



What is Leximancer

Automatic Content Analysis

Dr Andrew Smith

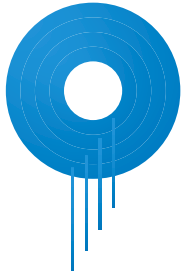


“Leximancer's entire purpose is to **let the data generate a transparent model** which can be interpreted by the analyst, so that this person may **efficiently conduct a sense making examination of conceivably vast amounts of text...**”

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.

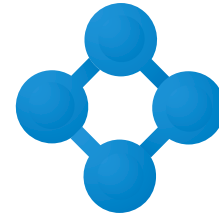
Mechanisms of Content Analysis



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.

Defining a concept

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.

```
rifle:
rifle -> 8.5763
[[c2766]] -> 5.192
telescopic -> 4.9546
ammunition -> 4.7806
serial -> 4.7806
[[mannlicher-carcano]] -> 4.6772
lifted -> 4.5586
[[Klein]] -> 4.4189
[[c2766_mannlicher-carcano]] -> 4.4189
millimeter -> 4.3385
wooden -> 4.3385
disassembled -> 4.3385
[[i.s.s]] -> 4.2489
[[western_cartridge]] -> 4.2489
[[carcano]] -> 4.2489
[[irving_sports_shop]] -> 4.2489
[[iss]] -> 4.2489
tuft -> 4.1475
bolt-action -> 4.1475
stored -> 4.1475
[[price]] -> 4.1475
[[slack]] -> 4.1475
[[mannlicher]] -> 4.0308
[[mauser]] -> 4.0308
sling -> 4.0308
butt -> 4.0308
[[a.hidell]] -> 4.0308
[[mannlicher-carcano_6.5]] -> 4.0308
rifleman -> 4.0308
underside -> 3.8931
[[cal]] -> 3.8931
wood -> 3.8931
four-power -> 3.8931
orange-yellow -> 3.8931
commit -> 3.8931
clip -> 3.8931
[[oswald_s_mannlicher-carcano]] -> 3.8931
[[lieutenant_day]] -> 3.8931
-millimeter -> 3.725
```

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')

Coded Text

- an example -

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

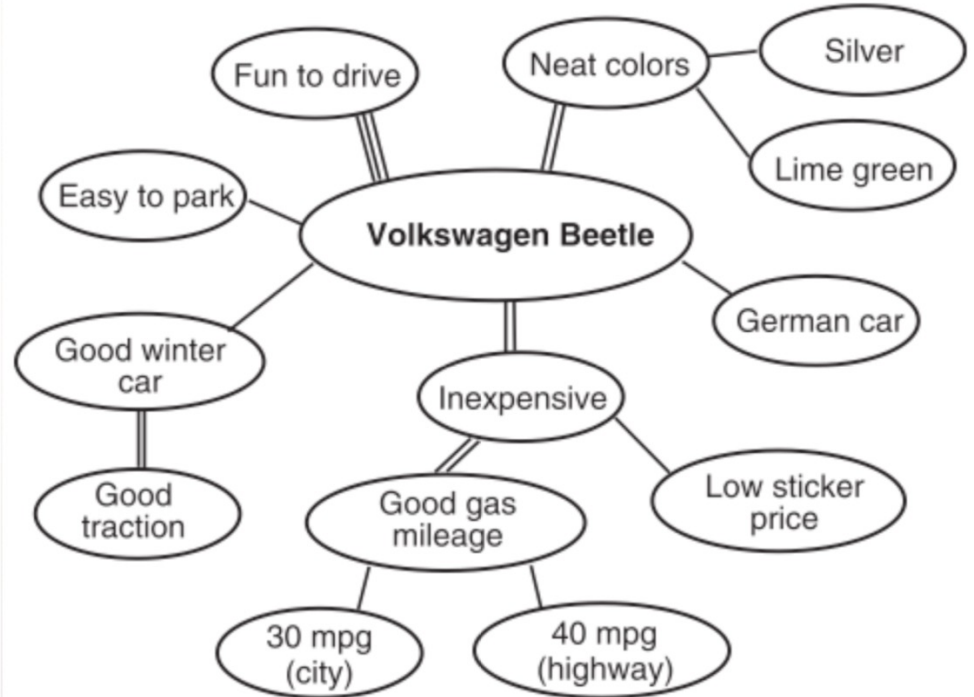
Cognitive Analysis

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

Cognitive Mapping Visualisation

A Concept Map

John, Loken, Kim, & Monga,
Journal of Marketing
Research, Vol. XLIII
(November 2006), 549–563)



