

CSC 5741

Lecture 1: Administrivia and Course Introduction

Lighton Phiri <lighton.phiri@unza.zm>

Department of Library and Information Science
University of Zambia

Lecture Series Outline

- Part I: Administrivia
- Part II: Course Introduction
- Part III: On Academic Activities
- Part IV: About Next Week

Lecture Series Outline

- **Part I: Administrivia**
 - Personal Introductions
 - Learning Outcomes
 - Course Structure
 - Prescribed Books
 - Tools and Services
 - Course Grading and Academic Dishonesty
 - Course Management
- **Part II: Course Introduction**
- **Part III: On Academic Activities**
- **Part IV: About Next Week**

Personal Introductions

- Your full names and preferred reference (first name, Mrs./Ms.Mr. X)
 - I prefer that you refer to me using my first name: Lighton
- Your formal education background
- What you are presently upto (Think what you do for a living)
- What you hope to get from CSC 5741

CSC 5741 Learning Outcomes

- Identify the key processes of data mining, data warehousing and knowledge discovery process
- Describe the basic principles and algorithms used in practical data mining and understand their strengths and weaknesses
- Apply data mining techniques to solve problems in other disciplines in a mathematical way
- Apply data mining methodologies with information systems and generate results which can be immediately used for decision making in well-defined business problems

Course Structure (1/6)

- CSC 5741 is a half course
- CSC 5741 will be run using a seminar session
 - One three hour-long lecture session per week
 - One seminar every fortnight.
 - Paper reading sessions every fortnight.
 - One class theory tests

Course Structure (2/6)

- Tentative Lecture series and session structure
 - Lecture session
 - Paper discussion
 - Seminar session
- We will tentatively spend two weeks on each CSC 5741 theme

TERM I	
Monday 28 th January, 2019 to Sunday 10 th March, 2019	On-line Registration
Monday 4 th February, 2019 to Friday 8 th February, 2019	Orientation of First Year students
Sunday 10 th February, 2019	Arrival of Returning Regular Students
Monday 18 th February, 2019 to Friday 31 st May, 2019	Lectures for Regular Students (15 weeks)
Monday 4 th March, 2019 to Thursday 7 th March, 2019	Graduation Week (First Graduation Ceremony)
Monday 11 th March, 2019 to Sunday 24 th March, 2019	Late Registration Period (2 Weeks)
Friday 26 th April, 2019	Social and Cultural Day
Monday 13 th May, 2019 to Friday 26 th July, 2019	IDE Students School Experience (11 Weeks)
Monday 20 th May, 2019 to Thursday 24 th May, 2019	Graduation Week (Second Graduation Ceremony)
Monday 3 rd June, 2019 to Friday 7 th June, 2019	Study Break and Post Graduate Seminar Week

<http://www.unza.zm>

Course Structure (3/6)

Create

March 2019 < > MON TUE WED THU FRI SAT

M	T	W	T	F	S	S
25	26	27	28	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31
1	2	3	4	5	6	7

Youth Day

11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Add calendar ::

My calendars ^

- Lighton Phiri
- Contacts
- Department of Library and...
- Department of Library and...
- ICT 1110, 2018
- LIS 4014 Project, 2018
- LIS 5310
- Reminders
- Tasks
- University of Zambia | CSC...
- University of Zambia | EDU...
- University of Zambia IUCT

MON TUE WED THU FRI SAT

11 12 13 14 15 16 17

18 19 20 21 22 23 24

25 26 27 28 29 30 31

1 2 3 4 5 6 7

11:00

12:00

13:00

14:00

15:00

16:00

17:00

18:00

19:00

20:00

21:00

00:00

07:00

08:00

09:00

10:00

11:00

12:00

13:00

14:00

15:00

16:00

17:00

18:00

19:00

20:00

21:00

09:00 – 13:00

Instructor Office Hours

Lecture: LIS 5310 [Class Cancelled]
Faculty Staff Industrial Action
13:00 – 15:00
3B

Lecture: CSC 5741 [Merged with Friday Session]
17:30 – 20:30
Computer Science
Conference/Seminar Room

