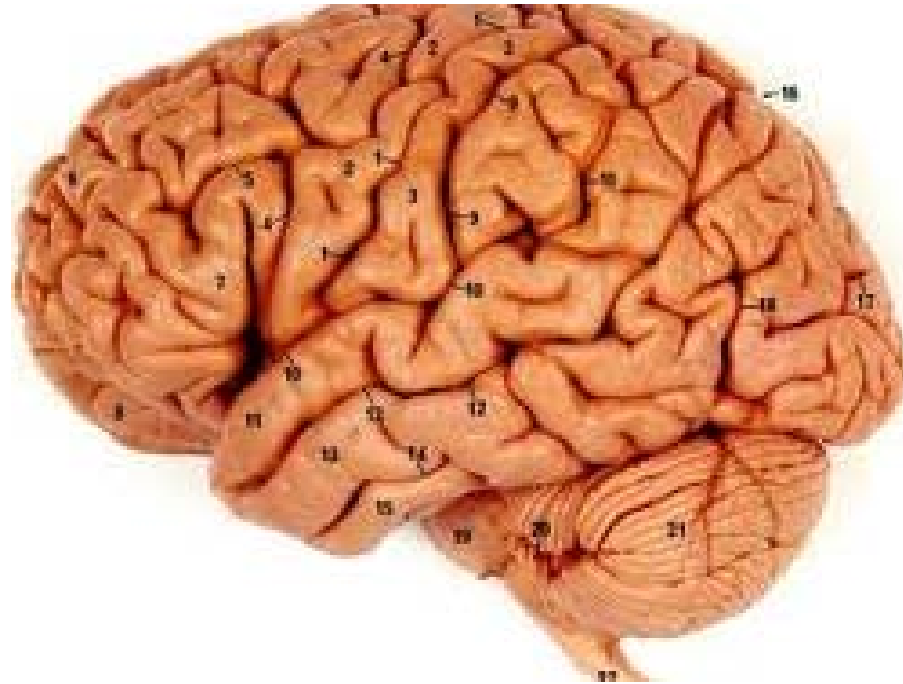




UniSA

QUALITATIVE DATA ANALYSIS – COULD I USE NVIVO OR LEXIMANCER?

Bruce Gurd



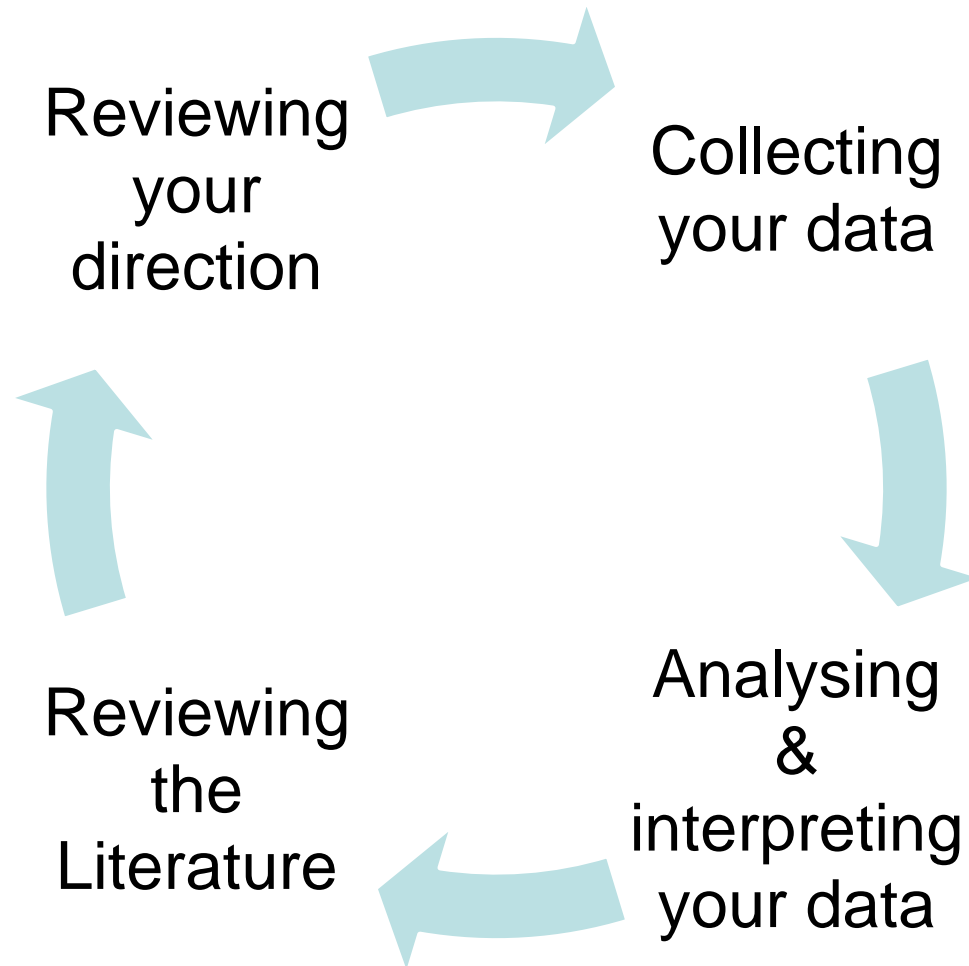
The best QDA tool

- May not be testing a hypothesis; building a theory from the data
- May commence analysis during data collection
- Less structured approaches (not less rigorous!)
- Research design may change during the project

