**Extract Structed Data of Books from Amazon and Google**

Lan Bai, Chaoqun Mei, Yuzhe Ma

March 22, 2018

**Abstract:**

**1 Data Sources**

We selected Amazon books and Google books as the two Web data sources for this project. Since there are billions of books on Amazon books and Google books, we narrowed down to fiction, health, and business books. For each book, the publisher, time, author, language, and title are extracted as a tuple.

**2. Data Extraction**

We extracted the structured data of books from Amazon and Google using Scrapy. Scrapy is a free and open source web crawling framework, written in Python.