What is Leximancer

Automatic Content Analysis

Dr Andrew Smith
“Leximancer's entire purpose is to let the data generate a transparent model which can be interpreted by the analyst, so that this person may efficiently conduct a sense making examination of conceivably vast amounts of text...”
What is Content Analysis

• A research tool used for determining the presence of words or concepts in collections of textual documents.

• Used for breaking down the material into a finite number of categories and relationships in order to quantify and visualise text.

• Required for building a reproducible and computable model of a very complex and intuitive information space.
Decide what proportion of the population of relevant content to sample, and how to sample.

Decide what segment of text to use as the basic unit of analysis. For example, Leximancer uses 2 sentence segments by default.

Decide what constitutes each of your concepts. This can be defined by a dictionary, or a code book, or can be found using an algorithm, such as latent semantic indexing or Leximancer.
An Example of a Latent Concept

An example of a latent concept thesaurus.

The system identified the single term rifle as being central to the discourse.

This seed word is then augmented with other directly and indirectly related terms and names from this data set.

The resulting weighted term list is employed as a classifier which automatically codes text segments.

Inspection shows various parts of a rifle as well as various types of rifle, and some associated names.
Mechanisms of Content Analysis

- **Code** (or classify, tag, or even mark) each text segment using the Code Book as a guide to decide what concepts are present in each segment.
- Count the code occurrences and co-occurrences.
- Assess the inter-coder reliability of your coders (for human content analysis), and measure the cross validity or stability of the final model.
- Make sense of the results, in the context of your research question.
Pride and Prejudice
by Jane Austen.

Chapter 4

When Jane and Elizabeth were alone, the former, who had been cautious in her praise of Mr. Bingley before, expressed to her sister just how very much she admired him.

('family' => '13.26','female' => '11.9')

"He is just what a young man ought to be," said she, "sensible, good-humoured, lively; and I never saw such happy manners! -- so much ease, with such perfect good breeding!"

('feeling' => '8.5','manners' => '7.55','male' => '9.35')

"He is also handsome," replied Elizabeth, "which a young man ought likewise to be, if he possibly can. His character is thereby complete."

('male' => '6.2')
Types of Content Analysis

**01 Conceptual Analysis**
Frequency counting.
A linear model, and not very expressive or contextualised.

**02 Relational Analysis**
Co-occurrence counting.
Second order.
More expressive and capable of sense disambiguation, but still rather disembodied.

**03 Cognitive Analysis**
Projection mapping.
Highly non-linear and very contextualised.
Visually oriented sense making → **Concept Map**
A Concept Map

A Content Analysis Engine

Leximancer is nothing more than a system which performs *most* of the steps of content analysis for the user.

You must still at least:

- Decide on your research question
- Select your data sample
- Make sense of the final model.
A Content Analytic Visualisation

Concepts and themes as building blocks to assemble a model of the meaning
<table>
<thead>
<tr>
<th>Application</th>
<th>Text Type</th>
<th>Outputs</th>
<th>Possible Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>Publications</td>
<td>Concept Map, Thematic Summary</td>
<td>Literature review, Characterisation of a field</td>
</tr>
<tr>
<td>Survey</td>
<td>Spreadsheet with mixed categorical and open ended</td>
<td>Category map and thematic summaries</td>
<td>Customer insight, qualitative research</td>
</tr>
<tr>
<td>Media</td>
<td>Media reports, Hansard, social, company data</td>
<td>Topical clustering of channels, thematic analysis</td>
<td>Issue analysis, environment mapping</td>
</tr>
</tbody>
</table>
If you are interested in learning more about content analysis, we recommend the following book:

This paper is sets out to validate Leximancer as a content analysis tool, using a comprehensive taxonomy of validation from the literature:


Some strong examples of how to use Leximancer to perform automatic content analysis:


Thank You!

Begin your journey here, free trial. Contact us: www.leximancer.com +61 434 506 651