both? Do most Mexican places serve empanadas?

  - Does this data support a proposal to add a Mexican restaurant?

Problem 2: Subjective Interpretation

- "You said that students want healthier food options. How do you know?"

  - "We counted 47 different responses that asked for healthier food."

  - "Like what?"

  - "Here's one: 'Food at Price gives me a stomachache.'"

  - "Is that about health, or do $5 tacos induce financial stress?"

Problem 3: Bias

- "You started the survey because you wanted a Mexican restaurant, so it's not surprising that you found students wanted one."

  - "How many wanted an Italian restaurant?"

  - "Uh, I don't know."

  - "Aha. Your analysis was biased."

Structured Analytic Methods

- Want: *structured* analytic method to *reduce* and *reveal* bias

- Time for some epistemology…

Epistemology

- Wikipedia: **Epistemology** (from Ancient Greek ἐπιστήμη *(epistḗmē)* 'knowledge', and *-logy*), or the **theory of knowledge**, is the branch of philosophy concerned with knowledge.

- What does it mean to know something?

  - Positivism: facts derived from data with logic

  - Interpretivism [1]:

    - "the important reality is what people imagine it to be"

    - "systems exhibit emergent behaviors not reducible to their component parts"

[1] K.-J. Stol, P. Ralph, and B. Fitzgerald, "Grounded theory in software engineering research: a critical review and guidelines," in *Proceedings of the 38th International Conference on Software Engineering*, New York, NY, USA, May 2016, pp. 120–131. doi: 10.1145/2884781.2884833.

Positivism Vs. Interpretivism

- Positivism: typical of natural sciences

  - My experience: most computer scientists are positivists

- Interpretivism: typical of social sciences

  - Qualitative methods come from *this* community

Two Approaches (of Many)

- Grounded Theory

- Thematic Analysis

Grounded Theory (Partial)

- Limit exposure to literature

- Treat everything as data

- Immediate and continuous data analysis

- Theoretical sampling (acquire data to fill gaps so far)

- Theoretical sensitivity (researcher must establish relationships between concepts)

- Coding (infer categories from data)

- Memoing (write memos about analysis while doing it)

- Constant comparison: compare and revise categories and interpretations

Grounded Theory (2)

- Pick one:

  - Classic / Glaserian

  - Straussian

  - Constructivist

My Opinion

- Grounded theory is great if you do not have preconceived notions

- But apply carefully…

Thematic Analysis

- Phase 1: familiarizing yourself with your data

- Phase 2: generating initial codes

- Phase 3: searching for themes

- Phase 4: reviewing themes

- Phase 5: defining and naming themes

- Phase 6: producing the report

Phase 1: Familiarizing Yourself With the Data

- Transcribe data

  - "The time spent in transcription is not wasted, as it informs the early stages of analysis, and you will develop a far more thorough understanding of your data through having transcribed it."

- Read it all

- Write initial ideas

Phase 2: Generate Initial Codes

- "Code" here means "Category" (ugh)

- "Codes identify a feature of the data (semantic content or latent) that appears interesting to the analyst, and refer to 'the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon'" (Braun and Clarke)

Coding Example

Data extract	Coded for
it's too much like hard work I mean how much paper have you got to sign to change a flippin' name no I I mean no I no we we have thought about it ((inaudible)) half heartedly and thought no no I jus- I can't be bothered, it's too much like hard work. (Kate F07a)	1. Talked about with partner 2. Too much hassle to change name

Coding Guidelines

• Code as many potential themes as possible

  • But can keep research questions in mind

  • Items can have multiple codes

• Can keep it simple (spreadsheet) or use fancy software (MaxQDA, ATLAS.ti)

Phase 3: Searching for Themes

- Group codes into categories

- Merge duplicates

- Build hierarchies

Phase 4: Reviewing Themes

1. Read collated extracts for each theme and ensure they all fit the theme

2. Consider each theme in context of the whole data set: re-read data set, and add missing codes

Phase 5: Defining and Naming Themes

- Can you describe the scope and content of each theme in a few sentences?

- If not, perhaps this is not a theme.

Phase 6: Writing a Report

- Explain each theme

  - Justify with examples (quotes)

- Give quantities when helpful ("28% of participants felt food from their cultures of origin was not represented at Price")

Practice

- Use the file on Canvas.

- Work in pairs to give codes to each line.

  - This is a real transcript. You can break lines apart if you like.