Definitions and files

**Book, secondary literature, research literature** – modern research books on ancient topics, from which I take the indices. Usually under copyright so we can only use a small part of the book (i.e., the index) for scholarly purposes.

**Source index, ancient text index, index locorum** – the part at the end of the book which keys ancient text references to a specific page in the book. Can be written in many different formats.

**Subject index, topic index** - the part at the end of the book which keys subjects to a specific page in the book. Usually includes also sub-subjects and sometimes sub-sub-subjects.

**Ancient texts, sources, works** – works written before 800 CE in the Mediterranean area, with authors known or unknown.

**Authors and titles of ancient texts** – this a complicated issue. Both authors and titles can be referred to in many various ways,

**Ref, reference** – a number (or roman numeral) or series of numbers pointing to a specific part of the ancient text. For example, Genesis 4.7 / IV.7 / IV:7/ 4,7 points to the book of genesis, chapter 4, verse 7. Different texts have different types of reference schemes.

**Hyphenation** – the common practice of shortening refs (and page numbers) with hyphens, e.g., 4.5-10 means chapter 4, verses 5 to 10. Sometimes this can be ambiguous if lacking data, for example 3.10-4.2 points to an unclear number of verses if we don’t know how many verses are in chapter 3.

**Fulltext** – the text itself of the ancient text, to which a reference points. Can refer to the text in the original language (Hebrew, Latin, Greek, Arabic) or to a translation to a modern language. Providing the users with fulltext in both modes and not only references will greatly enhance the value of the database for researchers. Many texts in the original language are available online through dedicated sites, some with with APIs. Translations are also available online for many texts, but only a small minority are with APIs.

**i.d.** – number used on our system to identify a specific ancient work.

**Files and database tables**

*Textsa* – a mysql table. Includes the following columns: ref,page,book bibliographic info,number,C. ~4,200,000 rows

ref – a reference in ancient text

page – a page in a modern book

book bibliographic info – the number of the modern book in the system

number – the i.d. of the ancient text in the system.

C – a record number between 1 and currently 4 million

For example a row with the following: 1.24,115,70,6284,3012323 means:

Ancient work with i.d. 6284 (Genesis) chapter 1 verse 24, is referenced in modern book number 70 on page 115, and this is record number 3012323.

The textsa database was built gradually from currently 15 textsa csv files, from *textsa1.csv* to *textsa15.csv* (so it can be rebuilt if needed).

*subjects* (currently texts\_subjects2.csv): a file of all subjects and the row numbers (C) to which they point in *textsa*. To take our example, we can know that Genesis 1.24 is related to the subject listed in the subjects file. ~45000 rows.

*abbreviations* (currently abbreviations4.csv): a file with information about ancient texts. Includes columns for variations on author name, title names, work codes/numbers on various websites when available, dates, original language, and an i.d. on our system. ~12,000 rows.

This is the (relatively simple) row for genesis:

6284,Hebrew Bible,,old testament,Bible,,,,,,Genesis,,,,,,,Gen.,,Gen,,Ge.,gen,,,,,,,,,,,,,,,,,Hebrew,-12,-3,,,,,,,

*titles* (currently titlesa.csv): a file mostly derived from *abbreviations*, includes the specific author and title used on the website for each work, as well as **i.d.** and date and language information. Also includes ~5000 authors and titles for additional works not appearing in *abbreviations*. These are works which were not identified so we have no further information on them. some of them may be mistakes, most other overlap with the identified works but to verify which someone has to go over them and see. The non-identified works received an i.d. of 50000 and above. The “number” column in *textsa* points to the number i.d. in *titles*.

This is the row for genesis:

6309.0,Hebrew Bible,-3,-12,,Hebrew,6284,Genesis

*bookreferences* (currently *bookreferences2.csv*): keys “book bibliographic info” numbers to the relevant book titles and also gives the google book code of each one. ~120 rows

Crn2.csv and crneeng.csv – files with links to fulltext. Will not explain because will probably be built from scratch.

emails.csv – a file which records what people enter in the search form. sometimes useful.

Flaskapp.py – the code for the app sits here. it uses templates/index.html which is the html for the main page.

db\_migration.py – loads information from the csv files to mysql database. This is in the home/moblid directory.