Professor Scott Althaus  
Office: 256 Computer Applications Building  
Hours: Tuesdays 1–3pm, and by appt.

E-mail: salthaus@illinois.edu  
Web: www.illinois.edu/~salthaus  
Phone: 217.265.7879

Course Objectives

The objectives of this course are threefold:

1. to teach a generic and multi-purpose method of quantitative content analysis that is commonly employed by scholars of mass communication and political communication to measure trends and discourse elements in news coverage;
2. to give students practical experience in all stages of quantitative content analysis, from protocol design to validity testing, reliability testing, coding, data entry, and data analysis;
3. to produce publishable research papers co-authored by participants in the course (the number of these papers will depend on the number of students taking the course and the scope of the projects we choose to undertake).

The first half of the semester will be spent on learning the methods of content analysis, while the second half of the semester will be spent on applying them.

The Practicum Project

In order to focus on the practical aspects of content analysis and to complete a major data collection effort by the end of the semester, students will either pursue their own content analysis projects (which must be approved by the instructor) or participate in a group project headed by the instructor.

Those students wishing to pursue independent content analysis projects as part of the course must have their proposed projects approved by the instructor and must have a well-developed literature review, including hypotheses and a draft codebook for the content analysis, ready for submission soon after the start of the class. This means that doing your own project (by yourself or in combination with other collaborators in the class) that is independent of the main class project(s) requires more outside work than taking part in the instructor's group project.

Everyone else will work on a research project designed by the instructor in advance of the semester to address a research question that is of broad topical interest, both popularly and in scholarly circles, and that has never been subject to systematic quantitative content analysis. This project is designed to yield a high-profile academic publication. This semester's project will produce a validity study that compares several different computer data-mining strategies for coding the evaluative tone of news stories (that is, positive versus negative coverage). Done in collaboration with Kalev Leetaru (Coordinator of Information and Research Technologies at the Cline Center for Democracy at UIUC, and a PhD student in the Graduate School of Library and Information Sciences), the project will focus on New York Times coverage of the Israeli-Arab conflict from 1946 to 2004, and will seek to determine whether computer-based sentiment analysis methods can produce tone results comparable to those produced by human coders. We have unprecedented access to the entire full text of the
New York Times over this period, so the scale of this project will be truly impressive. It will not only yield valuable tests of the validity of such methods relative to human coding, but will also produce unprecedented insights into the dynamics of American news coverage about Israel and Palestine over the last 60 years. TARGET JOURNALS: American Journal of Political Science; Political Analysis; Journal of Communication.

Required Reading
Students are required to obtain the following book, which is available at local bookstores:

- Kimberly A. Neuendorf. 2002. The Content Analysis Guidebook. Thousand Oaks, CA: Sage Publications (one non-circulating copy is in the reserve stacks of the Education Library, call number 301.01 N393c) A Web site for this textbook can be found at: http://academic.csuohio.edu/neuendorf_ka/content/

Students are also required to obtain a set of additional readings for class, most of which are available in electronic form on the course Moodle.

Course Moodle Site
This course has a Moodle site that will be the primary vehicle for receiving course assignments and distributing course-related materials in electronic form. The Moodle site can be accessed here (course enrollment required for access): https://courses.las.illinois.edu/

Assignments
Your final grade for this course will be determined by your performance on the following assignments:

- Reliability analysis assignment (2-3 page paper) (10% of final grade)
- Validity analysis assignment (2-3 page paper) (10% of final grade)
- Weekly participation in class discussions and lab sessions (20% of final grade)
- Weekly content analysis case quotas (20% of final grade)
- Lexicon review paper and presentation (20% of final grade)
- Technical report paper and presentation (20% of final grade)

Students pursuing independent content analysis projects for course credit will turn in final papers that include the following sections: literature review, research questions and hypotheses, methods and data, findings, and conclusions. These papers will be worth 50% of the student’s final grade.

The reliability analysis assignment requires you to assess the reliability of variables in a hypothetical two-coder reliability test (using “fake” data generated from an actual coding scheme—hat tip to Kimberly Neuendorf for this assignment idea). You will calculate rates of intercoder agreement and a variety of intercoder reliability statistics for each of the variables in the data set. You will then outline concrete steps to improve the reliability of the data and recommend whether the hypothetical project should proceed to the data collection stage or should do another round of reliability testing.

