Topic Modeling Application to Library Management Systems: The Latent Dirichlet Allocation Approach

Babafemi Richard Adebayo

Library Unit, National Open University of Nigeria, Lagos Nigeria

Abstract: **The adoption of Integrated Library Systems (ILS) became prevalent in the 1980s and 1990s as libraries began or continued to automate their processes. These systems enabled library staff work, in many cases, more efficiently than in the past. However, these systems are restrictive and has thus undergone changes over the years, making processes more efficient. One area of improved capabilities is that of "search", which in this paper, builds on integrating topic modeling as a new feature in modern integrated library systems. Users can now partake and explore new ways of resolving text classification and data exploration problems on a typical library management and recommender systems. This aims also at improving book search, browse and help in book-selection decision making, using machine learning tools as the major tweak to the traditional search. It is however hoped that further application of topic modeling will be sought and developed as its usage has continually grown in the past few years.**

Keywords: **Integrated Library System, Latent Dirichlet Allocation, Machine Learning, Searching, Topic Modeling.**

I. INTRODUCTION

1.0. Background of Study:

A **library** is an organized collection of sources of information and similar resources, made accessible to a defined community for reference or borrowing. It provides physical or digital access to material, and may be a physical building or room, or a virtual space, or both. A library's collection can include books periodical, newspapers, manuscripts, films, maps, prints, documents, microform, CDs, cassettes, videotapes, DVDs, Blu-rayDiscs, e-books, audiobooks, databases, and other formats [30].

Now what is an Integrated Library Management System? According to Wikipedia, An **integrated library system** (**ILS**), also known as a **library management system** (**LMS**), is an enterprise resource planning system for a library, used to track items owned, orders made, bills paid, and patrons who have borrowed.

An ILS usually comprises a relational database, software to interact with that database, and two graphical user interfaces (one for patrons, one for staff). Most ILSes separate software functions into discrete programs called modules, each of them integrated with a unified interface. Examples of modules might include:

- acquisitions (ordering, receiving, and invoicing materials)
- cataloguing (classifying and indexing materials)
- circulation (lending materials to patrons and receiving them back)
- serials (tracking magazine and newspaper holdings)
- the OPAC (public interface for users). [30]

Library Integration as it involves Topic Modeling:

Traditional library softwares have areas in form of modules in time past that has been worked on. However, little or nothing has been done in the integration of topic modeling into research techniques and overall search delivery.

Topic Modeling, in this case, Latent Dirichlet Allocation, as explained by [5] David M. Blei, are a suite of algorithms that uncover the hidden thematic structure in document collections. These algorithms help us develop new ways to search, browse and summarize large archives of texts.

It highlights new and fast search methods to extract topics from a large volume of texts or texts without a definitive topic, either by being unsure or otherwise.

II. LITERATURE REVIEW

2.0 Introduction:

Yongming Wang and Trevor A. Dawes [8] stated that since the late 1980s and early 1990s, the library automation system has gone from inception to rapid implementation to near ubiquitous adoption. But after two decades of changes in information technology, and especially in the last decade, the library has seen itself facing tremendous changes in terms of both resources and services it provides. On the resource side, print material and physical items are no longer dominant collections; electronic resources are fast outpacing physical materials to become the dominant library resources, especially in the academic and special libraries. In addition, many other digital format resources, such as digital collections, institutional repositories, and e-books have taken root. On the service front, library users accustomed to immediate and instant searching, finding, and accessing information in the Google age demand more and more instant and easy access to library resources and services.

However, Integrated Library Systems have been part of college and university computing systems since the early 1980s and would seem to be "old technology" and part of text-based mainframe systems. Today's ILS is a multifunction web-based multimedia content information management system, generally built on a standard relational database structure. While the system architecture remains grounded in bibliographical citations presented via structured indexes, the basis of these indexes is moving beyond the MARC fields designed for text information to include metadata description for multiple digital file formats and contents.

The one outstanding change in ILS that emerged during the 1990's is the linkage between bibliographical citations and the content they represent. This content initially came as Table of Contents records linked to the citations. It has now become linkage to the full content with that content going beyond text to sound, images, and full motion video. ILS provide indexing at the bibliographical unit level but also allow indexing within and across full content. Through common cross database indexing and inquiry structures, a single query can retrieve citations and content from multiple databases on variant software platforms. What makes the multi-database searching vital to efficient information retrieval is the combination of structured results sets, elimination of duplicate responses, and retrieval from indexing interior to content files [9].

MACHINE LEARNING

Over the past two decades, machine learning has become one of the main stays of information technology, and with that, a rather central, albeit usually hidden part of life. With the ever increasing amounts of data becoming available, there is good reason to believe that smart analysis will become even more pervasive as a necessary ingredient for technological progress.

Machine learning tasks can be of several forms:

In supervised learning, the computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs. Spam filtering is an example of supervised learning, in particular classification, where the learning algorithm is presented with email (or other) messages labelled beforehand as "spam" or "not spam", to produce a computer program that labels unseen messages as either spam or not.

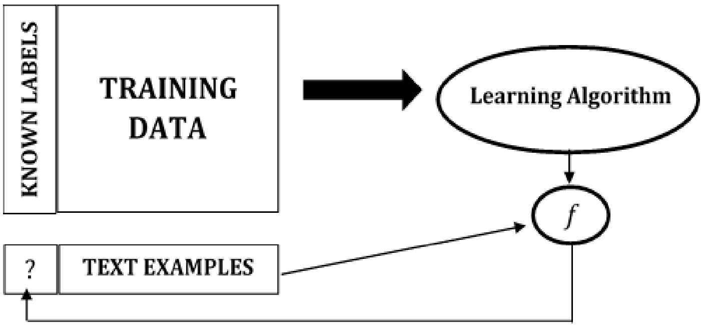


Fig 1: The General Supervised Learning Approach to Machine Learning

In **unsupervised learning**, no labels are given to the learning algorithm, leaving it on its own to groups of similar inputs (clustering), density estimates or projections of high-dimensional data that can be visualised effectively [25]. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end. **Topic modeling** is an example of unsupervised learning, where a program is given a list of human language documents and is tasked to find out which documents cover similar topics.