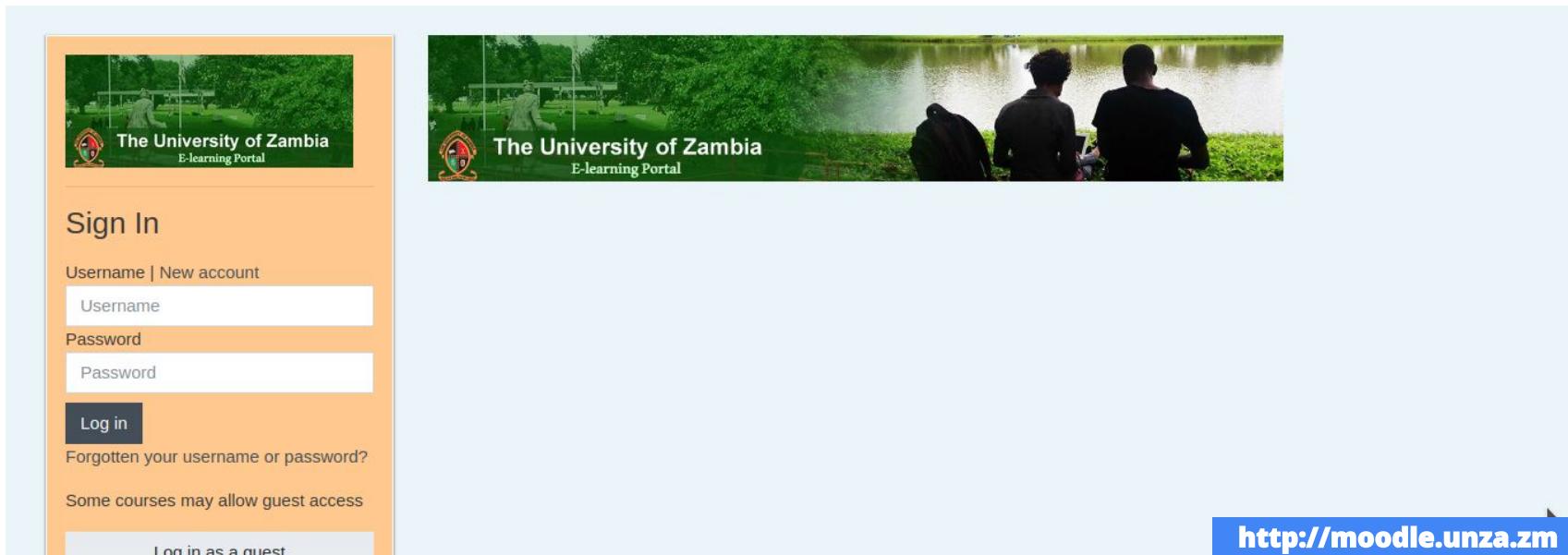
Lecture: LIS 5310: Academic Talk [Cancelled]
16:00 – 17:30
Library Auditorium

Lecture: CSC 5741
17:30 – 20:30

<https://goo.gl/4KLQfE>

Course Structure (4/6)

- Course Resources
 - All course resources will be made available on The Moodle.



The image shows two side-by-side screenshots. On the left is the sign-in page of The University of Zambia E-learning Portal. It features a header with the university's crest and the text "The University of Zambia E-learning Portal". Below this is a "Sign In" section with fields for "Username" and "Password", and a "Log in" button. There are also links for "Forgotten your username or password?" and "Some courses may allow guest access". At the bottom is a "Log in as a guest" button. On the right is a photograph of two students sitting by a lake, viewed from behind, looking at a laptop screen. The university's crest and name are overlaid on the top left of this image.

<http://moodle.unza.zm>

Course Structure (5/6)

- Course Resources
 - All course resources will be made available on The Moodle.