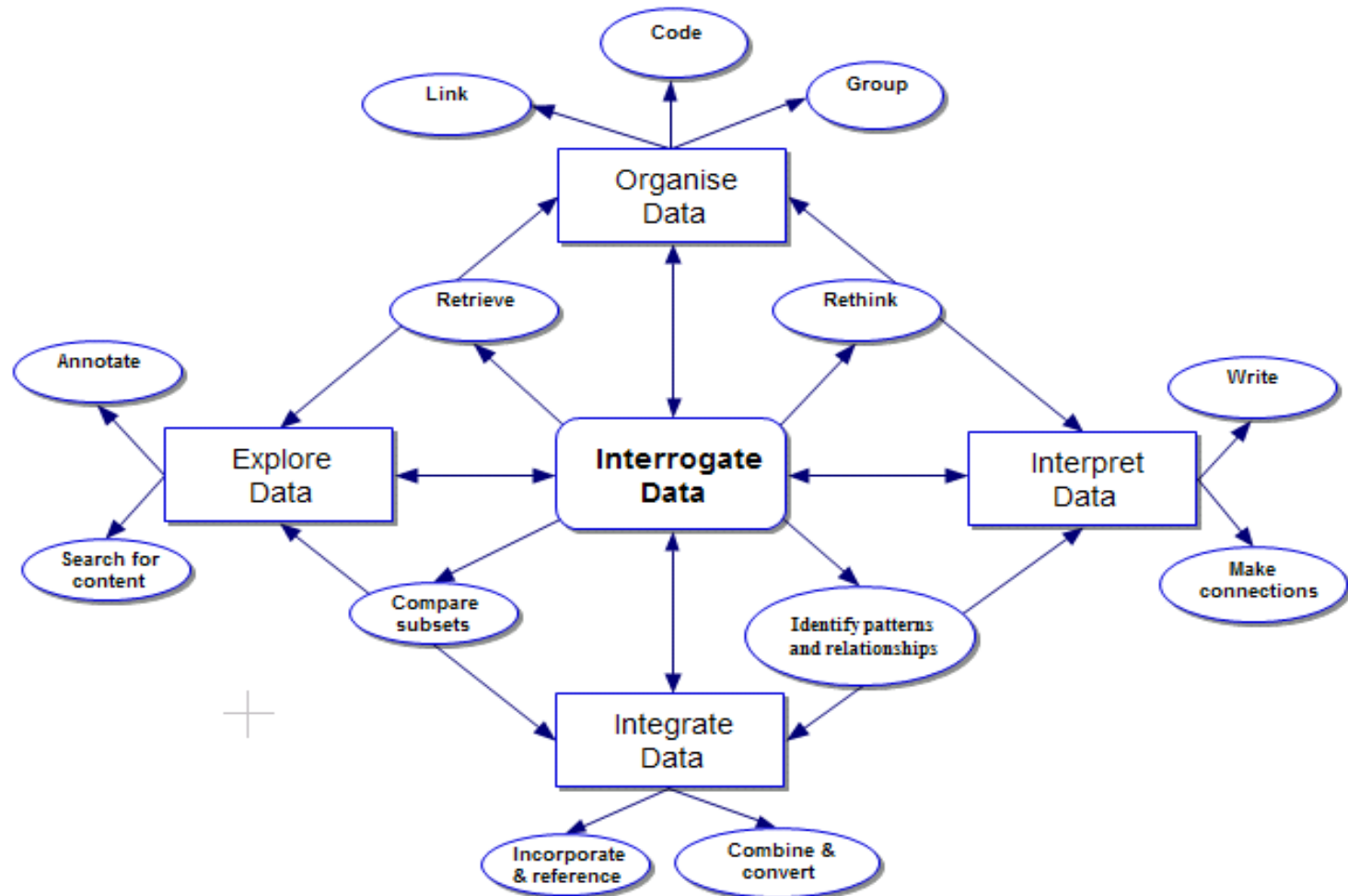
- Need to have a clear research design
- Care with data management
- Clear understanding of a rigorous process of data analysis
- Ethical issues

Where most of the work is done!!!

UniSA



Using Software in Qualitative Research (Lewins and Silver)



- Code and retrieve
- Text/Word searches
- Data organisation (e.g. socio-demographics)
- Searches of position of codes in data (e.g. co-occurrence, proximity etc.)
- Writing tools (memos, comments etc.)
- Output of coded segments, results of searches etc..