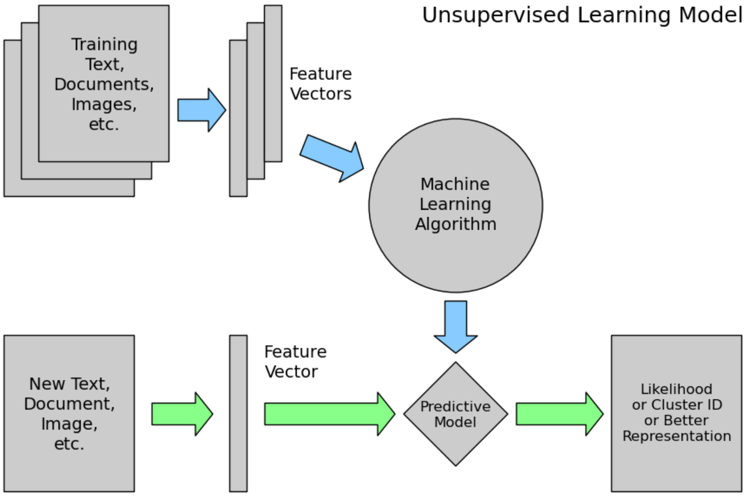

Unsupervised Learning Model
Training Text, Documents, Images, etc.
Feature Vectors
Machine Learning Algorithm
New Text, Document, Image, etc.
Feature Vector
Predictive Model
Likelihood or Cluster ID or Better Representation

Fig 2: An Unsupervised Learning Model.

In reinforcement learning, a computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle), without a teacher explicitly telling it whether it has come close to its goal or not [31].

APPROACHES TO MACHINE LEARNING

- Decision Tree Learning: It is a classic and natural model of learning. It is closely related to the fundamental computer science notion of "divide and conquer".

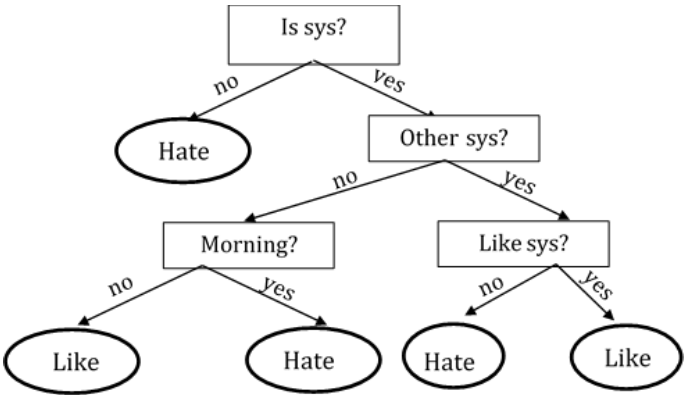


Fig 3: A Decision Tree for a course recommender system.

- *Association Rule Learning:* Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness [21].
- Artificial Neural Networks
- Inductive Logic Programming
- Support Vector Machines
- Clustering
- Bayesian Networks
- Reinforcement Learning
- Representation Learning
- Similarity and Metric Learning
- Sparse Dictionary Learning

TOPIC MODELING OVER TRADITIONAL SEARCH

Topic modeling is a relatively new, completely automated text mining technique that can extract semantic topics from any large collection of text items. These semantic topics can be useful as subjects and can be used to organize poorly categorized collections of objects, while traditional search only tries to match keywords against every word in a database, without necessarily employing intelligent coordination and arrangement of themes.

LDA AS AN APPROACH TO TOPIC MODELING

In many different fields we are faced with a ton of information: think Wikipedia articles, blogs, Flickr images, astronomical survey data, and we need algorithmic tools to organize, search, and understand this information. Topic modeling is a method for analysing large quantities of unlabelled data [23].

A topic is a probability distribution over a collection of words and a topic model is a formal statistical relationship between a group of observed and latent (unknown) random variables that specifies a probabilistic procedure to generate the topics - a generative model [23].

Latent Dirichlet Allocation (LDA) is arguable the most popular topic model in application; it is also the simplest.

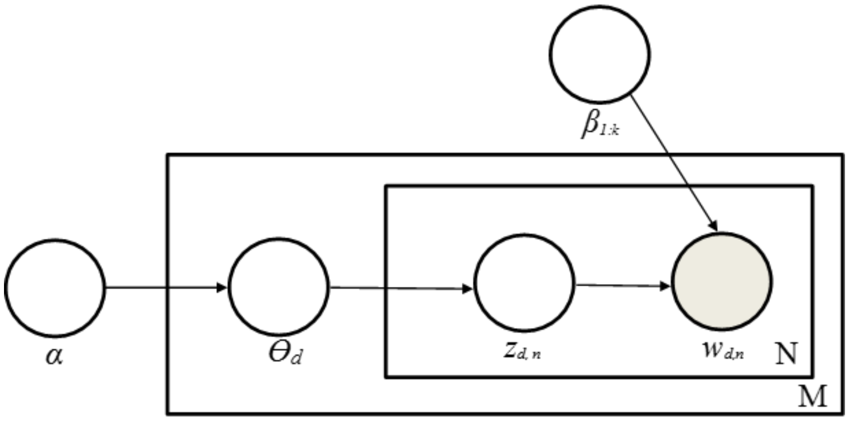


Fig 4: LDA Graphical model

First and foremost, LDA provides a generative model that describes how the documents in a dataset were created. In this context, a dataset is a collection of D documents. But what is a document? It's a collection of words. So the generative model describes how each document obtains its words. Initially, let's assume we know K topic distributions for our dataset, meaning K multinomial containing V elements each, where V is the number of terms in our corpus. Let β_i represent the multinomial for the *i-th* topic, where the size of β_i is V : $|\beta_i|$= V .

Given these distributions, the LDA generative process is as follows:

1. For each document:

(a) Randomly choose a distribution over topics (a multinomial of length K)

(b) For each word in the document:

(i) Probabilistically draw one of the K topics from the distribution over topics obtained in (a), say topic β_j

(ii) Probabilistically draw one of the V words from β_j

This generative model emphasizes that documents contain multiple topics. For instance, a health article might have words drawn from the topic related to seasons such as winter and words drawn from the topic related to illnesses, such as flu. Step (a) reflects that each document contains topics in different proportion, e.g. one document may contain a lot of words drawn from the topic on seasons and no words drawn from the topic about illnesses, while a different document may have an equal number of words drawn from both topics. Step (ii) reflects that each individual word in the document is drawn from one of the K topics in proportion to the document's distribution over topics as determined in Step (i). The selection of each word depends on the distribution over the V words in our vocabulary as determined by the selected topic, β_j. Note that the generative model does not make any assumptions about the order of the words in the documents, this is known as the bag-of-words assumption [23].

The central goal of topic modeling is to automatically discover the topics from a collection of documents. Therefore our assumption that we know the K topic distributions is not very helpful; we must learn these topic distributions. This is accomplished through statistical inference [23].

