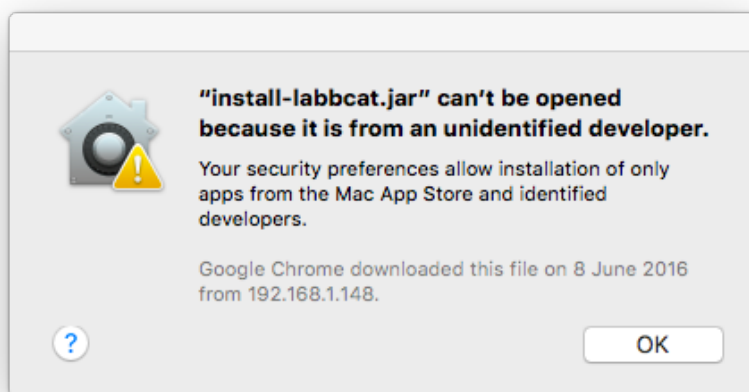## Exercise 1   Setting Up

In this exercise you will:

- Install the LaBB-CAT software

- Define corpora

- Define transcript types

- Define speaker meta data

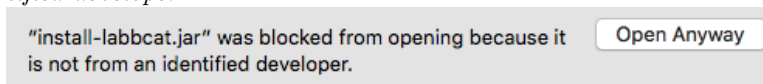After this you will have an empty LaBB-CAT database set up ready to upload transcripts into.

_____

1. You have a file called *install-labbcat.jar* – double click this file to start the installer.
   If you are using OS X, you may see a message that the file can't be opened:
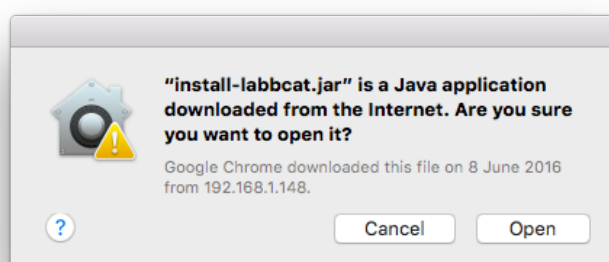

   If this happens:

   (a) Click the Apple icon in the top left corner of the screen.

   (b) Select *System Preferences*

   (c) Click *Security & Privacy*
       Near the bottom it says *"install-labbcat.jar" was blocked from opening because it is not from an identified developer.*
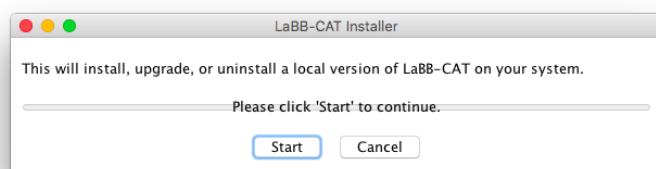

   (d) Click *Open Anyway*
       You may see another warning about the program being downloaded from the internet

   (e) Click *Open*

2. Click *Start* You will see the progress bar move as files are installed. Once this is finished, you'll see a message saying "Installation complete."

3. Click *Finished* to close the installer

The software is now installed. LaBB-CAT is a browser-based system, which means that it works as a mini web server on your computer, and you need to access it using your web browser.

Each time you want to use LaBB-CAT, you must start it up, and which you've finished, you close it down again.

To start LaBB-CAT, click the LaBB-CAT icon in your applications area.

- On Windows, open the *Start* menu and type LaBB-CAT.

- On OS X you will find LaBB-CAT in your *Applications* folder.

A window called "LaBB-CAT Server" will open, and after a short delay, your default web browser will open on a page called "LaBB-CAT" (The first time only, this page will initially display the LaBB-CAT licence).

Now that the software is installed, we will set up a basic structure for receiving data:

4. Start LaBB-CAT as described above.
This will open your web browser on LaBB-CAT's start page.

5. The start page has a link on it called "Where do I start?" – you may like to click on this link and read the first section, which explains a little about how to navigate around LaBB-CAT and where to find online help and hints.

6. Click back on the start page of LaBB-CAT (the page with the "Where do I start?" link).

Now we will set up some corpus names. . .

7. On the menu at the top, click the *corpora* link.
This page shows a list of current corpora, which only contains one corpus, called 'corpus'.