The validity analysis assignment requires you to assess the methodological soundness and appropriateness (in terms of both validity and reliability) of a major content analysis project that is available to scholars and used in your area of research. This is not an assignment to analyze the coding used in a particular paper, but rather to analyze a major content analysis data source that is already in use by multiple scholars in your field. You should choose a dataset that’s relevant to you and that you might actually use or engage with down the road. Applying methodological concepts learned in class, you’ll review the research questions and hypotheses that the dataset
was designed to address, and assess the quality of sampling, the operationalization of key content analysis variables, as well as the reliability of the coding scheme. Possible examples include the Wisconsin Advertising Project (http://wiscadproject.wisc.edu/); the Policy Agendas Project (http://www.policyagendas.org/); the Kansas Event Data System (http://web.ku.edu/~keds/); the Comparative Manifesto Project (http://www.wzb.eu/zkd/dsl/Projekte/projekte-manifesto.en.htm), Media Tenor (http://www.mediatenor.com/), and the Project for Excellence in Journalism’s News Coverage Index (http://www.journalism.org/by_the_numbers/datasets).

The **lexicon review paper and presentation** is designed to support the class project, and will be written by groups of students. The paper and accompanying PowerPoint presentation will have four goals: (1) review and explain the analytical strategy of a popular sentiment lexicon (to be assigned by the instructor); (2) identify the theoretical assumptions about communication that inform the lexicon’s use; (3) review the academic literature that has used the lexicon; and (4) assess the validity of the lexicon’s analytical strategy for sentiment analysis in political science and communication research. The assigned lexicons will later be tested against our class’s human coding and compared against the parallel results of other sentiment lexicons.

The **technical report paper and presentation** presents the results of an empirical test comparing human coding to an assigned lexicon, where both sets of results are drawn from the same underlying data. This assignment follows directly from the literature review assignment, and groups will retain the same lexicons for both assignments. The paper and accompanying PowerPoint presentation will have three goals: (1) assess how accurately the lexicon reproduces the tone scores made by of human coders; (2) to the degree that the lexicon misrepresents the tone score of human coders, assess how the lexicon’s analytical strategy may have been responsible for producing particular errors; (3) make recommendations on how to adjust the lexicon’s analytical strategy to better match the results of human coding.

**Authorship and Data-Sharing Agreement**

For those participating in the course’s collaborative data collection project, the following guidelines determine authorship credits for the specific projects detailed above:

- **Default order of authorship credit:** Kalev Leetaru and the course instructor will be listed as first authors (order to be determined), followed by each student (in alphabetical order) who participated materially in the collection of data, the analysis of data, or manuscript preparation for the particular project.

- **Modifications to this default order of authorship credit** can be made with the consent of the majority of persons claiming authorship in the project’s publication. Reasons for modifying the order of authorship include, but are not limited to: recognizing additional data collection effort above and beyond those of other co-authors, recognizing the contribution of special skills or additional time above and beyond those contributed by other co-authors, and recognizing additional effort given to preparing or revising the manuscript for publication. It is common for some participants to contribute only to data collection, while others go on to also put extra work into the manuscript and publication process—these differences in material contributions normally should be factors influencing the order of authorship. Authors can also be dropped from sharing credit in the project if the data they supply turns out to be flawed or otherwise unusable due to problems in the quality of data collection.

In addition, students taking this course and participating in its data collection project(s) also agree that once each research project (as defined above) yields its intended publication, all data collected for the project can be used freely by any of the project co-authors, who agree to cite the authors of the original publication as the source of the data. This means that every project participant is guaranteed authorship credit on the first publication, but not on any subsequent publications that might result from this data collection effort. This is
meant to encourage students to add further data to the cases or to use the original data for different projects without having to involve or get individual permission from all of the original co-authors.

