I. The Course

A vast expanse of information about what people do, know, think, and feel lies embedded in texts—from transcribed interviews, news and the academic cannon to constitutions, poetry and idling online chat. With the rise of literacy, and more recently of computers, scanners and the internet, there are both greater supply and demand for textual information, providing us with access to a greater variety of texts that reach further into the human condition, past and present. This class will begin by considering the nature of texts and the possibilities for (and limits to) analyzing society and culture through textual content. We will spend several sessions cultivating skills for close, interpretive text analysis (or so-called “qualitative text analysis”/QCA). Then we will survey techniques for counting and verifying discovered meaning across texts (classic “content analysis”). Finally, we will explore recent approaches from natural language processing and information retrieval that facilitate the discovery and assessment of meaning on larger scales.

The course is designed to assist those seeking to enhance or expand a small-scale coding project (think interviews and field notes), those hoping to wisely summarize large samples of text, or those in search of tools to sift through quantities of documents to find those meriting a closer read. The course balances an empirical paper with experimentation using techniques. Students will be evaluated based on their participation, a presentation/report on a chosen method and a final paper.

II. Readings

Readings will be circulated in class and through the course’s chalk site.

III. Course Requirements

A. READING and DISCUSSION (10%)
facilitate this, I expect students to come to class prepared with: a) one discussion question related to the reading and b) one application or extension of the method discussed.

B. EXERCISES (30%)

Students are also expected to complete several small assignments intended to invite exploration with the techniques we read about. The first three involve qualitative coding. The final three involve “experimenting” with three quantitative approaches to coding/analysis. For these final three, students can either use existing programs or, if they have programming experience (or want to experiment), write their own and turn in a brief (double-spaced 3-5 page) report of the results and a brief evaluation. These experiments will briefly sketch the mechanics of using the method, present results from a preliminary (“toy”) analysis with it, evaluate its drawbacks and identify scope conditions for its beneficial use. These must be turned in within one week of the day we discuss the method. Students are encouraged to key assignments to the substance of the final project. Students will be scheduled to present empirical detail from one or two (depending on class size) of these investigations in class for discussion.

C. FINAL PROJECT (60%)

Students will write a 15-20 page final paper in which they will perform textual analysis for social insight based on tools developed and explored over the course of the quarter. The paper can be primarily qualitative or quantitative (or equally both). If qualitative, however, it must incorporate at least one quantitative technique developed through the quarter. If quantitative, it must validate its inferences with qualitative assessments.

IV. Calendar of Reading Assignments

Week 1.
Jan.6: General introduction
Jan.8: Semantics, semiotics and theory of meaning


See also:

Week 2.
Jan.13:  Open Coding (or Finding Things)


Note: I recommend, but do not require the use of atlas.ti for projects in the first few weeks. Free editions for limited size projects are available at: http://www.atlasti.com/demo.php

See also:

Jan.15: (cont.)


See also:

Coding texts (1)

Week 3.
Jan.20:  Playing with Typologies and Refining Codes


See also:


Jan.22: (cont.)
Week 4.

Jan.27: Playing with Relationships: Memos, visualizations, and formalizations


See also:
Miles and Huberman. Qualitative Data Analysis: An Expanded Sourcebook: Diagramming chapters throughout.


Jan.29: Selective Coding (or Counting Things)


See also:
Miles and Huberman. Qualitative Data Analysis: An Expanded Sourcebook: 262-287.


Week 5.

Feb.3: Sampling text


See also:


Coding texts (3)

Feb.5: Scaling up: using machines; creating corpora


See also:

**Pitch Texts/Context for Final Papers**

Week 6.

Feb.10: Words & Phrases


See also:


Feb.12: Words, Phrases & Lexical Semantics (College Break)

Manning and Schütze. 1999. Chapter 7 (word sense disambiguation); 8.5 (semantic similarity).


See also:


Week 7.

Feb.17: Syntax & Semantic Grammars


See also:


Feb.19: Parsing & Information Extraction


Week 8.
Feb.24: Dimension Reduction: Vector Space models and Latent Semantic Indexing

Manning and Schütze. 1999. Chapter 15 (“Topics in Information Retrieval”)


See also:
Manning, Christopher, Prabhakar Raghavan and Hinrich Schütze. 2008. “Scoring, term weighting and vector space model.” Chapter 6 from *Introduction to Information Retrieval.*


Feb.26: Clustering Documents & Text


See also:
Manning, Christopher, Prabhakar Raghavan and Hinrich Schütze. 2008. “Flat Clustering” and “Hierarchical Clustering.” Chapters 16 and 17 from *Introduction to Information Retrieval.*
Week 9.
Mar.3: Text & Document Classification

Manning and Schütze. 1999. Chapter 16 Text Categorization

See also:


Mar.5: Text & Document Networks (& inferential-AI)


See also:

Week 10.
Mar.10: Text & Document Networks 2


Mar.12: no class (reading day)
**Final Projects due Friday, Mar.19 at 5pm**
Selected Content Analysis Bibliography:

**Texts / Surveys:**


**Theory:**


**Qualitative Content Analysis Method:**


**Standard Content Analysis Method:**


**Practice:**


