Introduction

- **What is Data Mining?** Data mining is extracting relevant data from huge data sets and using it in a way to help target services and materials to fit the needs of patrons.
- **Big Data:** Big data sets that are studied so you can find relationships, patterns, and trends related to materials like human behavior and interactions; usually too massive to analyze, store, or manage.
- **By using data mining, you have a better chance of predicting the outcome of a project that you can implement in your library.**
- **This process is also called Knowledge Discovery,** Pattern Analysis, Information Harvesting, etc.

Why do we Need Data Mining?

- **Because we have the data but we have not found the knowledge behind that data.**
- **With data mining, the information you find can help you find innovative ways to use your budget more efficiently.**
- **We can use data to track:**
  - Services and materials are needed
  - Services that can be blended or improved
  - Track trends to our patrons’ information needs and how to address those needs
  - Support Decision Making Processes

Where can we find Big Data?

- **Transactions done in the library and other reporting tools**
- **News**
- **Social Media (Facebook, Youtube, Twitter, etc.)**
- **Online Subscriptions**
- **Patron Requests for Information**

Examples of Data Mining in Digital Spaces

**Example of Bibliomining**

- **Article:** “Gaining Strategic Advantage Through Bibliomining: Data Mining for Management Decisions in Corporate, Special, Digital, and Traditional Libraries” by Scott Nicholson and Jeffrey Stanton
- **This article talks about understanding patron behaviors/interests and external data sources through data mining techniques, then using the information to make informed decisions on library materials and services (help predict the need of such resources).**

**Example of Web Mining**

- **2004 Article:** “Toward web mining of cross-language query translations in digital libraries” by Jenq-Haur Wang, et al.
- **Article is about the use of a query-translation engine (LiveTrans), which was developed by using a novel integration of dictionary resources and Web-mining approaches**
- **The search engine uses web resources to help translate information that are sent in by the user**
- **Many resources/articles are uploaded online in its original language, which makes it hard for people (who don’t already know the language) to access the information.**

Types of Data Mining

- **Bibliomining** is the use of data mining strategies to study library services.
- **Web Mining** is finding and extracting information with the data mining strategies on the internet (WWW) to find useful information.
- **Text Mining** is finding high quality information in text through the use of data mining strategies.

Selected Citations

- **Nicholson, S., & Stanton, J. (2003).** Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. \[http://surfaced.uw.edu/cgi/viewcontent.cgi?article=1112\&context=splpub\]