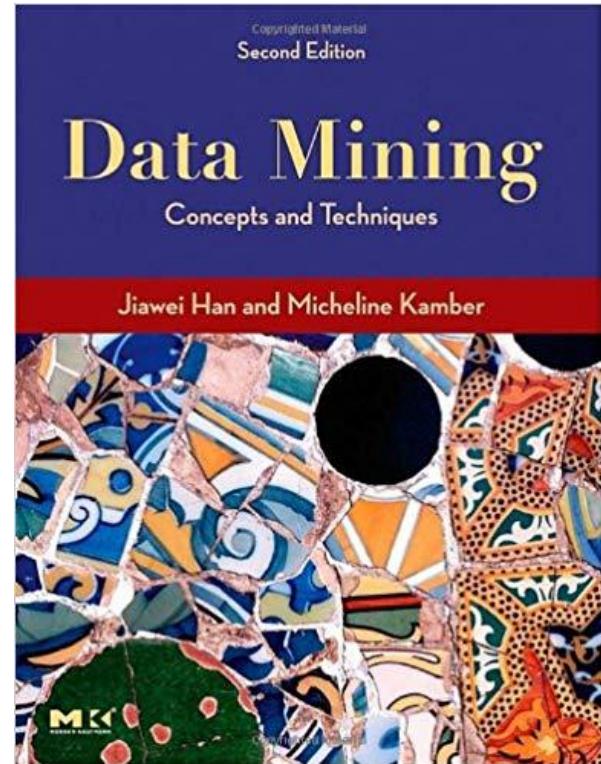
The screenshot shows the Moodle course page for LIS 5310, 2019. The top navigation bar is orange, featuring the UNZA Moodle logo, user icons, and a 'My Courses' dropdown. The course title 'LIS 5310 Information Systems and Technologies in Information Management, 2019' is prominently displayed in blue. A breadcrumb navigation bar at the top right shows 'Dashboard > My courses > LIS 5310, 2019'. On the left, a sidebar lists course links: Participants, Badges, Competencies, Grades, Dashboard, Site home, and Calendar. Below the title, a section titled 'Course Instructors' lists 'Lighton Phiri <lighton.phiri@unza.zm>' and 'Office: Room 515, School of Education Building'. At the bottom right, the URL 'http://moodle.unza.zm' is shown.

Course Structure (6/6)

- **Additionally, course resources will be disseminated as follows:**
 - Large files such as videos and software tools will be made available on the instructor's profile Website

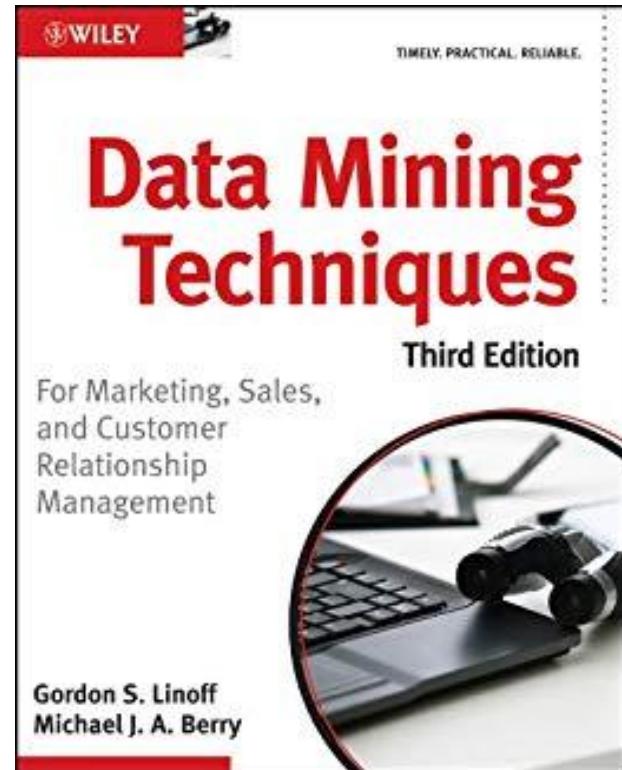
Prescribed and Recommended Textbooks (1/4)

- **Data Mining Concepts and Techniques**
 - J. Han and M. Kamber (2011)