To formalize LDA, first restate the generative process in more detail (compare with the previous description):

1. for each document:

(a) Draw a topic distribution, $\Theta_d \sim$ Dir (α), where Dir (.) is a draw from a uniform Dirichlet distribution with scaling parameter _

(b) For each word in the document:

(i) Draw a specific topic $z_{d,n} \sim$ multi(Θ_d) where multi(.) is a multinomial

(ii) Draw a word $w_{d,n} \sim z_{d,n}$

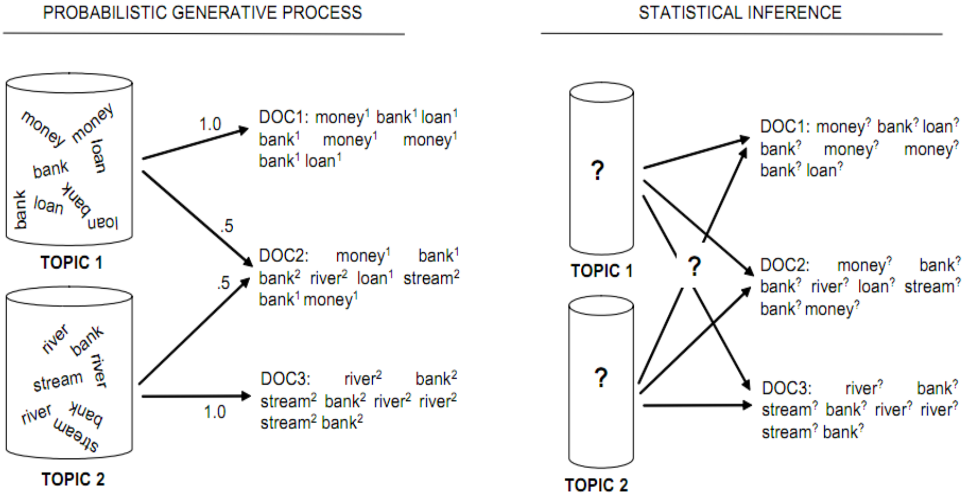

PROBABILISTIC GENERATIVE PROCESS
STATISTICAL INFERENCE
money money loan bank bank loan bank loan
TOPIC 1
river bank river stream river bank stream
TOPIC 2
1.0
.5
.5
1.0
DOC1: money^1 bank^1 loan^1 bank^1 money^1 money^1 bank^1 loan^1
DOC2: money^1 bank^1 bank^2 river^2 loan^1 stream^2 bank^1 money^1
DOC3: river^2 bank^2 stream^2 bank^2 river^2 river^2 stream^2 bank^2
?
TOPIC 1
?
TOPIC 2
?
DOC1: money$^?$ bank$^?$ loan$^?$ bank$^?$ money$^?$ money$^?$ bank$^?$ loan$^?$
DOC2: money$^?$ bank$^?$ bank$^?$ river$^?$ loan$^?$ stream$^?$ bank$^?$ money$^?$
DOC3: river$^?$ bank$^?$ stream$^?$ bank$^?$ river$^?$ river$^?$ stream$^?$ bank$^?$

Figure 5: Left: a visualization of the probabilistic generative process for three documents,

i.e. DOC1 draws from Topic 1 with probability 1, DOC2 draws from Topic 1 with probability 0.5 and from Topic 2 with probability 0.5, and DOC3 draws from Topic 2 with probability 1.

Right: In the inferential problem we are interested in learning the topics and topic distributions [26].

APPLICATIONS WHERE LDA HAS BEEN USED

a) LDA Based User-Tag Model for Automatic Image Geo-Tagging.

b) Applications of LDA to Document Modeling

c) Applications of LDA to Automatic Harmonic Analysis

d) Geometric Latent Dirichlet Allocation on a Matching e) Graph for Large-Scale Image Datasets

III. METHODOLOGY: INTEGRATING TOPIC MODELING (LDA) INTO WEB-BASED INTEGRATED LIBRARY SYSTEM

3.0 Introduction:

LDA as a topic modeling technique provides a generative model that describes how the documents in a dataset were created. In this context, a dataset is a collection of D documents. But what is a document? It's a collection of words [23].

Latent Dirichlet Allocation (LDA) is a generative probabilistic model of a corpus. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words [6].

LDA assumes the following generative process for each document **w** in a corpus *D*:

1. Choose $N \sim \text{Poisson}(\xi)$.

2. Choose $\Theta \sim \text{Dir}(\alpha)$.

3. For each of the *N* words w_n:

   (a) Choose a topic $z_n \sim \text{Multinomial}(\Theta)$.

   (b) Choose a word w_n from $p(w_n \mid z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

Several simplifying assumptions are made in this basic model. First, the dimensionality *k* of the Dirichlet distribution (and thus the dimensionality of the topic variable *z*) is assumed known and fixed. Second, the word probabilities are parameterized by a $k \times V$ matrix β where $\beta_{ij} = p(w^j = 1 \mid z^i = 1)$, which for now we treat as a fixed quantity that is to be estimated. Finally, the Poisson assumption is not critical to anything that follows and more realistic document length distributions can be used as needed. Furthermore, note that *N* is independent of all the other data generating variables *(Θ and **z**)*. It is thus an ancillary variable and we will generally ignore its randomness in the subsequent development [6].

HOW LDA CAN BE INTEGRATED INTO A WEB-BASED LIBRARY MANAGEMENT SYSTEM

A Library web-based system literally manages online textual and graphical resources, which in turn is made available to users for research and of course, extract information.

Some of the features of an E-Library are listed below:

This section outlines the main requirements that relate to the end users who borrow and download eBooks from the E-Library.

Online User Registration

Users must be able to create Patron or Student accounts by registering online. A suitable registration page or pages should be provided.

New users must go thru a verification process (TBD) to confirm their email address before their account is activated in the system.

User Login/out

Once verified, users must be able to log in to the portal.

Users must be able to change their password (but not their user name/ID)

There must be a way for users to retrieve a lost password

Users must be able to log out. There must be a session timeout mechanism that will automatically log the user out after a period of time (period TBD).

On logout, the contents of user's Select List and Cart are cleared. The contents of the Reserved List is preserved.

Catalog Search

Users must be able to search for eBooks

The system should provide interfaces for both "simple" and "advanced" searches

- "simple" should be a search on any one of fields *Author*, *Title*, or *ISBN*.
- "advanced" should provide an easy way to construct complex searches on multiple fields with different logic (e.g. AND, OR, NOT)

Catalog Quick Search

The system should provide several "quick search" options:

- Newest eBooks – returns a listing of the 10- 50 (configurable) most recently added eBooks
- Top 20 Titles – returns a listing of the 20 most frequently borrowed eBooks

○ The actual number should be configurable by an administrator

○ The UI should provide a pick list of values: 10, 20, 25, 50, 100

- Latest Returns – returns a list of 10-50 (configurable) most recently checked-in eBooks

○ Either explicitly checked in by a patron, or lending period expired freeing the eBook for loan

○ eBooks listed in this set must not have any reservations… must be available for immediate loan

Catalog Browsing

Users must be able to browse the eBooks in the catalog

Users must be able to select how they want to browse:

- Browse by Author

○ Must be able to select an alphabetical subset… e.g. author names beginning with A, or M for example

- Browse by Title

○ Must be able to select an alphabetical subset… e.g. author names beginning with A, or M for example

- Browse by genre (e.g. fiction, non-fiction, science fiction, etc.)