Silver's list of features

- Methodological underpinnings – sociology, psychology
- Software Architecture
- Textual data formatting
- Multi-media data
- Coding schema structures
- Margin display
- “Closeness” to data
- Beyond “code & retrieve”
- Querying, searching & auto-coding possibilities
- Output option

Data management

- “How shall I find a method to analyze 1,000 pages of interview transcripts?”
- Need appropriate systems for recording and filing data.
- Data needs to be kept for 5 years.



Qualitative Data Analysis



Benefits

- Efficiency with data management
- Allows line by line coding
- Memoing in course of analysis
- Create some structure out of codes e.g. tree structure in Nvivo

Costs

- claims to standardisation and reduction of diversity in approaches
- encouraging too much data and compromising depth for breadth. The increased efficiency of the computer encourages using higher volumes of data but may in fact reduce the researchers contact with the data.
- loss of context. Packages may lift text out of context and the researcher loses track of how the codes and concepts relate to the context of the data from which they were drawn.



- apparent claims to be more scientific. The use of a more structured approach clearly does not improve the “scientific” nature of the analysis. The coding and analysis processes may be more structured and efficient but the codes are still constructs of the researcher.

NVIVO

NVivo often includes

- Data records
 - Records of your thinking about the data (memos)
 - Coding items which show references to your data (so you can retrieve what you know about a topic, idea or relationship)
 - Variable type information about the case (demographics)
 - Results of interrogation of your data
 - Models showing relationships
- (Bazeley, QDA with Nvivo, 2007, p.26)

Lyn Richards 7 deadly sins

1. Software won't think
2. It distances researchers from data
3. It (only) does grounded theory
4. It's rigid, traditional method is flexible
5. It forces researchers to code
6. It forces top-down thinking
7. It is responsible for unplanned projects by unskilled researchers

Coding

- Coding from literature or in vivo
- Coder reliability
- Forming models or connections
- Exploring codes against base data

Slides kindly provided by Leximancer's
Julia Cretchley

LEXIMANCER

What does Leximancer do?

- Leximancer is text analytic software designed for analysing the conceptual content of natural language text data
- The program analyses text and displays the extracted information visually, in the form of a 'concept map'
- The concept map displays the main concepts in the text data, and depicts the relationships among concepts
- Using the concept map, the user can develop hypotheses and perform a directed search of the text
- In this way, Leximancer illustrates the conceptual structure of the texts

- Leximancer is able to efficiently quantify and explore large text documents based on a classification system of learned lexical concepts rather than only keywords (Grech, Horberry, & Smith, 2002).
- Extracted information is presented in a visual form known as a conceptual map (Leximancer Manual, 2009).
- Initially Leximancer examines the documents and selects a ranked list of important lexical terms based on word frequency and co-occurrence, referred to as concepts (Smith & Humphreys, 2006; Martin & Rice, 2007).
- Results are produced based on algorithms using a Bayesian approach to prediction (Davies, *et al.*, 2006; Smith & Humphreys, 2006; Martin & Rice, 2007; Chen & Bouvain, 2009; Leximancer Manual, 2009).

Applications

- Leximancer can assist in extracting meaning from voluminous and disparate bodies of texts.

Some examples include:

- Gauging changes in the tone and content of news articles and reports
- Coding free-text comments in staff or client surveys
- Analysing customer sentiment in forums and on-line reviews
- Researching group differences in culture or communication style
- Exploring the issues raised in interview or focus group transcripts

Approach


- Leximancer generates concepts that characterise the text data using information about the co-variation between high-frequency words
- The program extracts its own concepts from texts using statistical algorithms, and creates a specialised thesaurus of terms for each concept
- The user can also explore specific issues by making suggestions for concepts.

Advantages

- A specialised dictionary is built for each concept
 - concept definitions are well-grounded in the data, and
 - the application of inappropriate taxonomies is avoided
- The user need not formulate a coding scheme
 - this saves time, and
 - avoids the introduction of bias based on assumptions
- The coding can be done with a computer
 - this is efficient and reliable, and facilitates broader opportunities for interrogating the text.

