# Project - XML editor

## Idea

XML (Extensible Markup Language) is one of the most famous formats for storing and sharing information among different devices. Some text editors such as Sublime Text are able to parse such files and do some basic operations. In this project, you will work on developing a GUI (Graphical User Interface) based program to parse and visualize an XML file.

## Main requirements

* Building a GUI in which the user can specify the location of an input XML file.
* Checking the XML consistency: The input XML may have inconsistencies like missing any of the closing and opening tags or not matching tags.The program should be able to detect and visually show any errors in consistency. Optimally, the program will also be able to automatically solve the errors.

|  |  |
| --- | --- |
|  | **This input has 3 errors as indicated by the arrows.** |

* Formatting (Prettifying) the XML: the XML file should be well formatted by keeping the indentation for each level.

|  |  |
| --- | --- |
|  |  |

* Converting XML to JSON: JSON (Javascript Object Notation) is another format that is used to represent data. It’s helpful to convert the XML into JSON, especially when using javascript as there’s tons of libraries and tools that use json notation.

|  |  |
| --- | --- |
| **This is the result of converting the XML example input into JSON.** |  |

* Minifying the XML file: Since spaces and newlines (**\n**) are actually characters that can increase the size of an XML document. This feature should aim at decreasing the size of an XML file (compressing it) by deleting the whitespaces and indentations.
* Compressing the data in the XML/JSON file: You should come-up with a way to reduce the size of the file using a data compression technique. You can invent your own ad-hoc method for such compression. On the other hand, you can check how JSONH works and try to distill ideas from it. Finally, you can use a data compression technique such as byte pair encoding (<https://en.wikipedia.org/wiki/Byte_pair_encoding>).  
  The smaller the output file is, the more efficient your algorithm is.

## Sample input files

Some sample XML files can be found here: <https://drive.google.com/open?id=1FHRPUOEguIFw3jYausDRwplSiYtZANUQ>

These files are part of WordNet [[1]](#footnote-1) which is a way to define words in a dictionary and store the relations between different words.

## Submission Requirements

* The program should be working properly.
* The program should have a simple GUI window containing a set of buttons that are used to choose the operation that needs to be done on the XML file.
* The GUI should allow the user to choose the XML file that will be used.
* The GUI should allow viewing the input XML and the output of each operation as much as possible (If you face problems with large files then you can view only the first lines of the file).
* The GUI should allow the user to save the results of the operations to a new file.

## Why is the project worth doing

In this project, you will learn how to understand and parse XML and json files. Additionally, you will work on designing a GUI (Graphical User Interface) to visualize XML and json files. You should also work on designing the program such that it makes use of optimal data structures for the implemented features.

1. [↑](#footnote-ref-1)