Advanced Catalog Browsing

Users should be able to browse by Publisher

Users should be able to browse by Book Award

Users should be able to browse according to a combination of data, such as by *Author* + *Genre*, *Title* + *Year of Publication*, *Author* + *Book Award*, etc.

Create Select List

The Select List is similar to a shopping cart except that the checkout processes a loan of the eBooks in the list rather than a purchase.

- Each user must be able to save a list of eBooks they want to borrow – their "Select List"
- The Select List is only for the current session and is cleared when the session terminates

- Each eBook added to the Select List remains in the list for 20 minutes. During this time, the number of copies available for loan is reduced by 1 and the number available is shown to other users.

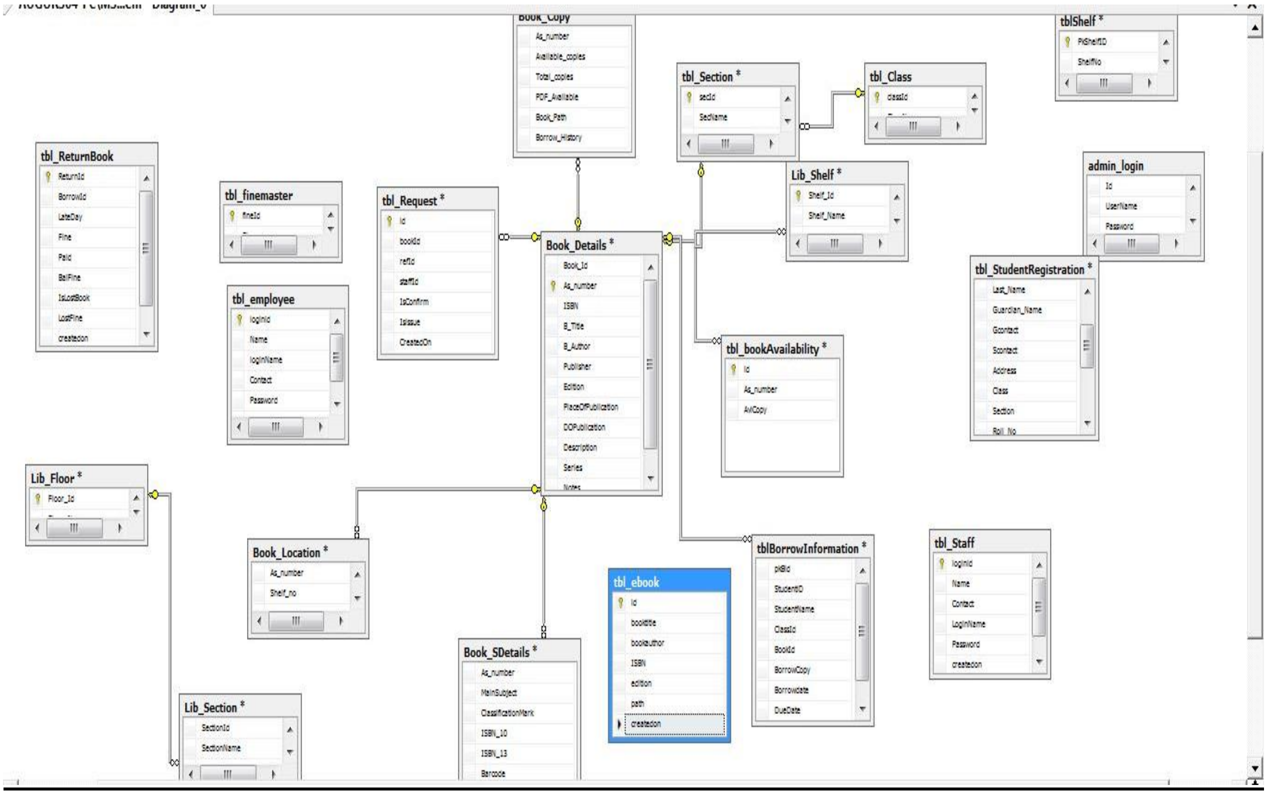
○ If the time expires and the eBook is removed from the Select List, then the number of copies available for loan is increased by 1 and the number available is shown to other users.

- The number of eBooks user may add to their Select List is limited to the checkout limit defined in the system configuration minus the number of eBooks patron has currently checked out.

○ If user currently has the maximum number checked out, then all *Add to Select List* links should be disabled for that user.

Above is an overview/some features of a traditional web-based electronic library system, which has proven effective, However, at this point, it pertinent to see how one can quickly extract topics from “about to be borrowed” books, because it has been observed that books with unrelated titles tend to have related contents. This, can be achieved using LDA right from the click of a butto

DATABASE SCHEMA DIAGRAM



IV. RESULT AND DISCUSSION

System Requirements

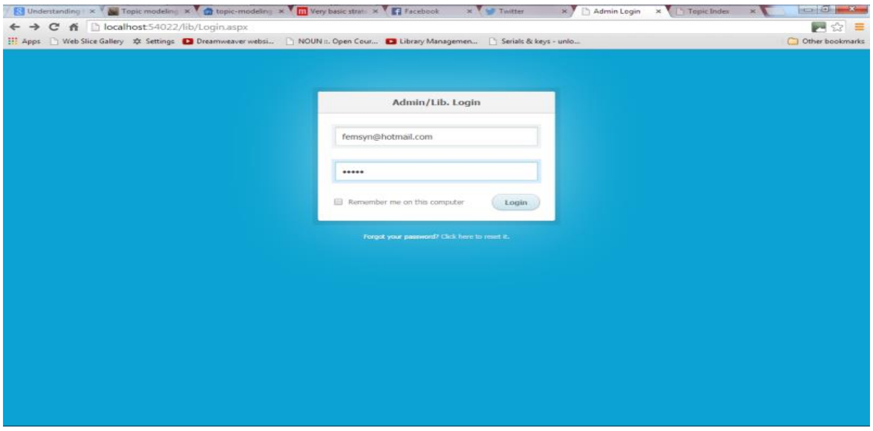
1. Hard Disk: 1G or more
2. Memory: 512MB or more
3. Processor: At least 1GHz
4. Operating System: Any OS, as long as it has browsing features. But preferably, Windows xp or later.
5. System type: 32 or 64bit
6. Software requirements: Any browser with internet connectivity. i.e. Mozilla, Internet Explorer, Google Chrome, Opera, Safari etc.