Phases of Processing

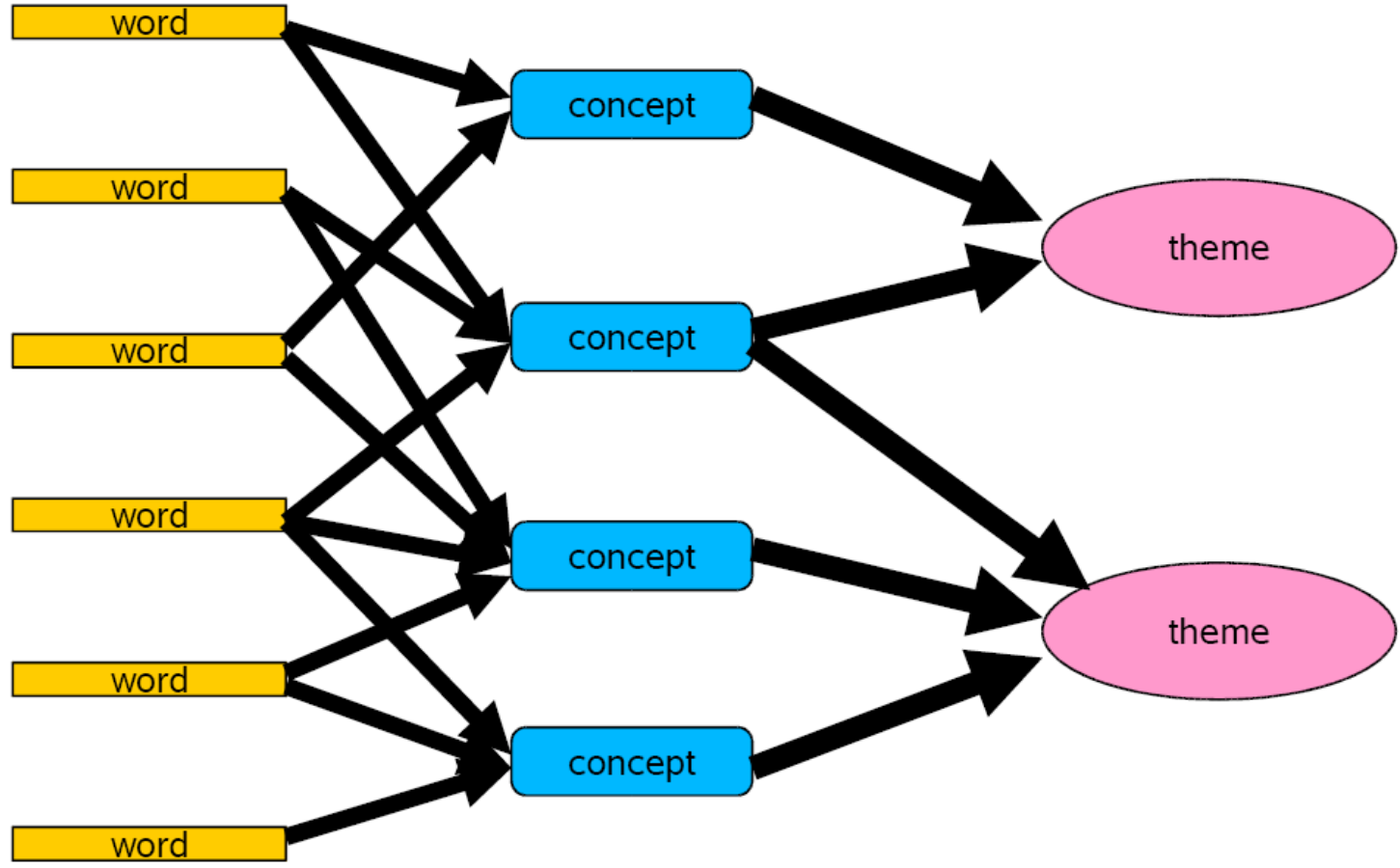
Project Control: ICL Explosion Enquiry

	Stage		Status
	<input type="button" value="Load Data"/>		● Ready
	<input type="button" value="Generate Concept Seeds"/>	<input type="button" value="+ Show Settings"/>	● Ready
	<input type="button" value="Generate Thesaurus"/>	<input type="button" value="+ Show Settings"/>	● Ready
	<input type="button" value="Run Project"/>	<input type="button" value="+ Show Settings"/>	● Ready

Pre-processing

- In this phase, Leximancer does the following:
 - Splits the information into sentences, paragraphs and documents
 - Removes non-lexical and weak semantic information ('stopword' list)
 - Identifies proper nouns and multi-word names
 - Performs a language test on each sentence

Concept Identification



Concept Editing

- Here the user can check and accept Leximancer's suggestions for (automatically-extracted) concept seeds
 - This is an exploratory approach that produces concepts that are completely grounded in the text,

OR

- The user can seed concepts manually (themselves) using a theory-driven approach
 - This allows users to define the concepts they are interested in quantifying
 - This approach enables users to test hypotheses and investigate specific issues

Thesaurus Development

- In this phase, Leximancer develops tailored definitions for the concept 'seeds' (starting words) contained in the concept editor
- The program generates a list of weighted terms (a thesaurus) that constitute evidence for the presence of each concept
- High-relevance evidence words occur commonly in contexts where the concept is discussed in the text, and rarely where it is not
- The words in a sentence must constitute sufficient evidence for a concept to be coded there
- A strict algorithm is applied in determining concept membership
- Sentences may contain several concepts.

