

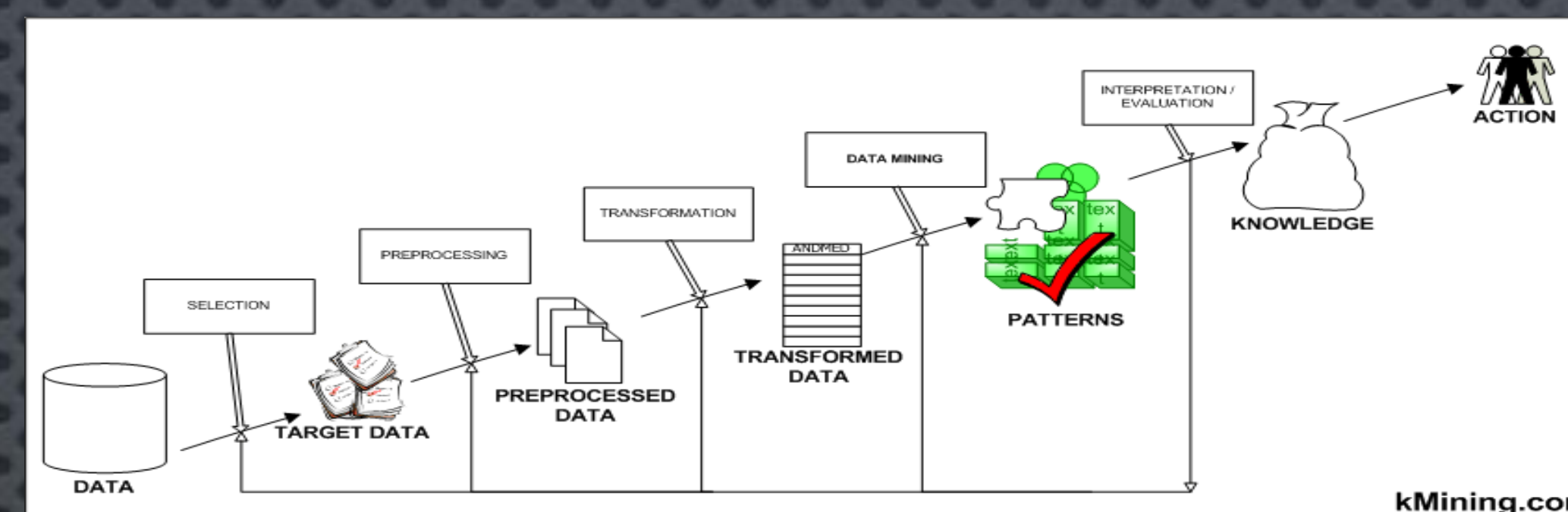
DATA MINING IN DIGITAL SPACES: INTRODUCTION TO THE BASICS

CHARLENE "CHARLIE" ARTHUR, MLISC CANDIDATE
UNIVERSITY OF HAWAI'I AT MANOA



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.



<http://bit.ly/2fx0uLo>

How Data Mining Works

1. Find the data relevant to your project
2. Preprocess the data: this is where you "clean" the data to make it useable
3. Transform the Data so that you can find the potentially useful and/or relevant data among the useable data collected (can combine info)
4. Data Mining: Look for useful patterns within the data to help you with your predictions for your project
5. Evaluate the information and your patterns based on your project parameters, and present it in an easy-to-understand format

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 though 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://surface.syr.edu/cgi/viewcontent.cgi?article=1112&context=istpub>
- Sultan M. Al-Daihani, Alan Abrahams (March 2016). *A Text Mining Analysis of Academic Libraries' Tweets.* The Journal of Academic Librarianship, Volume 42, Issue 2, <http://www.sciencedirect.com/science/article/pii/S0099133315003092>.
- Jenq-Haur, W., Wen-Hsiang, L., & Lee-Feng, C. (2004). *Toward Web mining of cross-language query translations in digital libraries.* International Journal On Digital Libraries, 4(4), 247-257. doi:10.1007/s00799-004-0091-y

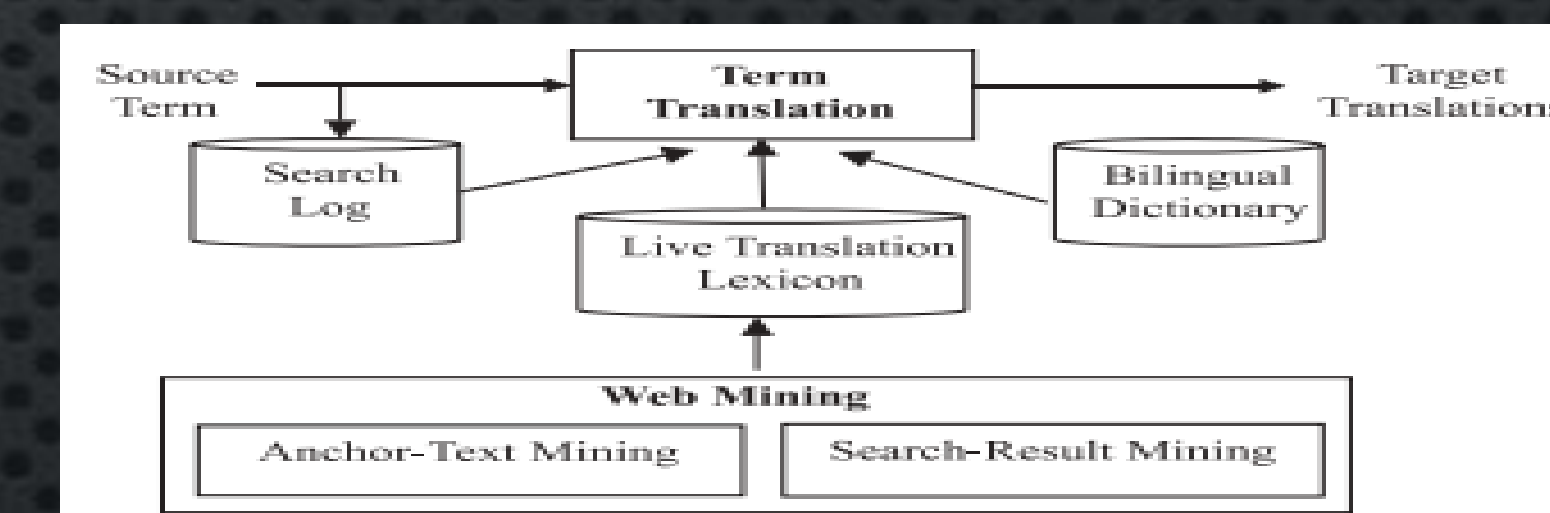
Examples of Data Mining in Digital Spaces

