**INTRODUCTION TO LEXIMANCER**

by Simon Moss

|  |
| --- |
| **Introduction** |

Suppose the university wants to invest in resources that could assist research candidates. To achieve this goal, one researcher might conduct a study to uncover the main research interests of research candidates at Charles Darwin University. To achieve this goal

* the researcher collects a set of texts—such as research proposals these research candidates have submitted
* the researcher subjects these texts to a software program called **Leximancer**
* the software program automatically uncovers a variety of themes—or sets of terms that often appear close to one another in the text
* one theme, for example, might revolve around communities—and include terms such as fisheries, languages, and sustainability
* a second theme might revolve around children—and include terms such as disease, development, education, and nutrition

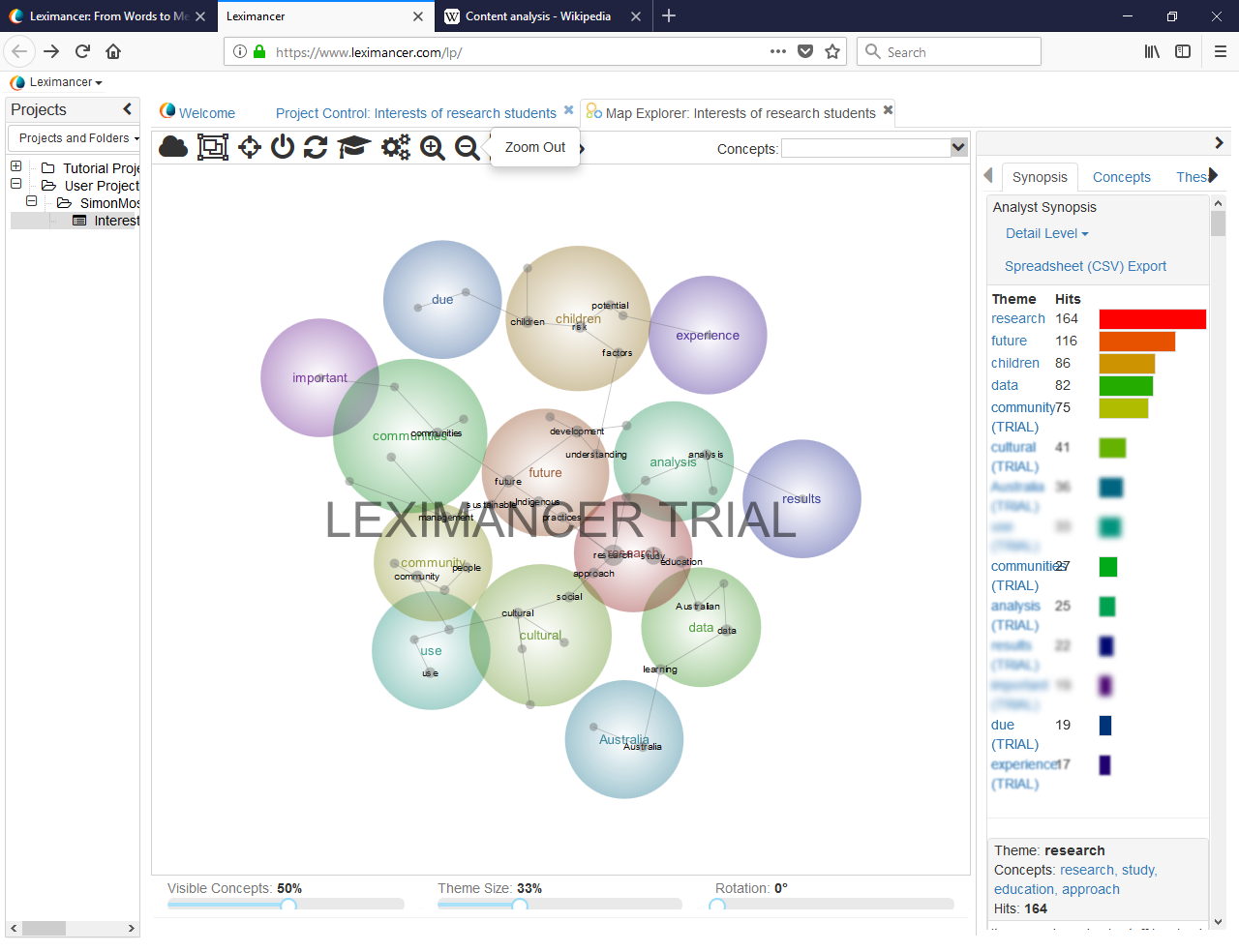
Consequently, the researcher might then utilize these insights to characterize the main resources the university needs to purchase. The university could first invest capital into research around the sustainability of communities. The university might then invest resources into research that explores childhood problems, and so forth.

**Purpose of Leximancer**

This example illustrates some of the features and benefits of Leximancer—a program that facilitates a methodology called content analysis. In essence, Leximancer can be used to uncover themes from texts, such as books, newspapers, emails, and other documents.

