Requirements
Last modified: October 31, 2016

Preamble

In the following document, the SI standards will be used for measurements:

- *1 second* will be abbreviated as *sec* or *s*;

- *1 byte* (abbreviated *B*) refers to *one octet*, or *eight bits*;

- *1 megabit per second* (abbreviated *1 Mbps*) refers to 10^6 *bits/sec*;

- *1 kilobit per second* (abbreviated *1 Kbps*) refers to 10^3 *bits/sec*;

- *1 Terabyte* (abbreviated *1 TB*) refers to 10^{12} *bytes*

The following definitions will also be used:

- *data decay* refers to the mutation of some stored data when it is not intended by the user.

Functional requirements (most important first)

1. The application should parse all XML tags required by the TEI standards, as well as those required by the **[x]**, **[y]** and **[z]** namespaces [1]

2. User should be able to search through a transcribed document [2]

    (a) User should be able to search for plain-text strings [3]

    (b) User should be able to search through documents by ID (volume, date, page) [4]

    (c) User should be able to search through documents by proper names [5]

    (d) User should be able to search through documents by annotations

3. User should see the search results and the ten words of context surrounding the result [6]

4. User should see the document ID in the search results

5. User should see the date of the document in the search results

6. User should be able to view a list of documents in chronological order [7]

7. The application should store the documents with a SHA-256 checksum

---

[1] **Rationale**: Nothing can happen without this

[2] **Rationale**: this is the main purpose of the software, and it would be basically useless without it. The sub-requirements are requirements as a part of this

[3] **Rationale**: Most people won't know the ID of what they want, or a specific name

[4] **Rationale**: This was specifically stated to be an important requirement

[5] **Rationale**: Not specifically stated by the LACR team

[6] **Rationale**: It's obviously as important to be able to see the search results as to be able to make a query, unless you're using your server as a space-heater

[7] **Rationale**: After searching, this is the other main point of our package

8. The application should periodically check the stored documents against their checksum so that data can be protected against *data decay*. [8]

9. User should be able to download the XML file of desired document [9]

10. User should be able to download the original scan of desired document [10]

11. The application should be able to interface with the CLARIN project

12. User should be able to turn *all* annotations "off" and "on" [11]

13. User will be provided with **[DOCUMENTATION]** [12]

Non-functional requirements (most important first)

1. The web-application will be licensed under **[LICENSE]** [13]

2. The web-application must be suitable for searching queries, browsing the XML documents, translations, interpretation and scanned pages for these user groups: [14]

   - secondary school children aged 11 years old or older;
   - University and college students;
   - researchers;
   - The wider public with average computer skills, and who are aged 11 years or older

3. The web-application must be accessible from Chrome version 50.+, Firefox version 45.+, Safari version 9.+ , Edge version 38.14393 and higher. [15]

4. It will be possible to extend the application such that translation and interpretation views can be added

5. The web application will be optimized for these screen resolutions: [16]

   - Extra small devices, Phones: less than 768px
   - Small devices Tablet: min-width:768px
   - Medium devices Desktops: min-width:992px
   - Large devices Desktops: min-width:1200px

---

[8] **Rationale**: Because these are public records, it is important they are always accessible and correct

[9] **Rationale**: Although considered important by the LACR team, it only makes sense when there is some way to search for a document

[10] **Rationale**: See above

[11] **Rationale**: This was considered a nice extra to have by the LACR team AFAIK

[12] **Rationale**: While we should be documenting things as we go, and it is important, there's no point having documentation for vapourware

[13] **Rationale**: If they don't agree with the license, then there's no sale

[14] **Rationale**: Being accessible to everyone was considered very important by the LACR team, and thus the usability requirements are the most important non-functional requirements

[15] **Rationale**: This is also important to accessiblity

[16] **Rationale**: Whilst it is important to accessability, it will mainly just impose annoyances on the user as opposed to making it unavailable to them

6. User should receive search results within 5 sec. if their device's connection to the Internet has a download speed of 1 Mbps, and their upload speed is 256 Kbps [17]

7. New data (XML files, translations, interpretations and scanned pages) should be imported via SFTP to dedicated folders for them. (via web interface, drag and drop function, filter out directories, only files which are supported) (functional requirement?) [18]

8. The system should be reliable and available at least 363 days in the year. [19]

9. It is acceptable to restart the system in the event of failure at most 1 time a week. [20]

10. The web-application should support up to 1000 concurrent users. This requirement is bound with minimal hardware requirement specification.

11. The system can be restarted at most once a week for maintenance purposes and applying software updates.

12. The system should be able to store single file with size up to 1 TB.

13. The system should be running on Ubuntu 16.04 LTS

14. The system should be installed and tested within one week by our team. [21]

---

[17] **Rationale**: See above
[18] **Rationale**: This is about usability for the LACR team, and is probably the part of the system they will have most interaction with
[19] **Rationale**: This also impacts usability to a large extent
[20] **Rationale**: See above
[21] **Rationale**: As this is a one-time thing, it's pretty unimportant