LING 516: Week 13, cl. 2

author: Sowmya Vajjala date: 5th April 2018 autosize: false

Today's Plan

I will describe an idea (involving databases, AND web-interface). We will walk through its implementation step-by-step. - This web application shows some information about a given file to the user, if that file is in the database - Example information: file content, some descriptive statistics about the file - If the file name is not in the database, the web-interface just shows a message saying that file does not exist in the database.

What I will give you:

a corpus of text files (in a folder)

What all do you need to do to make that idea work?

Part 1: Database with file content/statistics stored

What do we need?: A database with
–a) a single table called **filecontent** with columns: filename, content.
–b) a second table called **filedetails** that has columns: filename, num. chars, num. words, num lines, average word length

(create using db browser - you can use python if you want)

[Note: The filenames should match in both tables. You can have a single table with all these, but I **want** it this way.]

Part 1: Once we create empty tables, what next?

We need a program that populates these empty tables with the file content and other statistics.

- Today's task: a program, that populates Table filecontent in the database, using the files in the folder.
- additional exercise for evening attendees: a program that populates the second table in the database.

[You can ofcourse do the whole thing in one program! I am just splitting one task into two smaller tasks.]

Part 1: How should this work?

- for each file in that corpus folder:

- read the file content (How? How do you handle file path?)

- insert filename (without .txt, and without full path?), file content into the table filecontent. [Later: Calculate descriptive stats, insert those too!]

Part 2: UI - how should it look like?

- page1: ask user for filename.

- if file name exists in the tables:
- page2: show file content, details about the file etc

- else:
- page 3: show a message saying file does not exist in the db.
- For Today, let us focus only on showing filecontent table's output
- additional exercise: additional exercise: extend this to include information from Table 2 as well.

Other things you can explore:

how to show all files with more than 100 words etc.