|  |
| --- |
| **Example** |

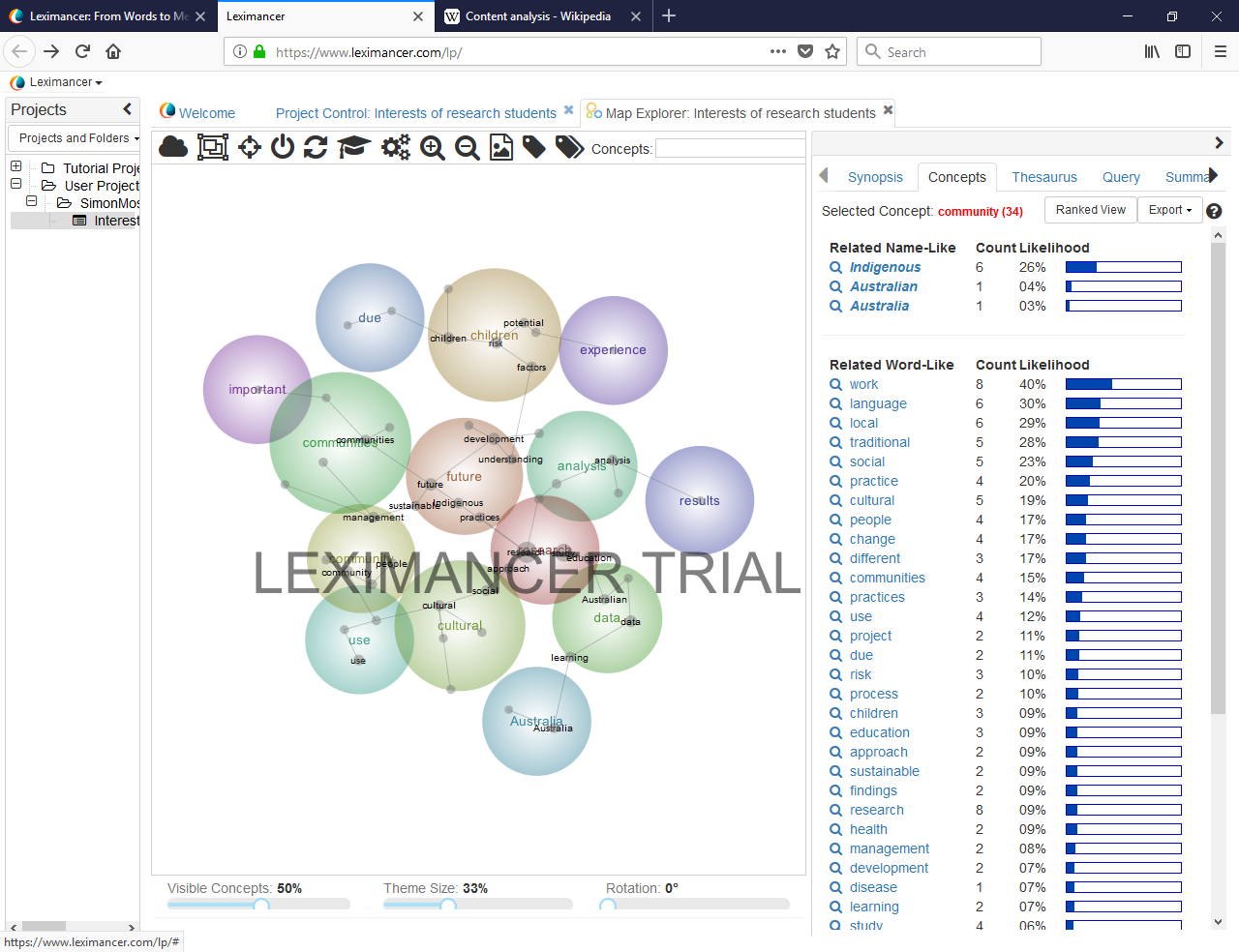
To introduce you to Leximancer, consider the following output. Note the word “LEXIMANCER TRIAL”, superimposed on the top, appears only because I have used a trial version of this software to generate the data.



The larger words, such as “important”, are names that Leximancer has assigned to these themes. The smaller black words are terms or concepts that correspond to these themes. For example, this display, called a concept map, shows that research candidates tend to be interested in research around

* communities
* the future
* children, and
* culture

Some of these themes, however, such as “important” or “results”, are not especially helpful to the researcher and, therefore, could be dismissed. The researcher could then clarify the terms and concepts that correspond to each theme, as shown below. This table indicates the term communities often appears close to Indigenous, work, language, local, tradition, and so forth. Accordingly, communities often refer to traditional, local Indigenous settings.



In essence, Leximancer assumes that terms that often appear close to each other in a text are typically related to each other in some sense. For example, the terms community, work, language, local, and tradition often appear in consecutive sentences. Consequently, these terms probably correspond to a shared notion or theme.

|  |
| --- |
| **Content analysis** |

To appreciate the benefits of Leximancer, you need to understand the methodology called content analysis. Actually, content analysis is not one methodology but a branch of methodologies. Despite this diversity, most variants of content analysis comprise several key features:

* Content analysis is primarily utilized to **examine texts** that comprise natural language, such as emails, advertisements, newspapers, and so forth
* The aim of content analysis is to unearth interesting **patterns** or regularities in communication
* To achieve this aim, these texts are assigned **codes** or labels
* This procedure is **systematic** in some sense

**Kinds of content analysis**

The kinds of content analyses vary considerably from one another. The following table distinguishes a few main kinds:

|  |  |
| --- | --- |
| Kind of content analysis | Example |
| **Conceptual analysis** is primarily undertaken to clarify the frequency or prevalence of particular themes or concepts. | * 20% of research candidates referred to communities * 30% of research candidates alluded to culture |
| **Relational analysis** explores the degree to which various concepts appear close to one another in the text | * Of the abstracts that revolved around communities, 25% also referred to culture |
| **Concept maps** depict the extent to which various themes or concepts are related to each other | * The display that Leximancer produces is an example of a concept map |

Leximancer generates concept maps automatically, except the researcher must obviously

* choose the research question
* decide on which texts to analyse, and
* interpret the meaning or implication of the results

**Content analysis versus thematic analysis**

A technique that resembles content analysis is thematic analysis. In both instances, researchers strive to identify themes from text. Indeed, some researchers utilize the terms content analysis and thematic analysis interchangeably. In practice, these approaches are similar. Yet, the following table highlights some historical differences between content analysis and thematic analysis (for a detailed discussion, see Vaismoradi et al., 2013).

|  |  |
| --- | --- |
| Content analysis | Thematic analysis |
| The researchers are particularly interested in a specific text, such as a historical encyclopedia. Their aim is to uncover the main themes in this text | The researchers are not especially interested in a specific text. The text is merely utilized to uncover interesting themes about a broader topic |
| The researcher tends to apply a systematic, methodical procedure to uncover the themes | The researcher, although systematic, might be more sensitive to the narrative, context, or setting when extracting themes |
| The researchers who conduct content analysis tend to be more interested in numbers—such as counting the occurrence of some theme | The researchers are not as interested in numbers |

|  |
| --- |
| **How to utilize Leximancer** |

**1 Access funding**

To access Leximancer, several options are available. In particular,

* you can check whether anyone in your research team can access this software
* you can request a free 7-day trial from <https://info.leximancer.com>. Choose “Products” and “Academic”. The 7-day trial includes most of the features, but obscures some of the output, so you cannot use the graphs or tables in publications
* you can purchase Leximancer yourself