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.



Pic taken from the article above

Example of Text Mining

- Article: "A Text Mining Analysis of Academic Libraries' Tweets" by A.M. Sultan & Alan Abrahams
- Article uses text mining strategies when looking at students' posts on social media to help evaluate and market their collections and services to better fit the needs and capabilities of the students.
- Topic tracking, information discovery, answering questions, etc.

Contact Information

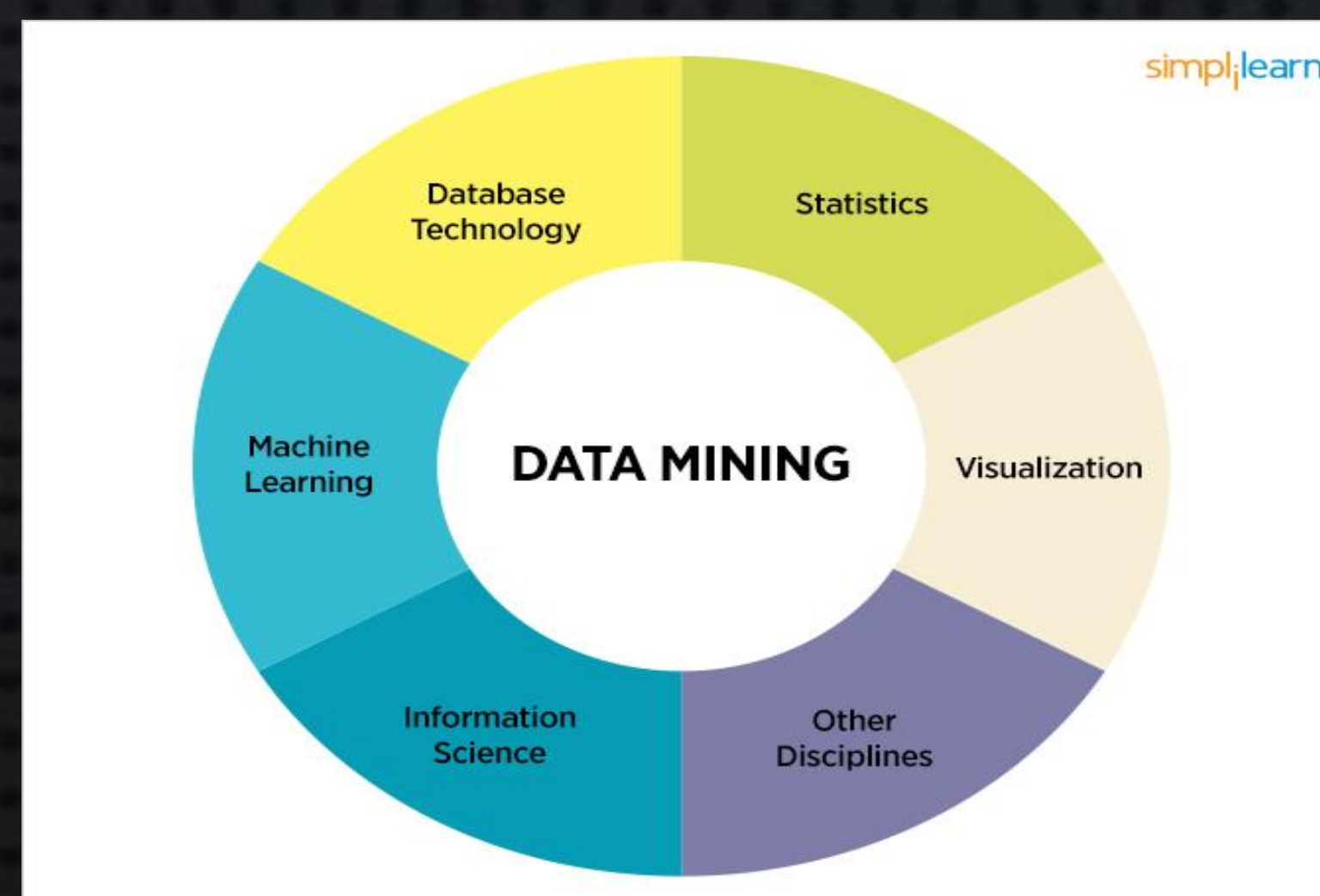
Charlene Arthur,
Library and Information Science Master's Candidate
caarthur@hawaii.edu

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



<http://bit.ly/2eK8EBE>