1. **Team members:**

Sajal Jain ([sjain68@wisc.edu](mailto:sjain68@wisc.edu))  
Zachary Huang ([zhuang333@wisc.edu](mailto:zhuang333@wisc.edu))  
CONG GU ([cgu38@wisc.edu](mailto:cgu38@wisc.edu))

1. **A description of the two Web data sources that you have selected.**

Goodreads and IMSA library.

Goodreads (<https://www.goodreads.com/>) is a "social cataloging" website helping users find books.

IMSA library (<https://vufind.carli.illinois.edu/vf-ims/>) is an Information Resource Center provided by Illinois Mathematics and Science Academy.

1. **A description of how you have extracted structured data from the two data sources.**

Manually identify the HLRT Wrappers and extracts attributes.

HLRT Wrappers work well because pages have the same string delimiters.

1. **describe the type of entity you extract, briefly describe the two tables, list the number of tuples per table.**

Book names, authors and publication years.

For Goodreads, there are 5934 tuples.

For IMSA library, there are 6740 tuples.

1. **the names of open-source tools you have used in this project stage and a brief description of what they do.**

No special open-source tool besides python 3 (High-level programming language) and urllib.request (Python module to receive HTML from url)