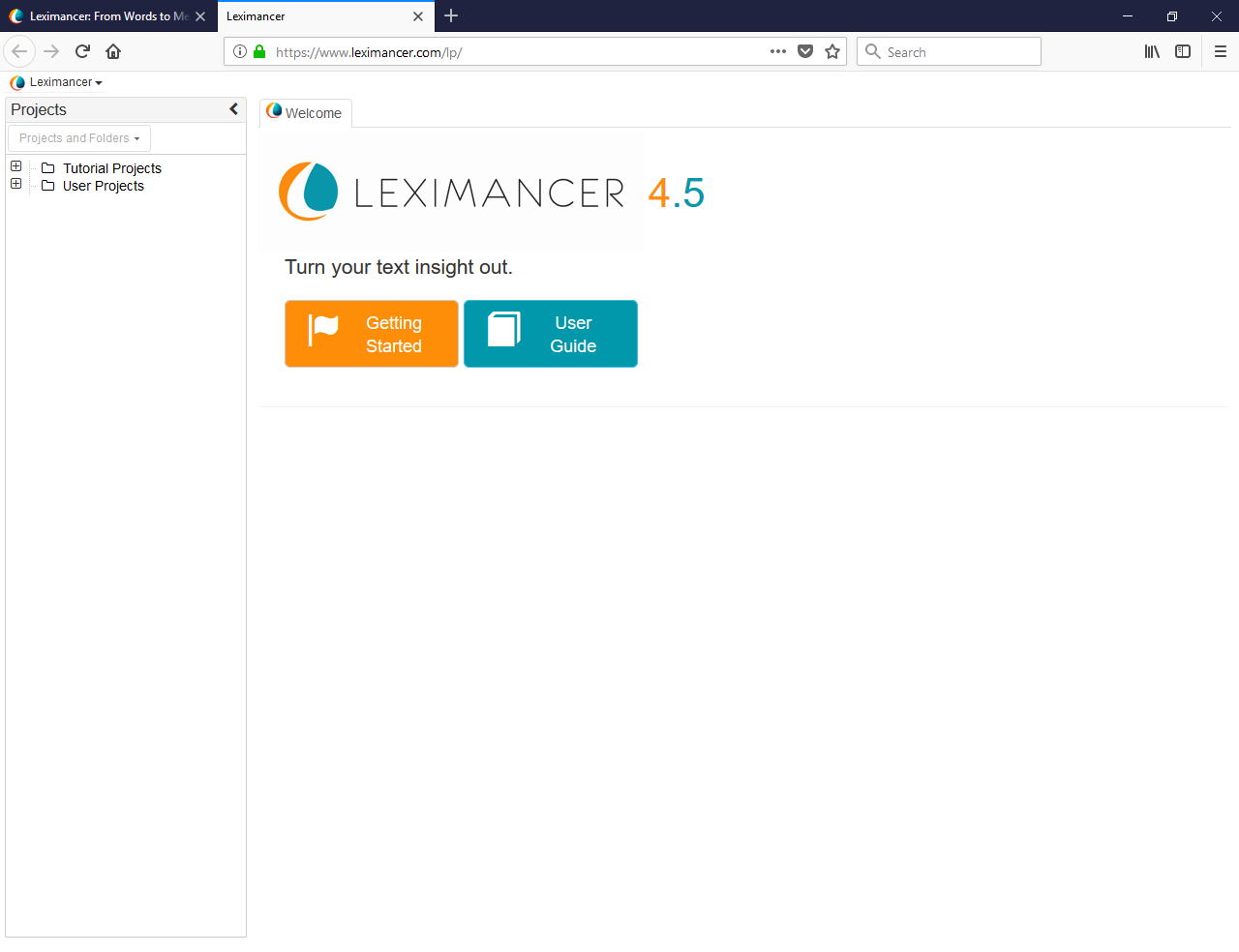
For academics and research candidates, Leximancer costs approximately $750 a year or $1500 for an ongoing licence plus GST. To access funding to pay these costs,

* contact the HDR administrator in your college or school
* indicate that you would like to purchase Leximancer from <https://info.leximancer.com>
* ask this person to organize the purchase for you—because the website requests a credit card number

**2 Open Leximancer**

Once you have requested a free 7-day trial or purchased the software, to activate Leximancer

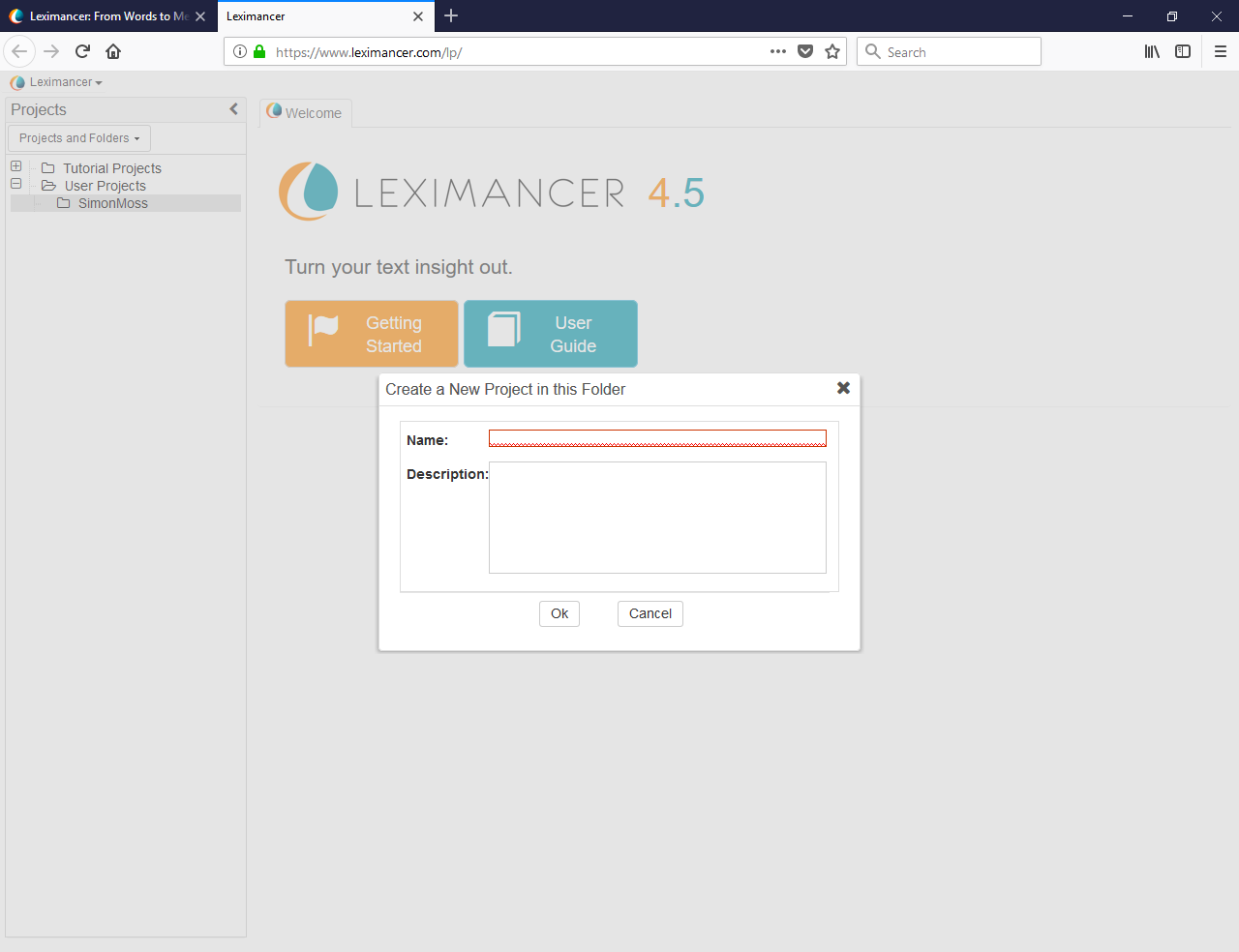
* visit <https://www.leximancer.com/lexiportal/>
* enter your username and password
* choose the button “Use portal” to generate the following screen