PROGRAM DESCRIPTION

The program called Integrated Library System with Topic Modeling is web based and was developed using C# and asp.net as it web framework. The back end is SQL based.

Getting on to the site… http://localhost:54022/lib/Default.aspx, you are greeted with the homepage, giving options to either log in as a Staff, Student or Admin.

To use the platform as a student or staff, the user must be created by the admin, after which log in details will be sent to the user's e-mail address, for onward log in.Username for admin is given as: femsyn@hotmail.comAlso, Password for admin is given as: 94295



localhost:54022/lib/Login.aspx

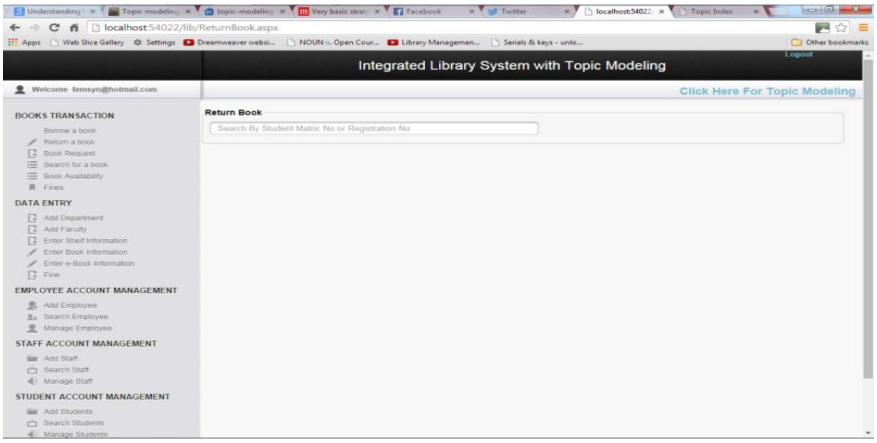
Admin/Lib. Login

femsyn@hotmail.com

•••••

□ Remember me on this computer

Homepage for Admin



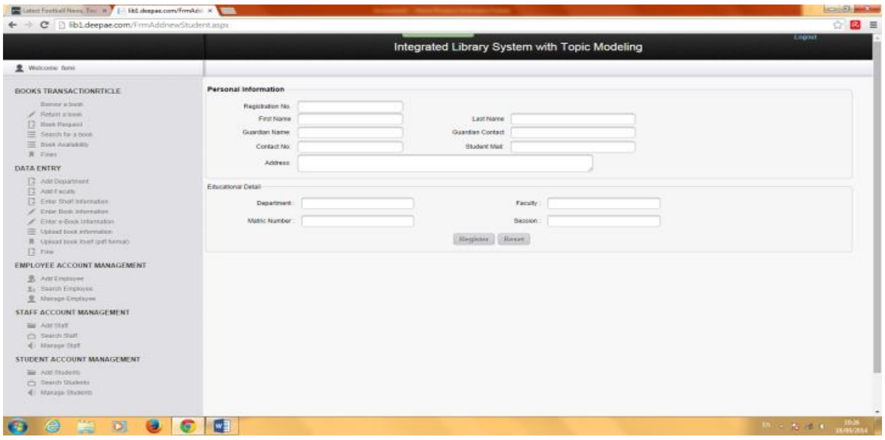
localhost:54022/lib/ReturnBook.aspx

Integrated Library System with Topic Modeling

Welcome femsyn@hotmail.com

Return Book

Adding a Student



lib1.deepae.com/FrmAddnewStudent.aspx

Integrated Library System with Topic Modeling

Welcome femi

Personal Information

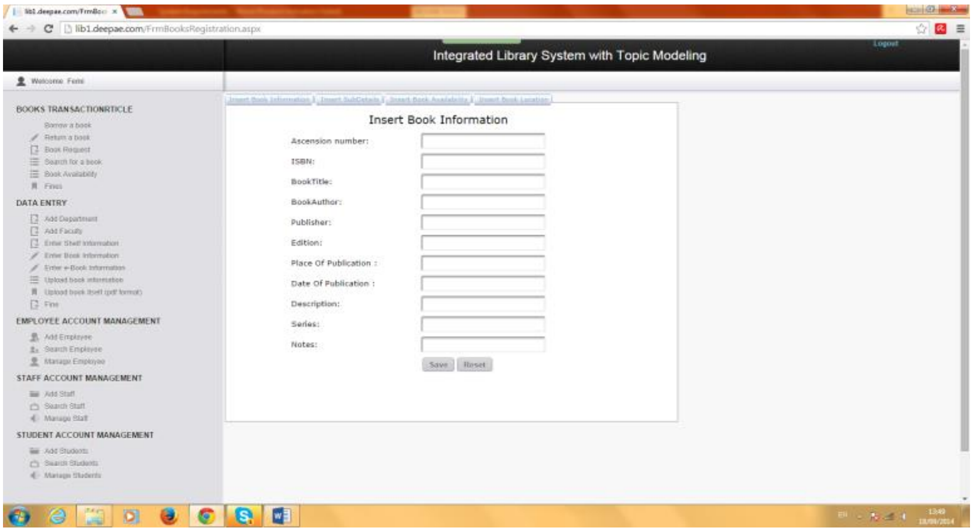
Registration No:
First Name: | Last Name:
Guardian Name: | Guardian Contact:
Contact No: | Student Mail:
Address:

Educational Detail

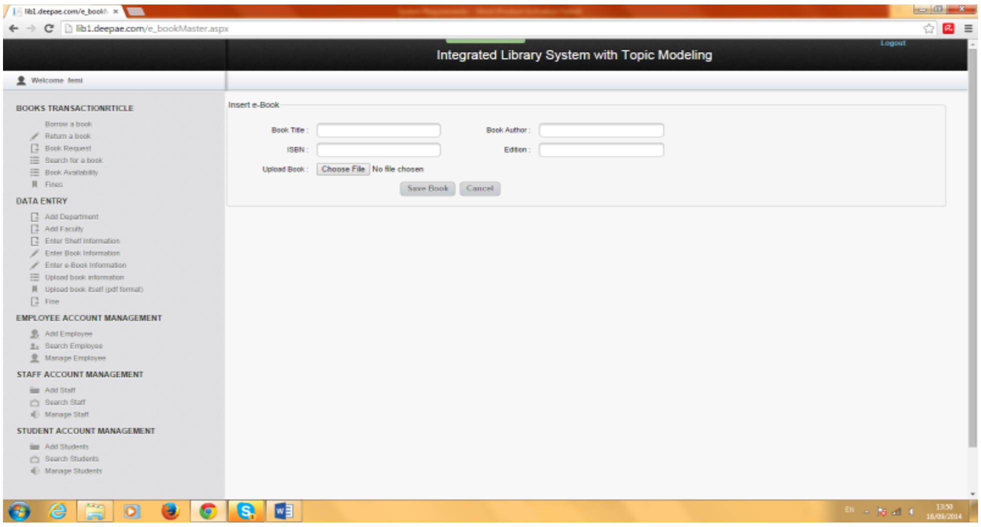
Department: | Faculty:
Matric Number: | Session:

After adding a Student, an e-mail is sent to the student's email address, containing the username and password.