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



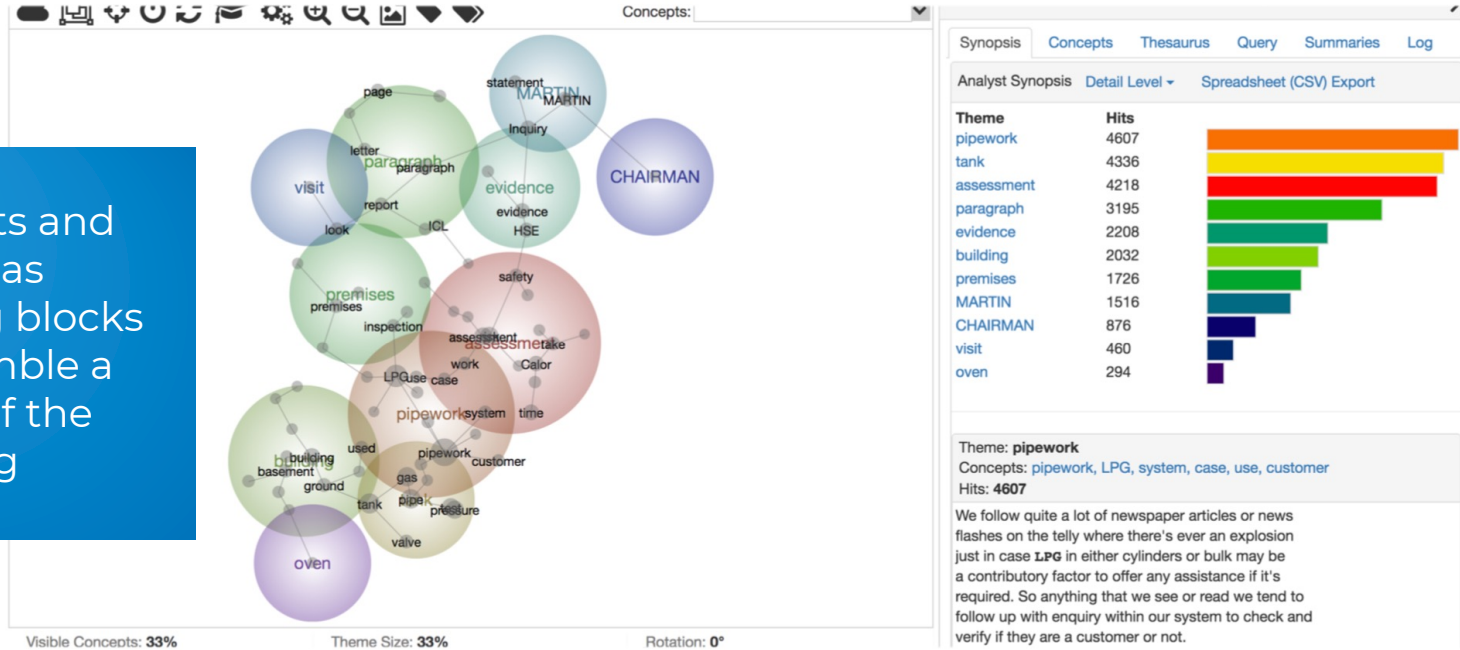
Select your
data sample



Make sense of
the final model.

A Content Analytic Visualisation

Concepts and themes as building blocks to assemble a model of the meaning



Application	Text Type	Outputs	Possible Projects
Research	Publications	Concept Map, Thematic Summary	Literature review, Characterisation of a field
Survey	Spreadsheet with mixed categorical and open ended	Category map and thematic summaries	Customer insight, qualitative research
Media	Media reports, Hansard, social, company data	Topical clustering of channels, thematic analysis	Issue analysis, environment mapping.

Future Reading

If you are interested in learning more about content analysis, we recommend the following book:

Weber, R.P. (1990) Basic Content Analysis. Newbury Park, Calif.: Sage Publications, 2nd ed.

Content Analysis with Leximancer

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

A. E. Smith and M. S. Humphreys (2006). Evaluation of Unsupervised Semantic Mapping of Natural Language with Leximancer Concept Mapping. *Behavior Research Methods*, 38 (2), 262-279.

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

Cheng, M., & Edwards, D. (2017). A comparative automated content analysis approach on the review of the sharing economy discourse in tourism and hospitality. *Current Issues in Tourism*. Advance online publication. doi: 10.1080/13683500.2017.1361908

Nunez-Mir, G. C., Iannone, B. V., Pijanowski, B. C., Kong, N., & Fei, S. (2016). Automated content analysis: addressing the big literature challenge in ecology and evolution. *Methods in Ecology and Evolution*, 7(11), 1262-1272. doi:10.1111/2041-210X.12602



Thank You!

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