Example Thesaurus

- This is the list of automatically-learned thesaurus terms for the *building* concept in the transcripts of a 12-day enquiry into an explosion at a plastics factory:
 - building 6.77
 - warrant 5.71
 - tower 5.42
 - collapsed 4.85
 - chimney 4.74
 - stability 4.68
 - pitched 4.61

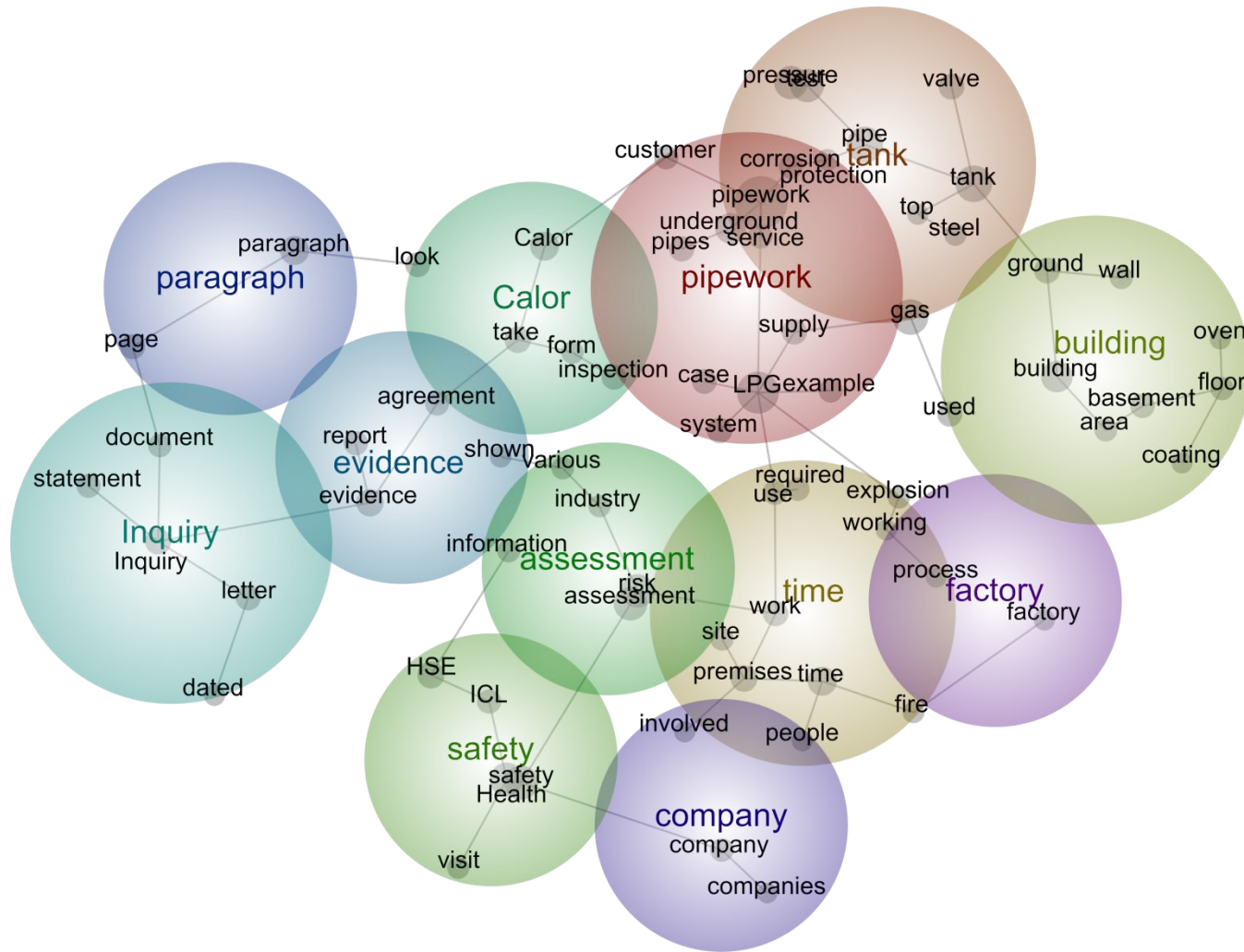
Concept Classification

- This phase is similar to manual coding in content analysis
 - The text is broken into blocks of 2 sentences (by default)
 - The blocks are tagged with the concepts that they contain
 - The co-occurrence between concepts is measured.

The Concept Map

- Leximancer uses concept frequency and co-occurrence data to compile a matrix of concept co-occurrences
- A statistical algorithm is then used to create a two-dimensional concept map based on the matrix
- Initially, concepts are dispersed randomly in the map space. Then the relationships between concepts act like attractive forces to guide concepts to their resting places
Concepts clustered together on the map if they are related, or tend to occur in similar contexts, thus the map offers information on the conceptual structure of the text
- The user can check the stability of the map (to assess reliability) by duplicating the map, randomising and re-clustering the concepts, then comparing between map copies.