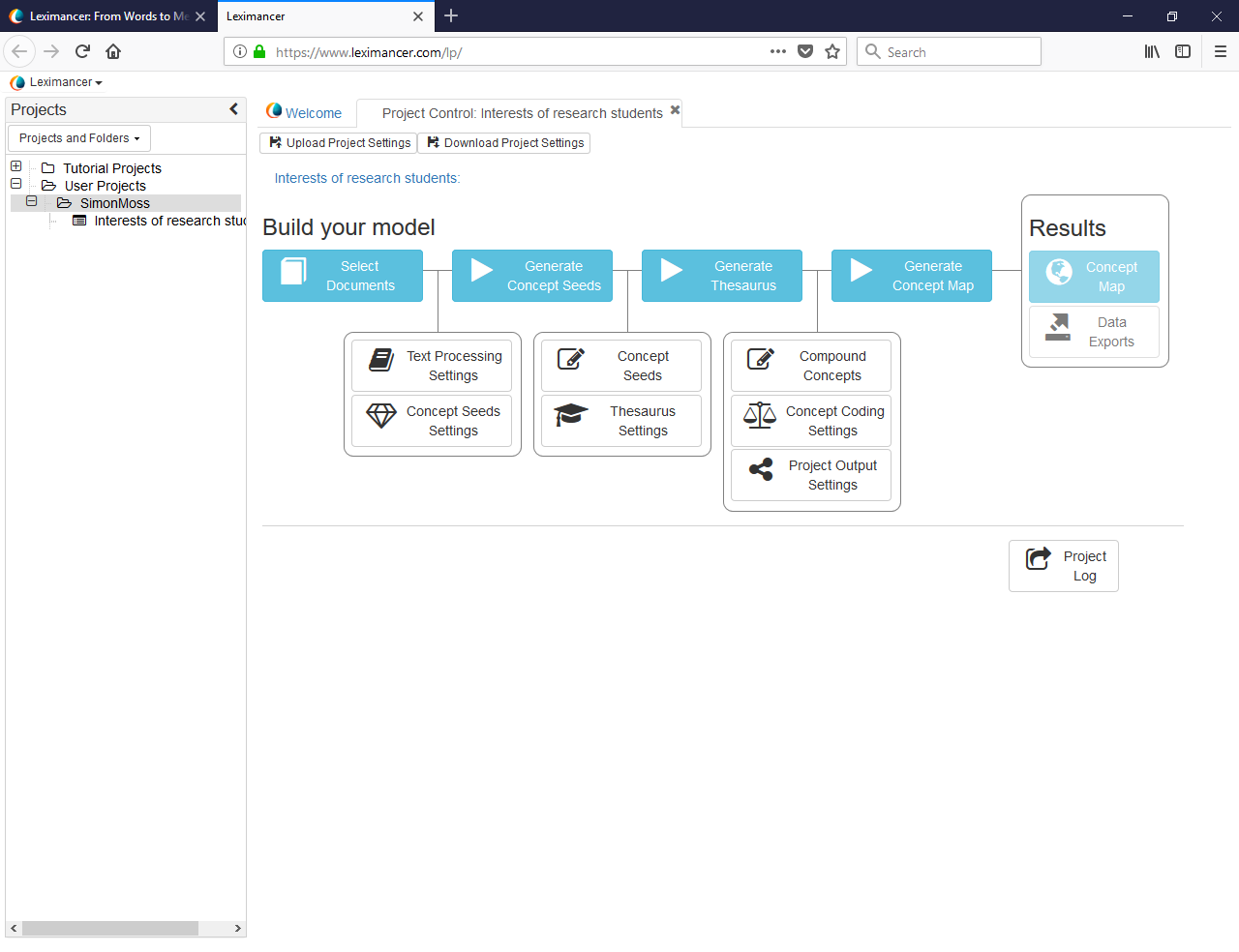
**3 Begin a project**

To begin a project

* click the + sign alongside “User Projects”. Your username will then appear under “User Projects”
* click your username to generate a set of options
* choose the option “create project” to generate the following screen



In the box called name, you can then assign a label to your project, such as “Interest of research students”. In the box called “Description”, you can record other details, such as the research questions you want to answer or the texts you want to analyse. After you press OK, the following box appears.