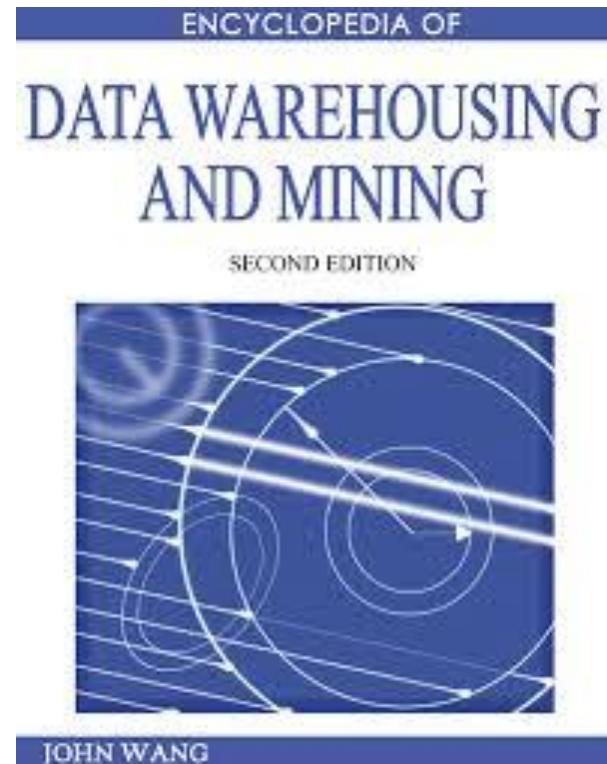
Prescribed and Recommended Textbooks (2/4)

- **Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management**
 - G. S. Linoff and M. J. Berry (2011)



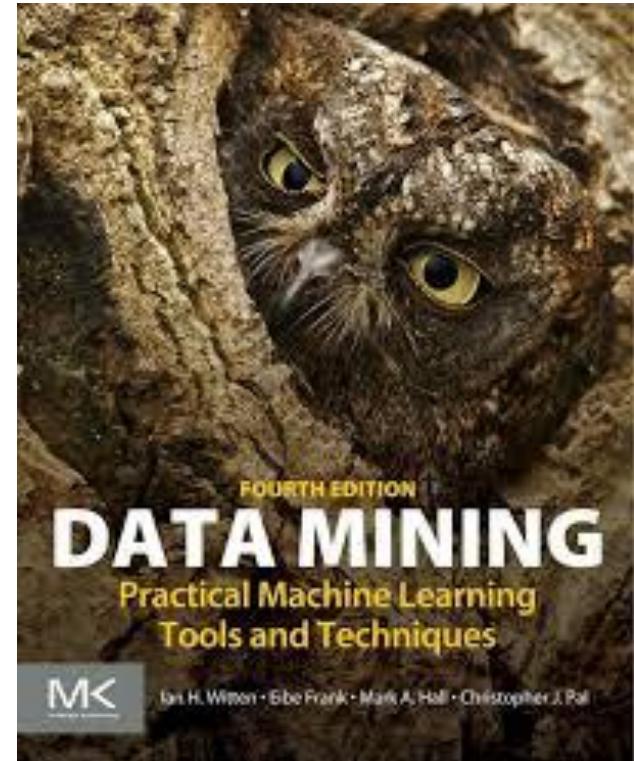
Prescribed and Recommended Textbooks (3/4)

- Encyclopedia of Data warehousing and Mining
 - J. Wang (2005)



Prescribed and Recommended Textbooks (4/4)

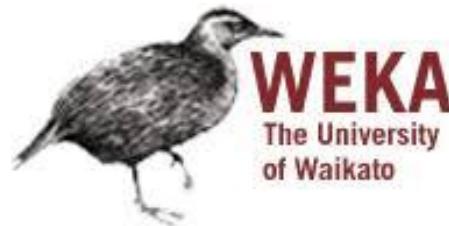
- **Data Mining: Practical Machine Learning Tools and Techniques**
 - I. H. Witten, E. Frank and M. A. Hall



Tools and Services (1/)

- **Tools and services**

- VirtualBox for creating virtual environments for running Ubuntu 18.04.
- Ubuntu 18.04 for running all practical-oriented activities and tasks.
- Python for scripting
- TensorFlow
- Weka



Course Grading (1/4)

- **10% Paper readings**
 - Paper summaries of peer-reviewed publications.
- **5% Seminar presentations**
 - Questions and discussions during seminars and reading sessions.
Marks awarded for participation.
- **5% Class participation**
 - Discussion in class.
- **20% Practical Projects**
 - Handson practical project assignments that will involve a project deliverable.