Example Map



Interpreting the Map

- The map offers five important sources of information about the text
 - The main concepts
 - The relative frequency of concepts
 - The co-occurrence of concepts
 - The centrality of concepts
 - Similarity of the contexts in which concepts occur (conceptual clustering)

- The concept map is linked to a text browser
- From the browser you can drill down into the text to:
 - Learn what the concepts refer to
 - Investigate the nature of concept relationships

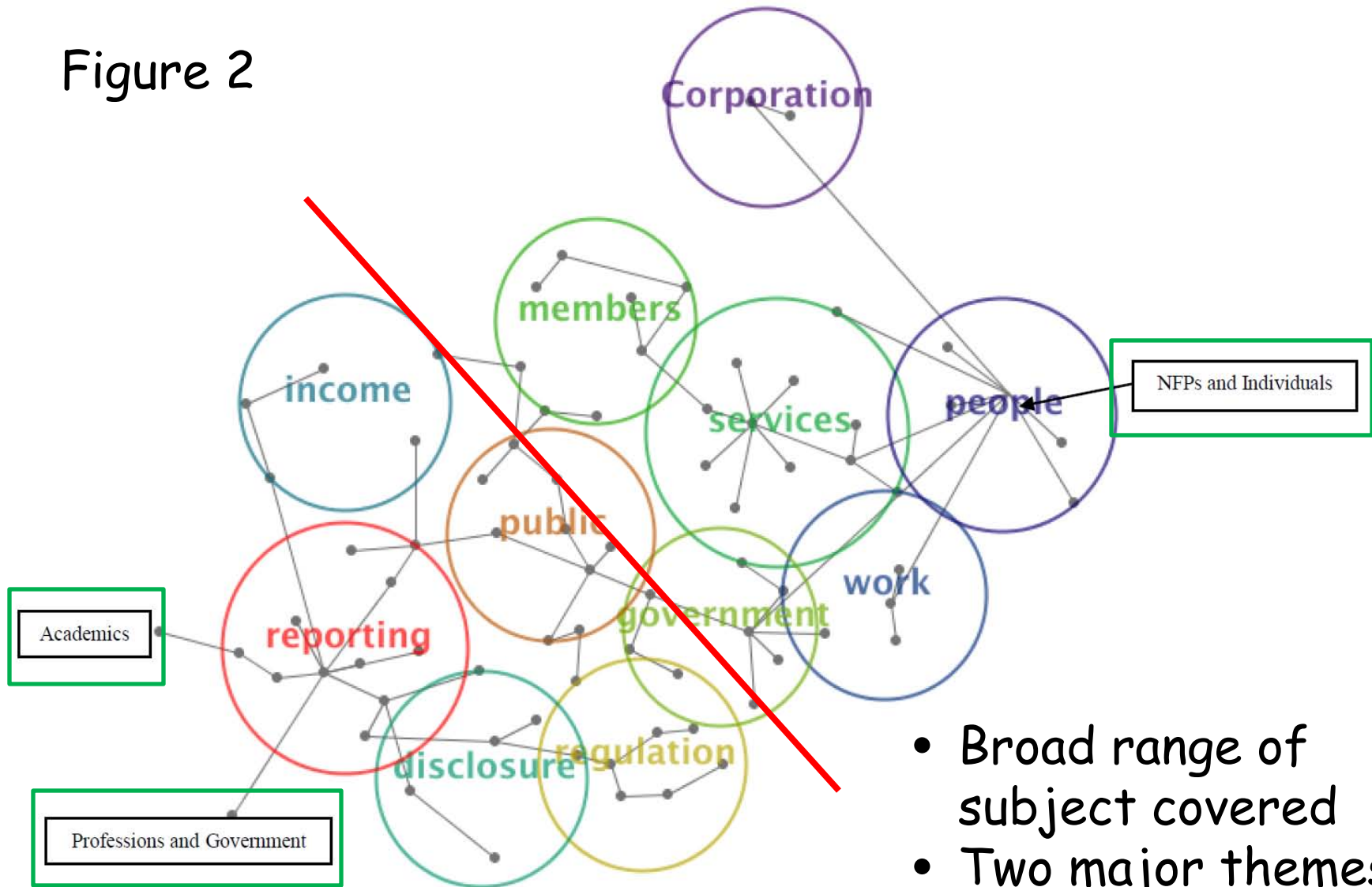
Summary

- Leximancer is a software tool for analysing the conceptual content of (natural language) text data
- By extracting semantic and relational information, Leximancer derives the key concepts and themes
 - Leximancer can extract concepts automatically to summarise the text, or
 - researchers can *seed* their own indicative concepts for confirmatory analysis.
- Leximancer uses the co-occurrence of words to produce a concept map that displays the data and shows interrelations among the main concepts
 - In this way, Leximancer summarises the conceptual structure of the text.