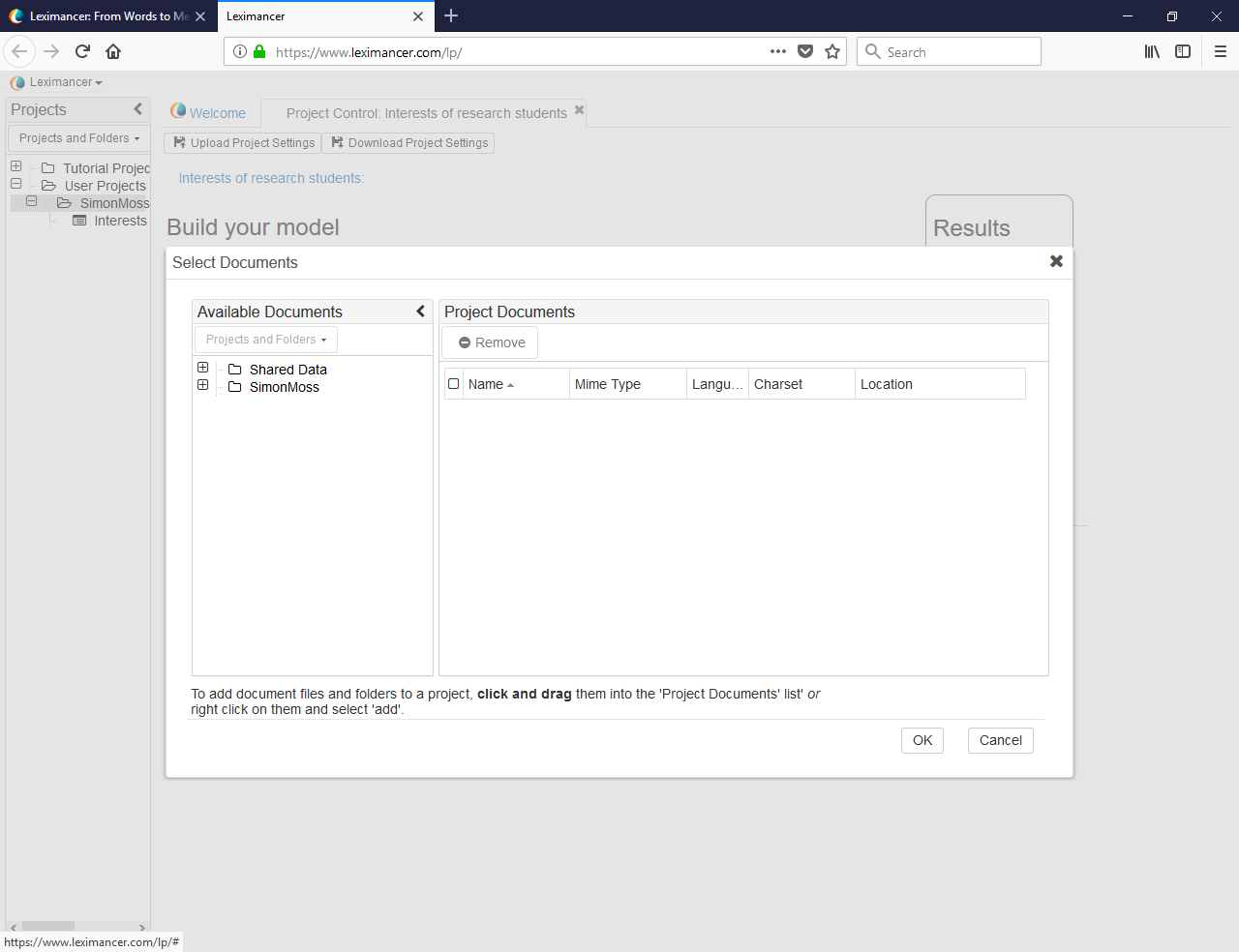


**4 Upload data**

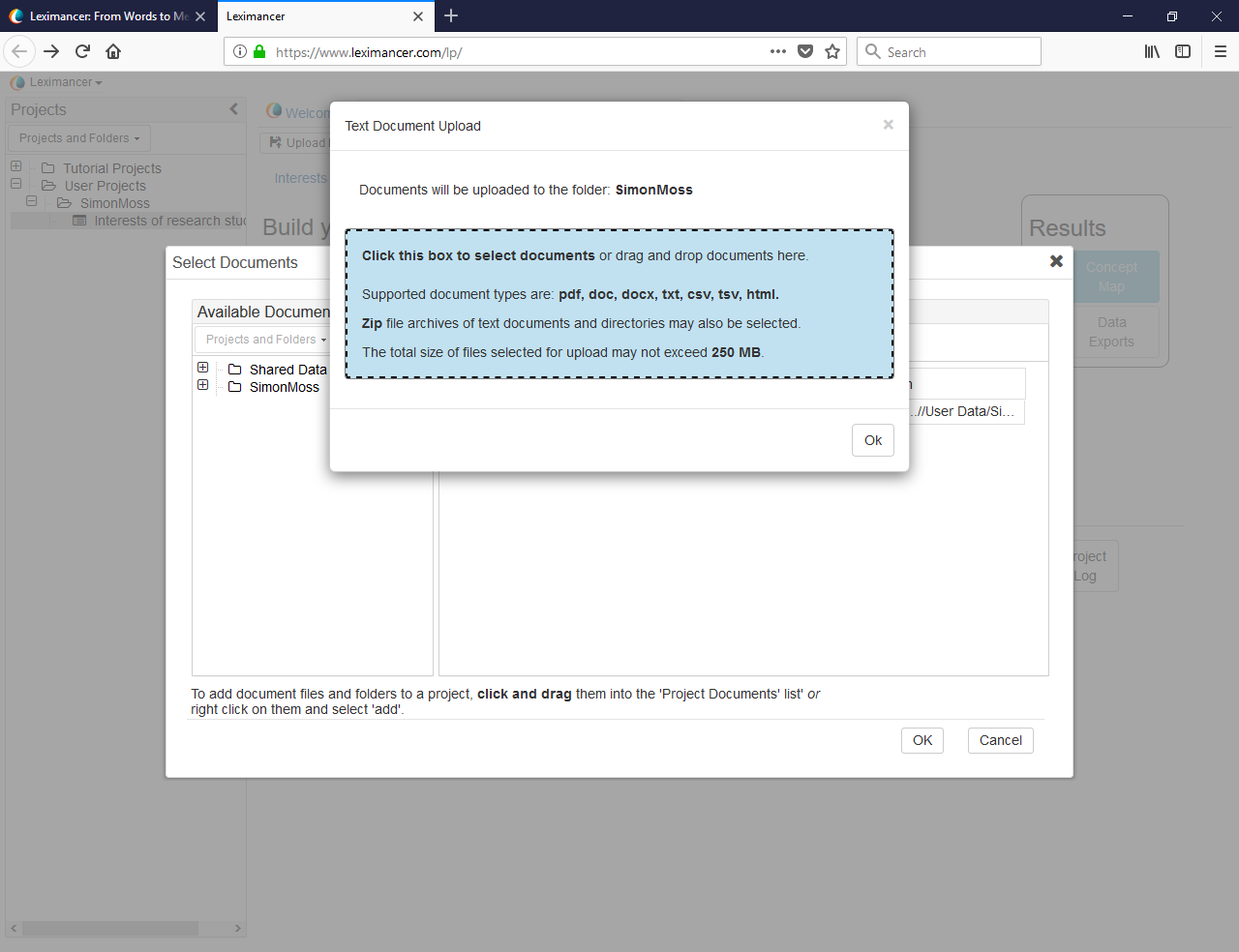
You then need to upload the data—the texts you want to examine. Leximancer can analyse a range of formats, including pdf, csv, Microsoft Word files, and Microsoft Excel files. However, whenever possible, convert these files to plain text files. That is, in Microsoft Word, Microsoft Excel, and many other programs, you can typically save these files as .txt files or similar. The reason is that

* text files are devoid of other features
* in contrast, other formats, such as pdf, might include distracting features, such as page numbers, headers, footers, or other information. Leximancer may incorrectly assume these page numbers, headers, or footers are part of the text.

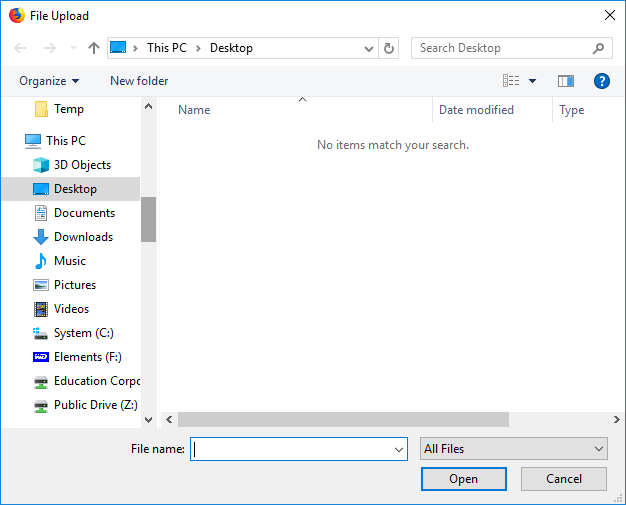
To upload these text files or other files, under “Build your model”, click “Select Documents” to generate the following screen



To upload the data, right click your username—such as “Simon Moss” in this example. A series of options will then appear; choose “Upload documents” to generate the following box



When you click this box, the files on your computer will appear, as illustrated in the following figure. You can then locate and choose the text files you want to analyse.