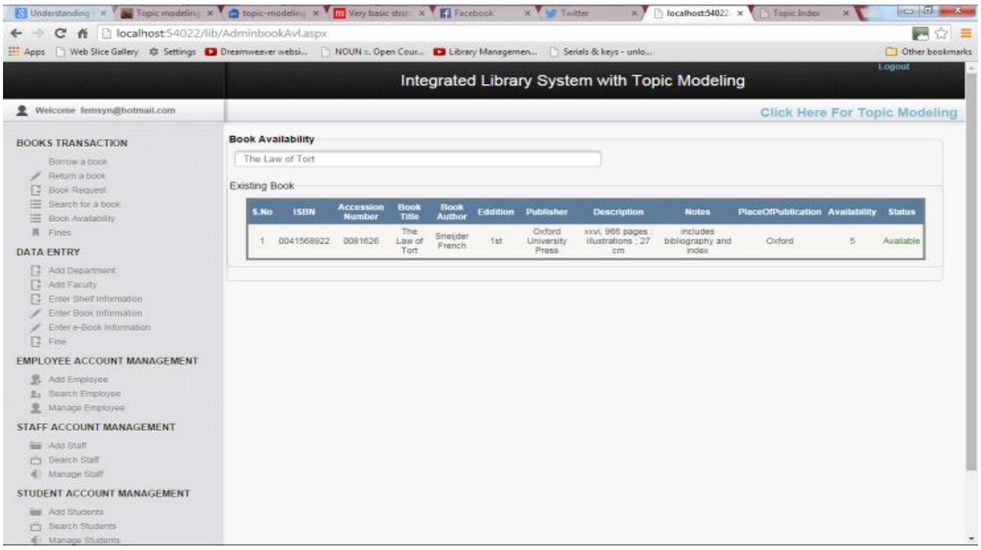
Enter Book Information tab



Enter/Upload E-Book Information



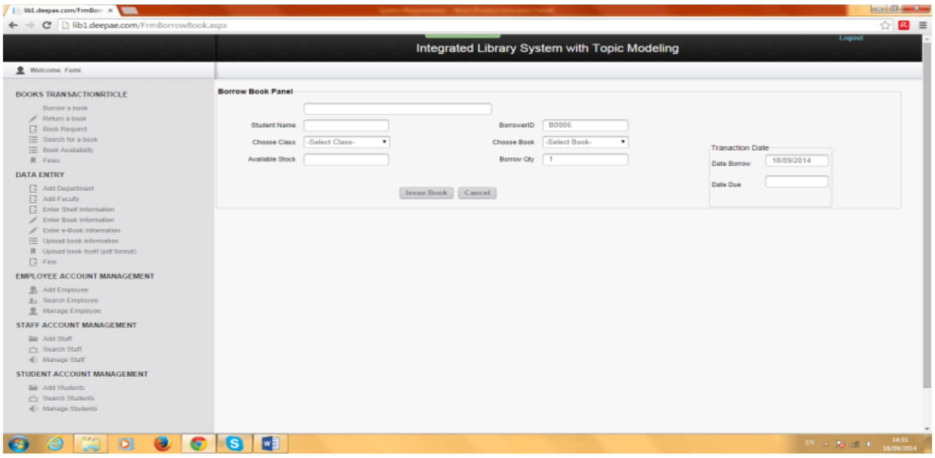
Searching for a book using keywords like Author's name, ISBN or Title


localhost:54022/lib/AdminbookAvl.aspx
Integrated Library System with Topic Modeling
Logout
Welcome femsyn@hotmail.com
Click Here For Topic Modeling
Book Availability
The Law of Tort
Existing Book

S.No	ISBN	Accession Number	Book Title	Book Author	Eddition	Publisher	Description	Notes	PlaceOfPublication	Availability	Status
1	0041568922	0081626	The Law of Tort	Sneijder French	1st	Oxford University Press	xxvi, 966 pages : illustrations ; 27 cm	includes bibliography and index	Oxford	5	Available

This outputs a summary and availability status of the said book. In this case – "The Law of Tort".

Borrow Book Panel

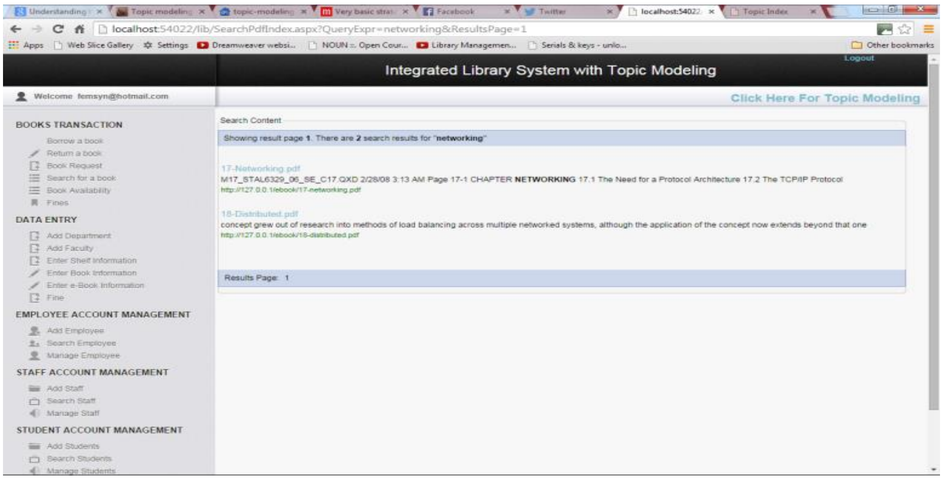


Borrow Book Panel

Student Name
Choose Class: -Select Class-
Available Stock
BorrowerID: B0006
Choose Book: -Select Book-
Borrow Qty: 1
Tranaction Date
Date Borrow: 18/09/2014
Date Due

Searching through book content i.e. pdf files

For example, a keyword – "networking" could be searched throughout the database, by mapping the keyword to every word present and output is given thus:



localhost:54022/lib/SearchPdfIndex.aspx?QueryExpr=networking&ResultsPage=1

Search Content

Showing result page **1**. There are **2** search results for "**networking**"

