**DATASET OVERVIEW**

The dataset has been taken from IIF. Collected by Cai-Nicolas Ziegler in a 4-week crawl from the [Book-Crossing](http://www.bookcrossing.com/) community with kind permission from Ron Hornbaker, CTO of [Humankind Systems](http://www.humankindsystems.com/). It contains 278,858 anonymous users (anonymized but with demographic information) providing 1,149,780 ratings (explicit / implicit) for about 271,379 books. Data is comma-separated values (CSV). Since the dataset is of huge size, github is not allowing me to upload the dataset.

The Book-Crossing dataset comprises 3 tables.

* BX-Users: Contains the users. Note that user IDs (`User-ID`) have been anonymized and map to integers. Demographic data is provided (`Location`, `Age`) if available. Otherwise, these fields contain NULL-values.
* BX-Books: Books are identified by their respective ISBN. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given (`Book-Title`, `Book-Author`, `Year-Of-Publication`, `Publisher`), obtained from Amazon Web Services. Note that in case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours (`Image-URL-S`, `Image-URL-M`, `Image-URL-L`), i.e., small, medium, large. These URLs point to the Amazon web site.
* BX-Book-Ratings: Contains the book rating information. Ratings (`Book-Rating`) are either explicit, expressed on a scale from 1-10 (higher values denoting higher appreciation), or implicit, expressed by 0.

**Reference:** Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, Georg Lausen; Proceedings of the 14th International World Wide Web Conference (WWW '05), May 10-14, 2005, Chiba, Japan. To appear. Here is the source link of dataset: http://www2.informatik.uni-freiburg.de/~cziegler/BX/