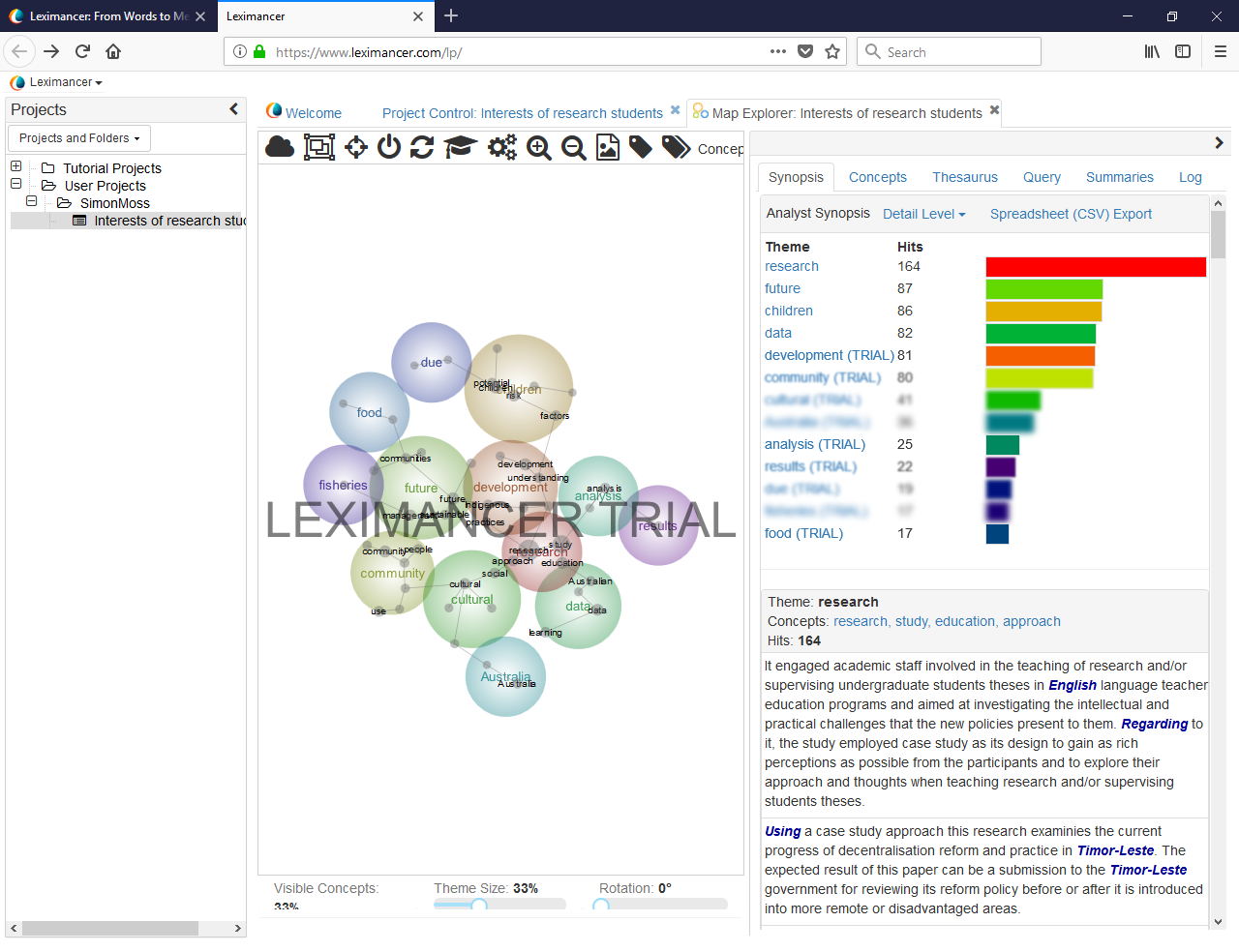
**Execute the analysis**

To analyse these data

* you might need to press OK to return to the screen titled “Build your model”
* if permitted, you can click “Concept map”
* or, if this option is not accessible, click “Generate concept seeds”, “Generate thesaurus”, “Generate concept map”, and then “Concept map”
* Alternatively, you can adjust some of the defaults before you press these buttons

**Examine the output: synopsis**

This procedure will generate output that resembles the following screen. Note, some of the information in the right panel is obscured, because this output was generated by a trial version.



To help you interpret this output

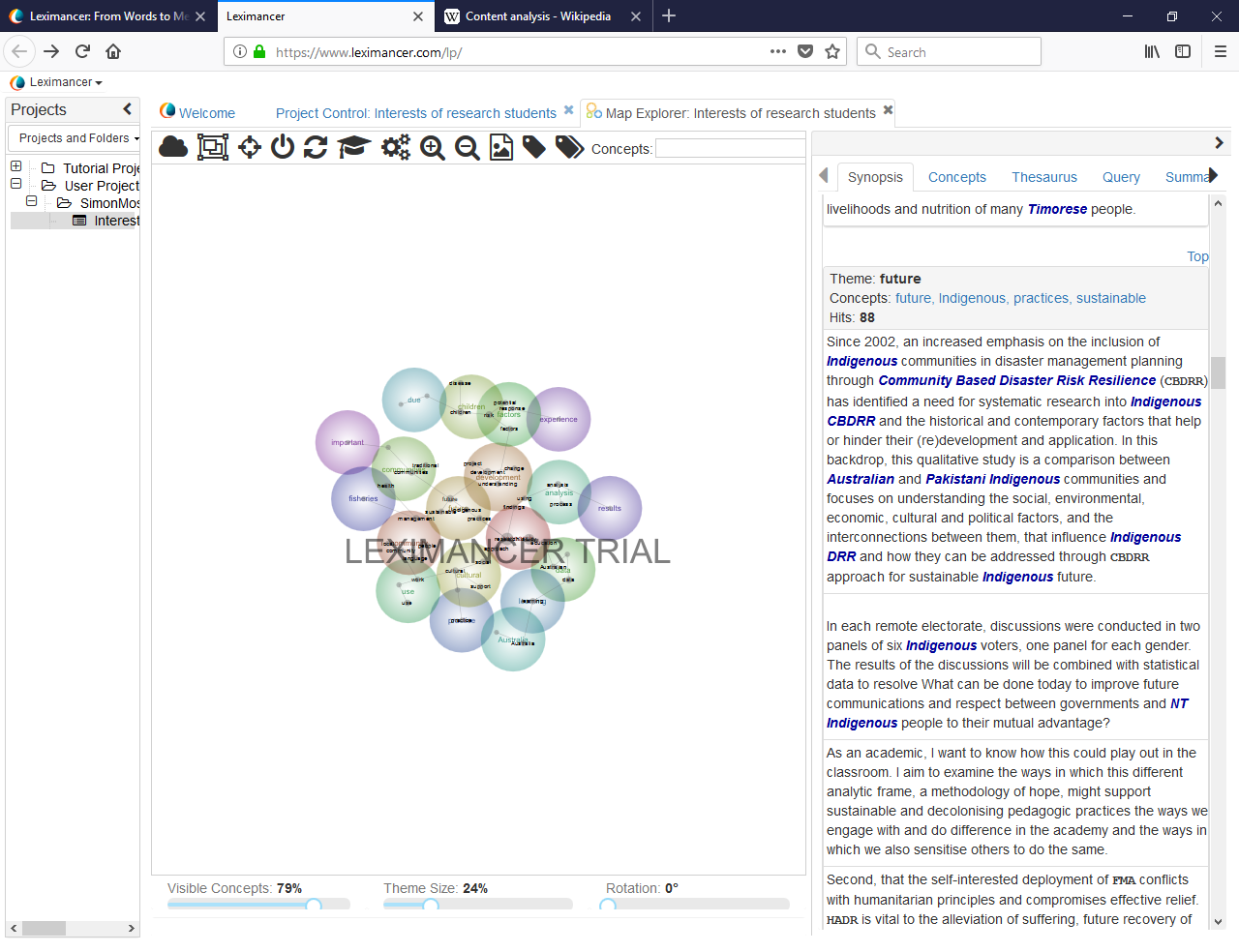
* the bubbles in the central panel represent the themes, such as fisheries and food.
* bubbles that are closer together represent themes that tend to appear close to one another and thus might be related to each other.
* to illustrate, fisheries and food might be associated with each other
* below this graph are options that can be adjusted; for example, you can increase the number of small words or concepts to clarify the contents of each theme