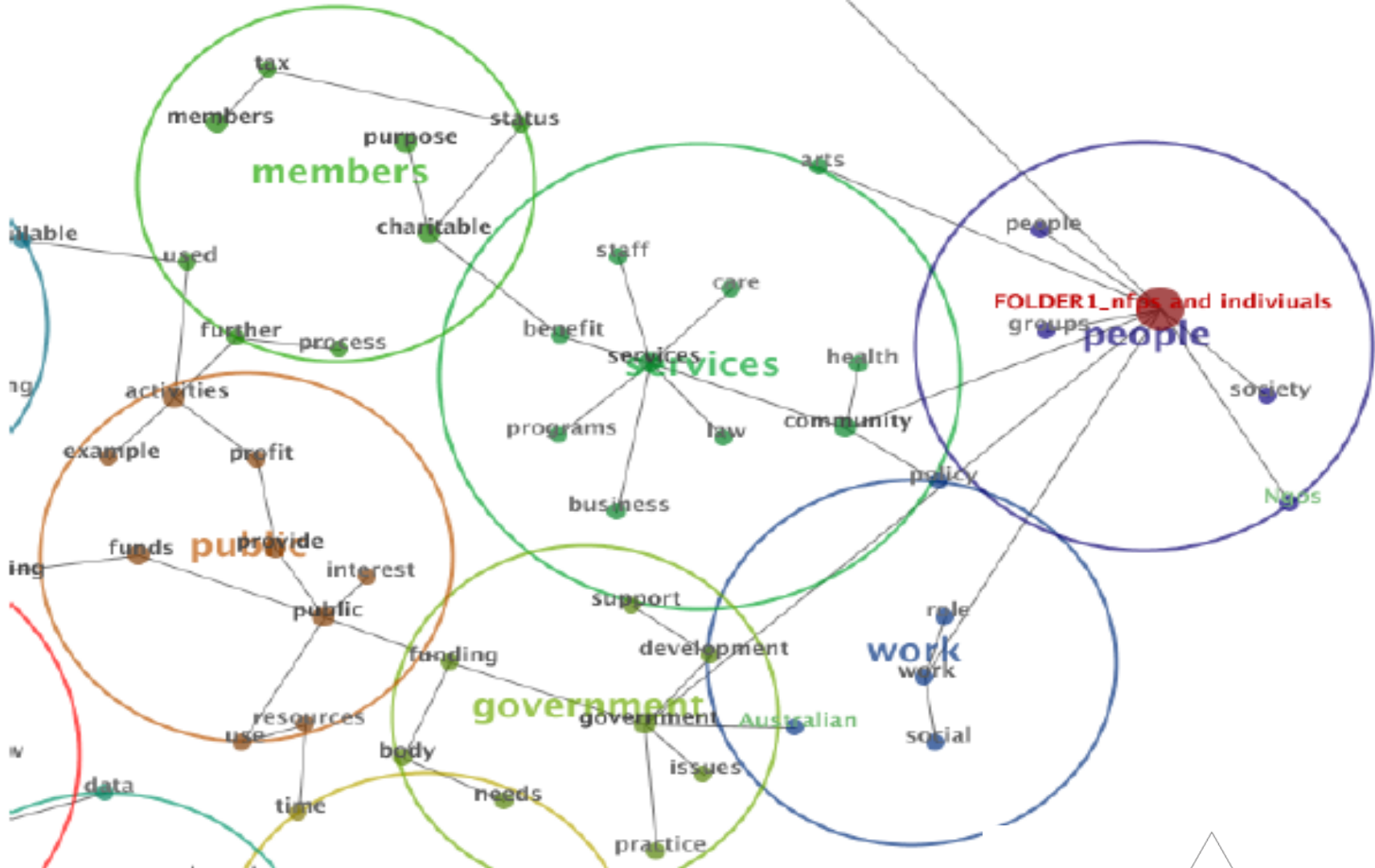
AN EXAMPLE

- Considered the submissions made to the 2008 Australian Senate Economics Committee in response to the inquiry into the *Disclosure regimes for charities and no-for-profit organisations*
- Respondents to the Senate inquiry may not be representative of the whole NFP sector and that their responses may not be as candid as those provided in an interview setting.
- Responses are from those entities interested enough in the issue to write a submission and submit it to the inquiry.
- Responses cover a broad range of NFP entities; small and large, from a range of sectors (e.g. education, health, social services) and from a range of different geographical locations in Australia as well as individuals, academics, government departments and professional bodies.
- The sampling bias is towards those NFPs and individuals that are committed enough to respond to the call for submissions.