### Tentative Course Schedule

<table>
<thead>
<tr>
<th>Tuesdays</th>
<th>Thursdays</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>8/24</strong></td>
<td><strong>8/26</strong></td>
</tr>
<tr>
<td>Overview of Course and Introduction to Content Analysis Projects</td>
<td>The History, Applications and Analytical Logic of Content Analysis</td>
</tr>
<tr>
<td></td>
<td><em>In-class:</em> begin discussing sampled news stories to define tone and the two actors for which tone will be assessed</td>
</tr>
<tr>
<td><strong>8/31</strong></td>
<td></td>
</tr>
<tr>
<td>What Is Content Analysis Trying to Do?</td>
<td></td>
</tr>
<tr>
<td><em>In-class:</em> discuss assigned readings; finish discussing the first batch of sampled news stories to define tone and the two actors for which tone will be assessed</td>
<td><em>9/2</em></td>
</tr>
<tr>
<td></td>
<td>Before class: read the second batch of sampled news content with the draft coding scheme in mind</td>
</tr>
<tr>
<td></td>
<td><em>In-class:</em> discuss the sampled news content to define tone and the two actors for which tone will be assessed</td>
</tr>
<tr>
<td><strong>9/7</strong></td>
<td></td>
</tr>
<tr>
<td>Validity I: Generalizability</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>9/9</em></td>
</tr>
<tr>
<td></td>
<td>Before class: read the third batch of sampled news content with the draft coding scheme in mind</td>
</tr>
<tr>
<td></td>
<td><em>In-class:</em> refine draft codebook</td>
</tr>
<tr>
<td><strong>9/14</strong></td>
<td></td>
</tr>
<tr>
<td>Validity II: Operationalization</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>9/16</em></td>
</tr>
<tr>
<td></td>
<td>In-class: Finalize codebook; determine sampling frames for content analysis project; draw sample for first reliability test</td>
</tr>
<tr>
<td><strong>9/21</strong></td>
<td></td>
</tr>
<tr>
<td>Validity III: Reliability</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>9/23</em></td>
</tr>
<tr>
<td></td>
<td>In-class: Discuss results of first reliability test. If necessary, refine coding scheme, and draw new sample; otherwise begin coding</td>
</tr>
<tr>
<td><strong>9/28</strong></td>
<td></td>
</tr>
<tr>
<td>Validity IV: Databases and Content Proxies</td>
<td></td>
</tr>
<tr>
<td>Reliability analysis assignment due</td>
<td><em>9/30</em></td>
</tr>
<tr>
<td></td>
<td>In-class: Work on second reliability test</td>
</tr>
<tr>
<td><strong>10/5</strong></td>
<td></td>
</tr>
<tr>
<td>Introduction to Data Mining Approaches</td>
<td></td>
</tr>
<tr>
<td><em>In class:</em> Discuss results of second reliability test. If necessary, refine coding scheme, and draw new sample; otherwise begin coding</td>
<td><em>10/7</em></td>
</tr>
<tr>
<td></td>
<td>In-class: Interesting examples session</td>
</tr>
<tr>
<td>Date</td>
<td>Topic</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>10/12</td>
<td>Sentiment Analysis I</td>
</tr>
<tr>
<td></td>
<td>Validity analysis assignment due</td>
</tr>
<tr>
<td>10/19</td>
<td>Sentiment Analysis II</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>10/26</td>
<td>Network Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>11/2</td>
<td>Automatic Topic Categorization</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>11/9</td>
<td>Lexicon Review Presentations</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>11/16</td>
<td>Comparative Lexicon Performance Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>11/23</td>
<td>NO CLASS—THANKSGIVING BREAK</td>
</tr>
<tr>
<td>11/30</td>
<td>Special Topic (To Be Determined)</td>
</tr>
<tr>
<td>12/7</td>
<td>Special Topic (To Be Determined)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Reading List**

8/26  *The History, Applications and Analytical Logic of Content Analysis*

CAG chapters 1, 2, 3, and 9

8/31  *What Is Content Analysis Trying to Do?*


9/7  **Validity I: Generalizability**

CAG chapter 4 ("Message Units and Sampling")


**For further reading:**


9/14  **Validity II: Operationalization**

CAG chapters 5 and 6


**For further reading:**


9/21  **Validity III: Reliability**

CAG chapter 7 ("Reliability")


Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2002. Content analysis in mass


**Useful Web resources:**

Andrew Hayes’ Kalpha Macro for SPSS and SAS: http://www.comm.ohio-state.edu/ahayes/SPSS%20programs/kalpha.htm

Deen Feelon’s ReCal System for Calculating Intercoder Reliability: http://dfreelon.org/utils/recalfront/

Matthew Lombard’s Intercoder Reliability Page: http://astro.temple.edu/~lombard/reliability/

PRAM Intercoder Reliability Software Documentation (Software will be distributed from the course Moodle): http://academic.csuohio.edu/neuendorf_ka/content/pram.html

**For further reading:**


**9/28 Validity IV: Databases and Content Proxies**

**Databases**

CAG resource 2 (“Using Nexis”)


**News Indices**


**News Abstracts**


**10/5 Introduction to Data Mining Approaches**


**Useful Web sites:**

CAG resource 3, “Computer Content Analysis Software” and software links available in the Content Analysis Guidebook Online web page at [this location](http://www.snsoroka.com/lexicoder/resources.html)


Amsterdam Content Analysis Toolkit (AmCAT) Software: [http://content-analysis.org/](http://content-analysis.org/)


Software Environment for the Advancement of Scholarly Research (SEASR) Software: [http://seasr.org/](http://seasr.org/)

**For further reading:**


**10/12 Sentiment Analysis I**


10/19 **Sentiment Analysis II**


10/26 **Network Analysis**


11/2 **Automatic Topic Classification**