On the right side, you can access more specific information about each theme and concept. You can choose between several tabs on the top, such as Synopsis, Concepts, Thesaurus, Query, and Summaries. To illustrate

* **Synopsis** specifies how often the main themes, such as future or community, were mentioned.
* For example, the terms or concepts that belonged to the theme called research were used 164 times.
* Likewise, the terms or concepts that belonged to the theme called future were used 87 times.

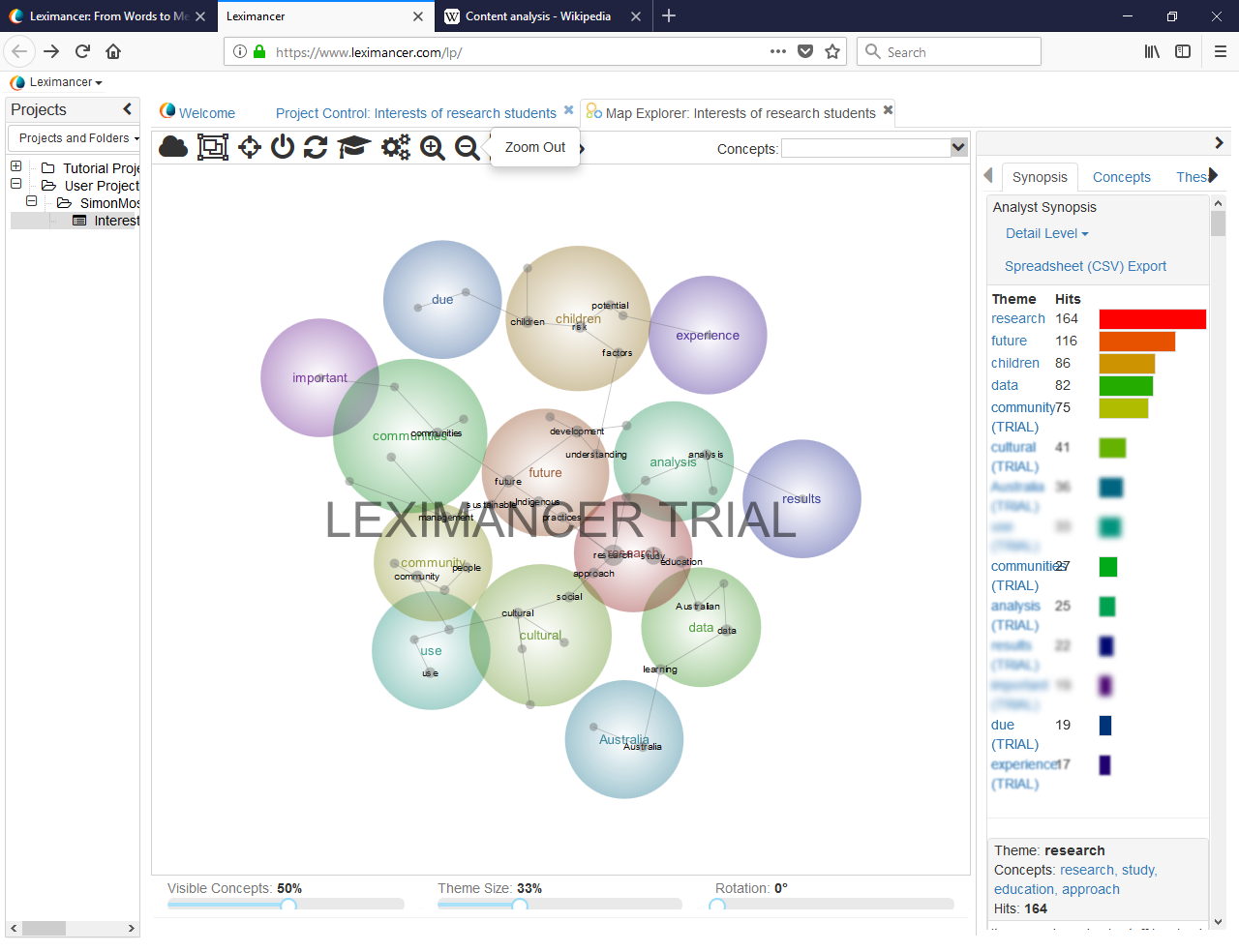
But, which terms or concepts correspond to these themes? To answer this question

* simply click a specific theme in the list, such as “future”
* the computer will then scroll down to generate the following output on the right side
* this information indicates that future corresponds to the terms future, Indigenous, practices, and sustainable. This theme may thus revolve around the future sustainability of Indigenous practices.
* the following paragraphs include the texts that referred to this theme