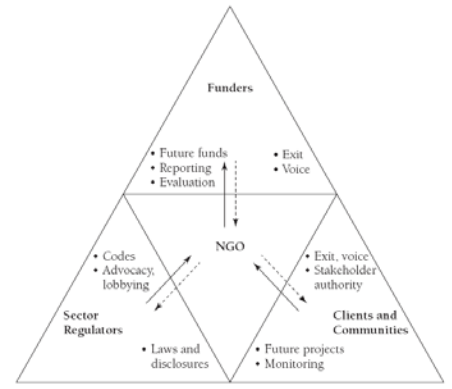
Figure 2



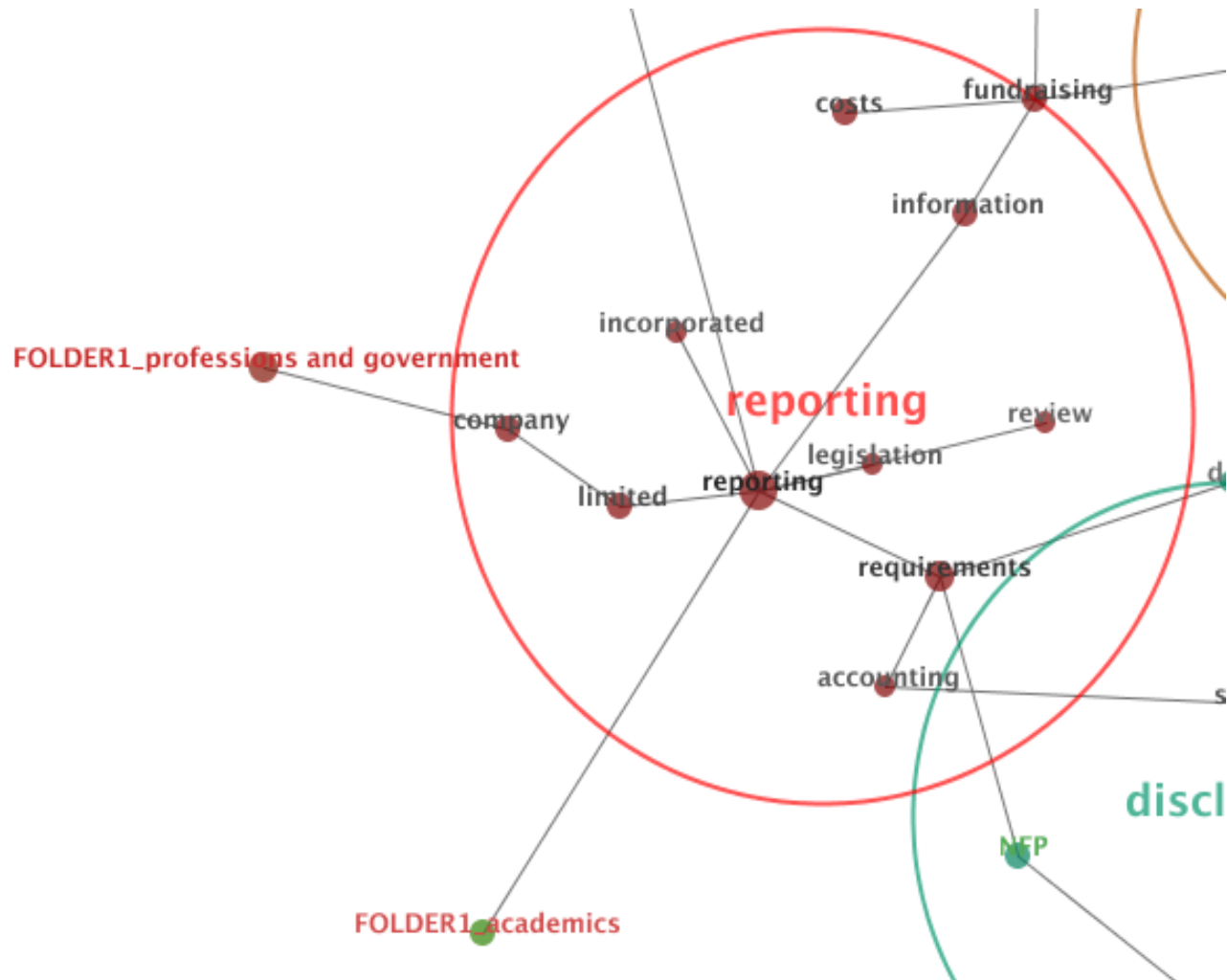
- Broad range of subject covered
- Two major themes emerge
- Shows focus of three groups



NFP's and individuals focused on the clients and communities section of Ebrahim's model



NGOs and reporting





NVivo cp. Leximancer

Leximancer and Nvivo

- Editing phase – remove, merge, insert concepts
- While Leximancer enables the linkage of key words and concepts it does not assist in teasing out the richness of the ideas.
- NVivo software used for qualitative coding, a ‘process by which segments of data are identified as relating to, or being an example of, a more general idea, instance, theme or category’ (Lewins & Silver, 2007, p. 81).
- The key concepts emerging from the high level Leximancer analysis may therefore be explored using NVivo.

Some comparisons

	Leximancer	NVivo
Fundamental approach	Psychology – Andrew Smith (U of Q)	Sociology – Lyn Richards (Latrobe)
Coding	Words – creating a thesaurus	Ideas – must be self-coded
Connection of ideas	Proximity	Context
Conclusions on research questions	Weight of observations	Connections The disconfirming case
Data	Anything but video	Anything.. (9.1 handles pdfs even better)
Nature of data	Large volume of reports on same subject	Diffuse but able to enlighten research question