Course Grading (2/4)

- **20% Class Theory Tests**
 - One 90 minutes-long class test will be held towards the end of Term #1
- **40% Final Examination**
 - The final examination is based on the entire course outline.

Course Grading (3/4)

- Final grading is based on a 60/40 split
 - You MUST pass both the continuous assessment and examination.

1	Student ID	Paper #1 [1%] Paper #2 [1%] Paper #3 [1%] Paper #4 [1%] Paper #5 [1%] [...]	Paper #10 [1%] Paper Total [10%]	Seminar #1 [1%] Seminar #2 [1%] Seminar #3 [1%] Seminar #4 [1%] Seminar #5 [1%] [...]	Seminars Total [10%] Z Z Mini Project [20%] Z Z Class Test [20%] Z Z Continuous Assessment [60%] Z Z Final Examination [40%]
2	XXXXXXXXXXXX	N N N N N N N N	N N N N	N N N N N N N N	Z Z
3	YYYYYYYYYYYY	Z Z Z Z Z Z Z Z	Z Z	Z Z Z Z Z Z Z Z	Z Z

Course Grading (3/4)

- Final grading is based on a 60/40 split
 - You MUST pass both the continuous assessment and examination.

1	Student ID	Paper #1 [1%]	Paper #2 [1%]	Paper #3 [1%]	Paper #4 [1%]	Paper #5 [1%]	[...]	Paper #10 [1%]	Paper Total [10%]	Seminar #1 [1%]	Seminar #2 [1%]	Seminar #3 [1%]	Seminar #4 [1%]	Seminar #5 [1%]	[...]	Seminars Total [10%]	Z Z Mini Project [20%]	Z Z Class Test [20%]	Z N Continuous Assessment [60%]	Z N Final Examination [40%]
2	XXXXXXXXXXXX	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Z Z				
3	YYYYYYYYYYYY	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z	Z N	Z N	Z N	Z N	Z N

Course Grading (4/4)

GRADE	DESCRIPTION	SCORE RANGE	GRADE POINT
A+	DISTINCTION	86–100	5
A	DISTINCTION	75–85	4
B+	MERITORIOUS	70–74	3.5
B	CREDIT	65–69	3
C+	CREDIT	55–64	2.37
C	PASS	50–54	1.5
D	FAIL	<49	0

Academic Dishonesty

- Every assessment submitted must be your own work. Academic dishonesty of any form is considered very seriously.
 - NOTE: Any form of academic dishonesty (plagiarism, copying, cheating etc) will result in a ZERO mark for the entire continuous assessment score.

Course Management (1/2)

- Instructor: Lighton Phiri and TBA (possible Invited Talks)
- Email: lighton.phiri@unza.zm
- Office: Room 515, Fifth Floor, School of Education Building
- Office hours: Friday 09H00–13H00
 - Alternatively, schedule an appointment via email after checking free/busy slots on my calendar (<https://goo.gl/6kHrnA>)

Course Management (2/2)

- Communication exclusively done electronically
 - The Moodle, Course Mailing List and Email

The screenshot shows a course management interface with a navigation bar at the top. The navigation bar includes a logo, a search bar labeled "Search for messages", and various icons for messaging, actions, filters, and user settings.

The main content area displays a group page for "CSC 5741: Data Mining and Warehousing". The page indicates that the group is shared privately and contains 2 of 2 topics. A message box at the top states: "In May of 2019, we'll be merging and deprecating some of our settings to make group management easier." with a link to "Learn more".

The group description for CSC 5741 mentions it is a graduate-level course in Computer Science, addressing concepts, skills, methodologies, and models of data warehousing and data mining. It lists several bullet points detailing the course content and objectives:

- * The students are introduced to appropriate techniques for designing data warehouses for various business domains and concepts for potential uses of the data warehouse and mining opportunities.
- * CSC 5741 also provides students with fundamental concepts and algorithms in knowledge discovery process such as data pre-processing, data mining and post-process evaluation.
- * The course aims to expose learners to the principles of these techniques and appreciate their strengths and applicability to solve problems in daily life.