8. Underneath the 'corpus' corpus, there's a form that you can fill in to add a new corpus. Fill in the following information:

   (a) *name:* QB

   (b) *language:* (leave this as 'English')

   (c) *description:* Quakebox recordings

9. Click the *New* button to add the 'QB' corpus.
You should see a message at the top of the page saying "Record created" and now the 'QB' corpus is in the list, under the 'corpus' corpus.

10. Add another corpus called UC with the description Campus recordings.

11. We won't actually be using the corpus called 'corpus', so we want to delete it. To do this, click the *Delete* button below and to the right of the 'corpus' corpus in the list.

12. You will be asked "Are you sure you want to delete this record?" You are sure, so click *OK*.
This will 'cross out' the corpus in the list.

13. Now click the *Save Changes* button that has appeared in the top right corner of the page.
The row will be deleted from the list.

Now you have some corpora set up with the names you've provided.

The data we are using is a collection of stories about peoples' experiences during the devastating earthquakes that hit the Canterbury region of New Zealand in 2010 and 2011. Some recordings are interviews, where an interviewer asks the participant questions, and others are monologues. Now we're going to set up these two transcript types . . .

14. Click on the *transcript types* menu option.
    You will see a list of transcript types, although there's currently only one type in the list, called 'interview'.

15. Underneath this, fill in the empty box with the word: `monologue`

16. Click the *New* button.
    You will see a message at the top saying "Record created", and you will notice that now the list has two transcript types, 'interview' and 'monologue'.

Now that we have both corpora and transcript types, we're going to set up meta-data options for the participants in our corpus . . .

17. Click on the *participant attributes* menu option.
    You will see a list of fields or attributes that participants (or speakers) can have. There are currently only three attributes:

    - 'Gender' – i.e. whether the participant is male or female
    - 'Birth Year' – i.e. the year the participant was born
    - 'Notes' – general arbitrary notes

    For our corpus data, we don't have the exact year of birth. Instead we have the age of the participant, defined by various age group categories.

18. Each of the three attributes already in the list has a form specifying its definition, and below all three, there's an empty form, which we will now fill in with the following information:

    - *Category:* 'General'
    - *Attribute:* `age_category`
    - *Label:* `Age`
    - *Description:* `Age Category`
    - *Type:* 'select' (this is because we want to be able to select from a list of possible values)
    - *Style:* leave all the three options un-ticked
    - *Order:* `4` (this will position the 'Age' attribute between the 'Birth Year' attribute, whose order=3 and the 'Notes' attribute, whose order=100)
    - *Searchable:* 'Searchable'(so that it appears on the search page)
    - *Access:* 'Public'

    If you want more information about what each of these are for, check the online help for this page

19. Click the *New* button.
    You will now see the new 'Age' attribute below the 'Birth Year' field. Now we need to define the options for it. . .

20. Underneath the 'Age' attribute details, there's a button called *Options*; press it.
    This shows a (currently empty) list of options for the 'age_category' attribute.

21. In the blank row, fill in the following details:

    - *Value:* `18-25`
    - *Description:* `18-25 years`

22. Click the *New* button to add the option.

23. In the blank row at the bottom, fill in the following details:

    - *Value:* `26-35`
    - *Description:* `26-35 years`

24. Click the *New* button to add the option.

25. Similarly, add the following age categories:

    - `36-45`
    - `46-55`
    - `56-65`
    - `66-75`
    - `76-85`

26. Add a final option:

    - *Value:* `85+`
    - *Description:* `85 years or more`

27. Lastly, add a 'default' option for participants with missing data:

    - *Value:* (Leave the box blank)
    - *Description:* `[not specified]`

28. Now click the *participant attributes* option on the menu to return the list of all attributes.
    We are going to add a few more attributes, but they will be 'free text' fields without predefined options.

29. Add another attribute, called `ethnicity`. For 'Type' select 'string', and make it 'Searchable'.

30. Similarly, add the following 'Searchable' 'string' attribute:

    - `languages_spoken` – a list of languages they speak

    . . . and the following 'Not Searchable' attributes:

    - `grew_up` – what country they grew up in
    - `grew_up_region` – what region of New Zealand they grew up in
    - `grew_up_town` – what town or city they grew up in

31. Lastly, as we will not be using it, delete the 'Birth Year' attribute.

Now you have an empty database for which you've:

- created two corpora, QB and UC,

- created a new transcript type, so that we can have monologues as well as interviews, and

- created some new attributes for participants, so we can record the ages of our speakers, their place of origin, and the languages they speak, in addition to their genders.