CS 638 Project Stage 2 Report

Zhiwei Fan   Lingfeng Huang   Fang Wang
zfan29@wisc.edu   lhuang58@wisc.edu   fwang64@wisc.edu

October 18, 2016

Source Websites and Relevant Documentation

1. Source Websites:
   Barnes&Noble: *http://www.barnesandnoble.com*
   Goodreads: *https://www.goodreads.com*

2. Corresponding Tables:
   tableA.csv: table transformed from **Barnes&Noble**'s data
   tableB.csv: table transformed from **Goodreads**'s data

3. Original Schema:
   tableA(id: integer, title: string, year: integer, pages: integer, day: integer, publisher: string, month: integer, authors: string, isbn3: string)
   tableB(id: integer, title: string, isbn: string, pageCount: integer, author: string, publisher: string, date: string)

Schema Transformation

According to *Original Schema* in previous section, tableA and tableB originally have different schemas. After careful consideration, we decide to transform both tables to be in the consistent schema shown as the following:

table(id: integer, title: string, author: string, publisher: string, pages: integer, year: integer, month: integer, day: integer, isbn: string)

Specific Steps of Transformation

Most attributes in *tableA* remain the same. The only change made to *tableA* was *renaming* attribute isbn3 to *isbn*. For B, the attribute *pageCount* was renamed to *page*; attribute *date* has been properly parsed and split into three separated attributes *year*, *month* and *day* by running appropriate python scripts on the original csv files *tableA.csv* and *tableB.csv*.

Set of Attribute: S

According to the transformation steps we have described in the previous part, the attributes of the consistent schema in *set S* is:
id, title, author, publisher, pages, year, month, day, isbn.