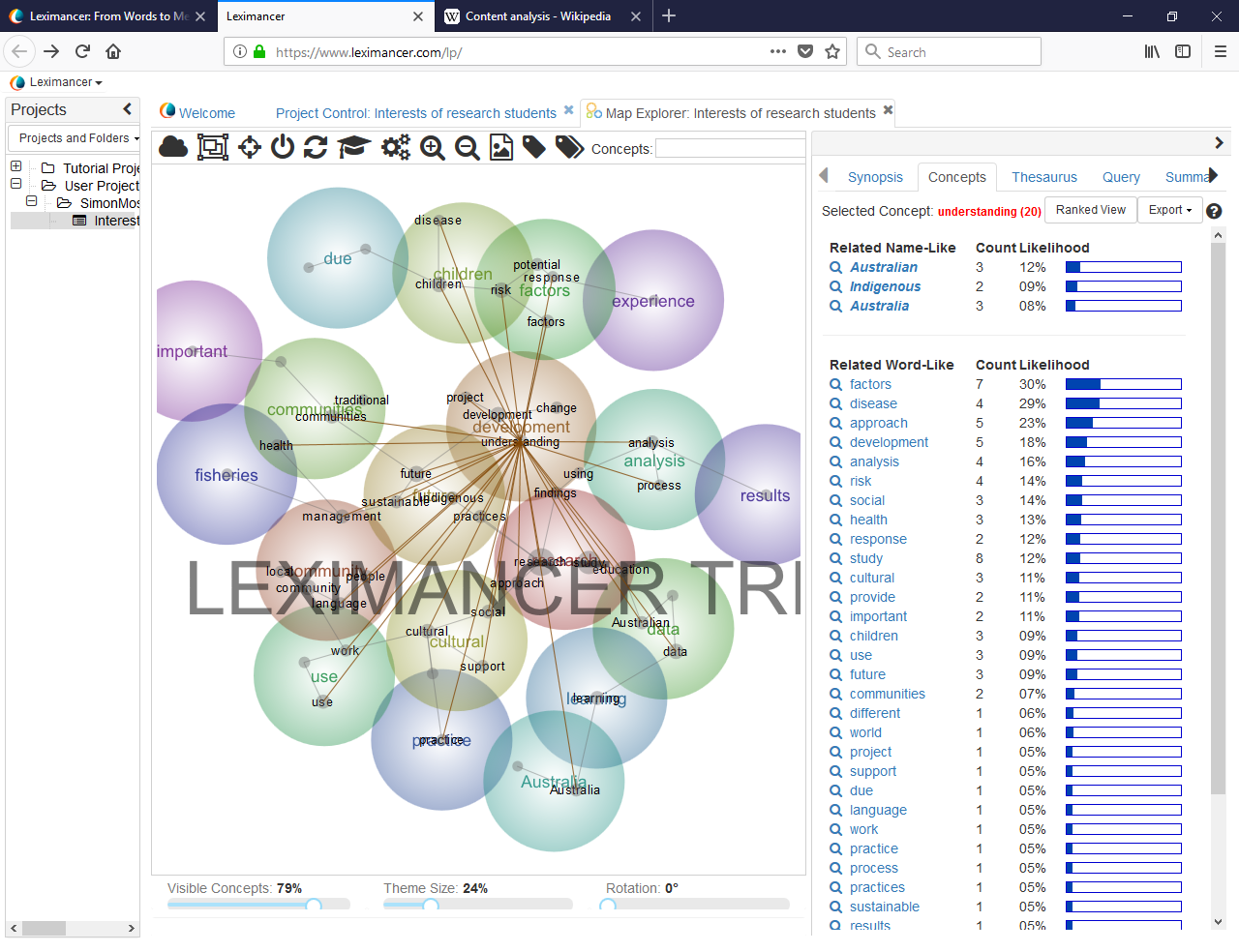
**Examine the output: concepts**

The synopsis tab presents information about each theme—sets of terms or concepts that tend to appear close to each other. In contrast, the **concepts** tab presents information about each term or concept. This tab helps researchers define these concepts. To use this tab, click one of the smaller black words—the concepts—on your concept map, such as “understanding”



You will then receive the following information on the right panel. This information indicates that

* “understanding” appeared 20 times
* the word tended to appear close to “factors”, “disease”, and “approach”—and thus understanding, in this context, may primarily involve ways to clarify the factors that promote disease.



|  |
| --- |
| **Other considerations** |

This document has merely introduced the fundamentals of Leximancer. To learn about this application more comprehensively, you could

* experiment yourself with this tool
* read the manual

Nevertheless, this section offers some other helpful details.

**Underlying rationale**

So, what is the rationale or logic that underpins Leximancer? In essence, to analyse the data

* by default, Leximancer examines pairs of consecutive sentences
* Leximancer then counts the number of times particular words, such as community and fisheries, appear in the same pair of sentences.
* words that often appear in the same pair of sentences are assumed to be related
* actually, the algorithm that Leximancer utilizes is not publicly available but utilizes Bayesian statistics
* to facilitate analysis, Leximancer will learn to discard words that appear evenly across the text and, therefore, are probably functional rather than informative words—such as “the”, “to”, “and”, “her”, and so forth. The exclusion of these words is called stop-word filtering

Thus, the results are dependent purely on the relative location of words. Leximancer does not utilize any pre-existing knowledge about the words. Indeed, Leximancer can be applied to any human language.

**Names of themes**

In the previous examples, Leximancer uncovered several themes—sets of words that appear close to another—such as community, culture, and research. Leximancer merely utilizes one of the words in this set to name this theme. Consequently, these names might not be suitable. To illustrate

* one theme might comprise the words Asia, Europe, Africa, and Antarctica
* Leximancer might name this theme Asia
* researchers, therefore, will typically change the name of this theme to the notion that all these words share, such as Continents

**Options**

Many settings can be adjusted. For example

* if the text comprises many very small sentences, rather than examine each consecutive pair of sentences at a time, Leximancer might examine three, four, or more consecutive sentences at a time
* you can add words to the stop-list—the words that Leximancer will disregard
* if you use a csv file, you can add tags or codes to segments of data. For example, you might add codes to all text that revolves around complaints. Leximancer will then determine which terms or concerns are most related to these segments of data
* you can increase the number of themes and concepts that Leximancer will extract
* you can merge similar terms or concepts, such as Australia and Australian

|  |
| --- |
| **Previous illustrations** |

To help you decide whether Leximancer might be relevant to your research, consider the following table. This table summarizes previous studies that have utilized Leximancer—and thus demonstrates the versatility of this application.

|  |
| --- |
| Study |
| Leximancer uncovered the range of errors that mariners commit on a ship, derived from safety incident reports (e.g., Grech et al., 2002) |
| Leximancer uncovered the main themes that were discussed in the Journal of Cross-Cultural Psychology, such as values and acculturation (Cretchley et al., 2010) |

|  |
| --- |
| **References** |

Cretchley, J., Rooney, D., & Gallois, C. (2010). Mapping a 40-year history with Leximancer: themes and concepts in the journal of cross-cultural psychology. Journal of Cross-Cultural Psychology, 41(3), 318-328.

Grech, M. R., Horberry, T., & Smith, A. (2002). Human error in maritime operations: Analyses of accident reports using the Leximancer tool. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46(19), 1718-1721.

Vaismoradi et al. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nursing & Health Sciences, 15(3), 398-405.