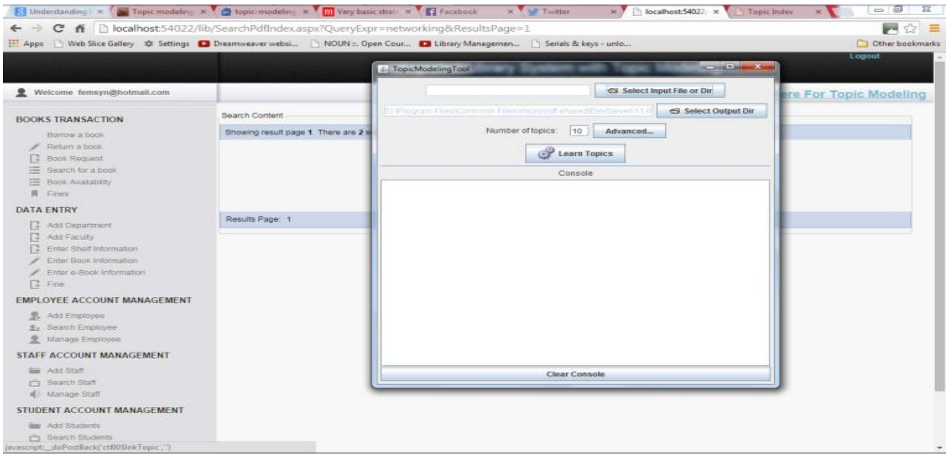
17-Networking.pdf
M17_STAL6329_06_SE_C17.QXD 2/28/08 3:13 AM Page 17-1 CHAPTER **NETWORKING** 17.1 The Need for a Protocol Architecture 17.2 The TCP/IP Protocol
http://127.0.0.1/ebook/17-networking.pdf

18-Distributed.pdf
concept grew out of research into methods of load balancing across multiple networked systems, although the application of the concept now extends beyond that one
http://127.0.0.1/ebook/18-distributed.pdf

Results Page: 1

Patrons can go ahead to view the source files in pdf format and can optionally run topic modeling on it. At every point of these webpages, users can take advantage of topic modeling and use, right from the click of a button.

Performing Topic Modeling



localhost:54022/lib/SearchPdfIndex.aspx?QueryExpr=networking&ResultsPage=1

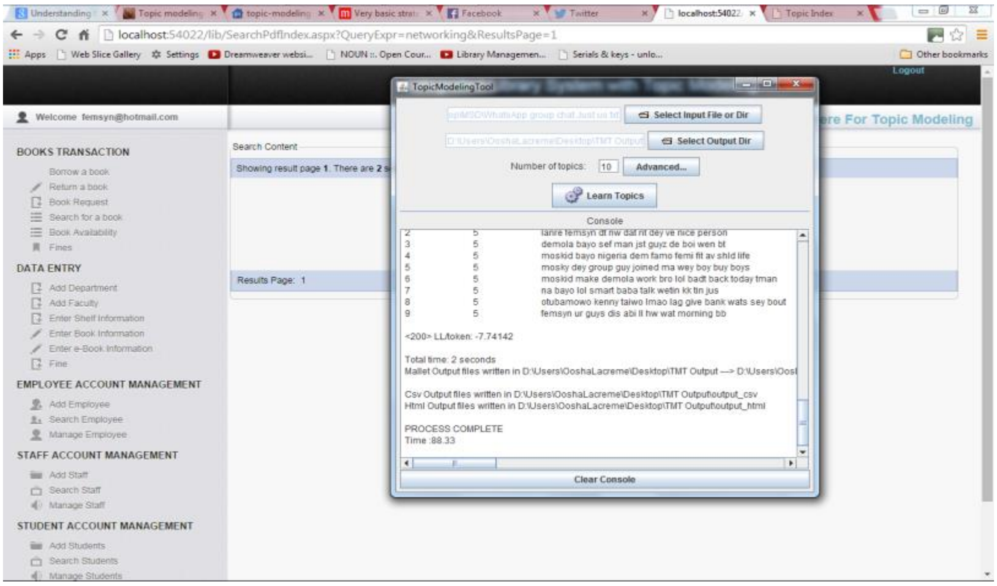
D:\Program Files\Common Files\microsoft shared\DevServer\11.0

Number of topics: 10

javascript:__doPostBack('ctl00$lnkTopic','')

Files are inputted, variables and the output path set. Training is done on the input document and output given for analysis. N.B.: The length of time spent "learning topics" is dependent on the size of document being worked on.

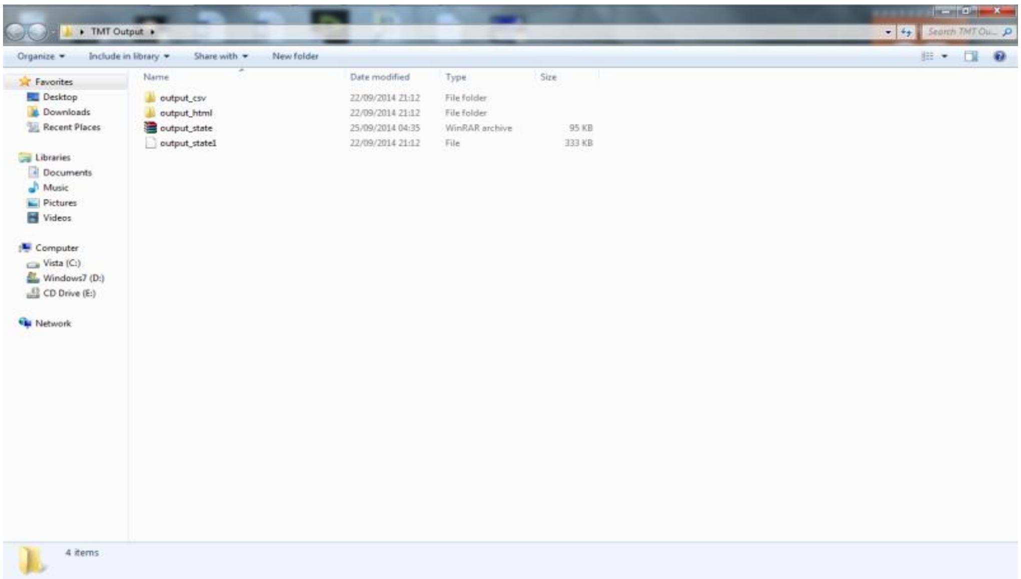
Output



Number of topics: 10

```
2           5          lanre femsyn dt nw dat nt dey ve nice person
3           5          demola bayo sef man jst guyz de boi wen bt
4           5          moskid bayo nigeria dem famo femi fit av shld life
5           5          mosky dey group guy joined ma wey boy buy boys
6           5          moskid make demola work bro lol badt back today tman
7           5          na bayo lol smart baba talk wetin kk tin jus
8           5          otubamowo kenny taiwo lmao lag give bank wats sey bout
9           5          femsyn ur guys dis abi ll hw wat morning bb

<200> LL/token: -7.74142

Total time: 2 seconds
Mallet Output files written in D:\Users\OoshaLacreme\Desktop\TMT Output ---> D:\Users\Oos

Csv Output files written in D:\Users\OoshaLacreme\Desktop\TMT Output\output_csv
Html Output files written in D:\Users\OoshaLacreme\Desktop\TMT Output\output_html

PROCESS COMPLETE
Time :88.33
```

The Output Folder is created and further analysis done.



Name	Date modified	Type	Size
output_csv	22/09/2014 21:12	File folder	
output_html	22/09/2014 21:12	File folder	
output_state	25/09/2014 04:35	WinRAR archive	95 KB
output_state1	22/09/2014 21:12	File	333 KB

5. RECOMMENDATION

Further application of topic modeling should be sought and developed as its usage has continually grown in the past few years.

6. CONCLUSION

It has been seen that this approach will not only prove productive, information-wise, but also help in time management. That is, instead of flipping through already borrowed books (or tons of pages) for hours or even days, only to find out that the content is irrelevant, topic modeling should be employed to do topic extraction for well-informed decisions.

7. CONTRIBUTION TO KNOWLEDGE

User can now partake and explore new ways of search on future library management systems that are useful for basic tasks such as classification, novelty detection, summarization, similarity and relevant judgments.

REFERENCES

[1] Anick Jesdnun, A., n.d. GPS adds dimension to online photos. [Online] [Accessed 19 January 2014].

[2] Anick, J., 2013. GPS adds dimesion to online photos, s.l.: s.n.

[3] Anon., n.d. Perceptions 2008: an International Survel of Library Automation. [Online] Available at: www.library technology.org [Accessed 17 08 2014].

[4] Arenas-Garcia, J. et al., 2007. Unveiling Music Structure via plsa Similarity Fusion. s.l., IEEE International Work shop on Machine Learning for Signal Processing, pp. 419-424.

[5] Blei, D. M., n.d. Princeton University. [Online] Available at: http://www.cs.princeton.edu/~blei/topicmodeling. html [Accessed 1 August 2014].

[6] Blei, D., Ng, A. & I., J. M., 2003. Latent Dirichlet Allocation. Journal of MAchine Learning Research, Volume 3, pp. 993 - 1022.

[7] Codes, G. T. Q., n.d. [Online] [Accessed 2013 October 2013].

[8] DAwes, Y. W. a. T. A., 2012. The Next Generation Integrated Library System: A Promise Fulfilled. Information Technology and libraries, p. 84.

[9] Deddens, M., 2002. Overview of Integrated Library Systems, Cincinnati: Univeristy of Cincinnati.

[10] Eric, G. & Wang, X., n.d. Spatial Latent Dirichlet Allocation, s.l.: s.n.

[11] Hamby, R., McBride, R. & Lundberg, M., 2011. South Carolina's SCLENDS Optimizing Libraries, Transforming Lending. Compuers in Libraries, 31(8), pp. 6 - 10.

[12] Hoffman M., D. B. a. P. C., 2008. Content-based Musical Similarity Computation using the Hierarchical Dirichlet Process. s.l., ICMC.

[13] J. Diane, H., n.d. Latent Dirichlet Allocation for Text, Images and Music. San Diego, Department of Computer Science, University of California.

[14] Kessler, C. & Krumhansl, E., 1983. Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys. Psychological Review, Volume 89, pp. 334 - 368.

[15] Kochtanek, T., 2002. Library Information Systems: From Library Automation to DIstributed Information Access Solutions. Westport, Connecticut: Meckler.

[16] Kozyk, S., 2013. Itegrity Group. [Online] Available at: http://www.itegritygroup.com/asp-net-advantages/[Accessed 18 8 2014].

[17] Krumhansl, C., 1990. Cognitive Foundations of Musical Pitch. Oxford: Oxford University Press.

[18] Microsoft Developer Network, n.d. Microsoft Developer Network. [Online] Available at: http://msdn.microsoft. com/en-us/library/aa664274(v=vs.71).aspx[Accessed 18 8 2014].

[19] Microsoft, n.d. Microsoft. [Online] Available at: http://www.microsoft.com/technet/prodtechnol/WindowsServer 2003/Library/IIS/ad2a22b9-135c-432a-bc9f-c67f074242b7.mspx?mfr=true[Accessed 18 8 2014].

[20] Philbin, J., Sivic, J. & Zisserman, a. A., 2011. International Journal of Computer Vision, 95(2), pp. 138-153.

[21] Piatetsky-Shapiro, G. & Frawley, W. J., 1991. Discovery, analysis, and presentation of strong rules. In: Knowledge Discovery in Databases. Cambridge, MA: AAAI/MIT Press.

[22] Rashmi, K. V. & Giulia, F., n.d. LDA Based User-Tag Model for Automatic Image Geo-Tagging. California, Department of Electrical Engineering and Computer Science, University of California.

[23] Reed, C., 2012. Latent Dirichlet Allocation: Towards a Deeper Understanding, s.l.: Colorado Reed.

[24] Ross, G., 2012. Search Queries in Library Databases. [Online] Available at: http://www.library.illinois.edu/hpnl/ guides/searching.html[Accessed 21 9 2014].

[25] Simon, P., 2013. Too Big to Ignore: The Business Case for Big Data.. s.l.:Wiley.

[26] Steyvers, M. & Grifiths, T., 2007. Probabilistic topic models.. In: Handbook of latent semantic Analysis. s.l.:s.n., pp. 424 - 440.

[27] Teh, Jordan, M. & Blei, M. B. a. D., 2006. Hierarchical Dirichlet Processes. Journal of the American Statistical Assocation, 101(476), pp. 1566 - 1581.

[28] Wallace, P. & Pitkin, G., 1991. Library Systems Migration: An Introduction. Westport, Connecticut: Meckler.

[29] Wikipedia, 2014. SQL. [Online] Available at: http://en.wikipedia.org/wiki/SQL[Accessed 18 8 2014].

[30] Wikipedia, 2014. Wikipedia. [Online] Available at: en.wikipedia.org/wiki/Library[Accessed 12 July 2014].

[31] Wikipedia, n.d. Machine Learning. [Online] Available at: http://en.wikipedia.org/wiki/Machine_learning[Accessed 18 8 2014].

[32] Wikipedia, n.d. Wikipedia. [Online] Available at: en.wikipedia.org/wiki/integrated_library_system[Accessed 10 August 2014].