At the bottom of the page, there are buttons for "Edit welcome message" and "Clear welcome message", and a reminder message: "Reminder: We Have Class at 17H30 T".

A blue banner at the bottom of the page contains the URL <https://groups.google.com/a/unza.zm/d/forum/csc5741>.

Q & A Session

- **Comments, concerns and complaints?**

Lecture Series Outline

- **Part I: Administrivia**
- **Part II: Course Introduction**
 - Contextualising Data Mining and Warehousing
 - CSC 5741 Themes and Topics
 - Hints on Paper Summaries
 - Reading List
 - Planned Activities
- **Part III: On Academic Activities**
- **Part IV: About Next Week**

Contextualising Data Mining and Warehousing (1/9)

- **Past CS@ UNZA Dissertations**
 - Lillian Muzyece (2019). Automatic Weather Prediction
 - Soft Mulizwa (2019). Automatic Customer Segmentation for effective Targeted Campaigns
 - Friday Chazanga (2019). Automatic Number Plate Recognition
- **Current CS@ UNZA Dissertations**
 - Simon Hawatichke Chiwamba (2019—). Machine Learning Automated Image Capture and Identification of Fall Armyworm
 - Francis Chulu (2019—). Automatic identification and early warning and monitoring web based system of fall Armyworm

Contextualising Data Mining and Warehousing (2/9)

- Automatic classification of scholarly research
 - Automatic generation of metadata
 - Automatic reclassification of digital objects

```
▼<header>
  <identifier>oai:dspace.cbu.ac.zm:123456789/6
  <datestamp>2011-08-18T08:59:44Z</datestamp>
  <setSpec>hdl_123456789_23</setSpec>
</header>
▼<metadata>
  ▼<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/orexample-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/orexample-instance">
    ▼<dc:title>
      Customer service management in the retail
    </dc:title>
    <dc:creator>Atanga, Muyenga</dc:creator>
    <dc:subject>Banks</dc:subject>
    <dc:subject>Retail banking</dc:subject>
    <dc:subject>MBA THESIS</dc:subject>
    <dc:subject>Customer service</dc:subject>
    <dc:description>v,116p.</dc:description>
    <dc:description>Copperbelt University, School of Management</dc:description>
    <dc:date>2011-07-19T14:32:14Z</dc:date>
    <dc:date>2011-07-19T14:32:14Z</dc:date>
    <dc:date>2011-07-19</dc:date>
    <dc:type>Thesis</dc:type>
```

Contextualising Data Mining and Warehousing (3/9)

- University of Zambia Ranking Committee Research Report
 - Mining for scholarly output on the Web

Staff ID	Title	URL	Publication Date		Except
007132.csv	The moral injury symptom scale-military version	http://scholar.google.com/https://E+018 None http://scholar.google.com/scholar?cites=5193522035196824996&as_sdt=2005&sciodt=0,5&sdt=0,5&as_ylo=2018&as_yhi=2019 None The purpose of this study was to develop a multi-dimensional measure in intervention studies that target moral injury (MI) in Veterans and Active Duty Military.			
007132.csv	Patient-reported outcomes in cirrhosis: A scoping review of the literature	http://sep.29756 2018 8 5 1.0605112746671039028&hl=en&as_sdt=0,5&as_ylo=2018&as_yhi=2019 None Patients with cirrhosis have significant dissatisfaction with the care they receive. However, these patient-reported outcomes (PROs) are not robust ...			
007132.csv	Executive summary of the Merck Animal Health Veterinary Wellbeing Study	http://scholar.31 2018 1 2 1.23665324378367E+019 None http://scholar.google.com/scholar?cites=1236653243783663220&hl=en&as_sdt=0,5&as_ylo=2018&as_yhi=2019 None Still, questions remain as to how widespread it is, and whether it has risen to crisis levels. To date, no definitive assessment tool for veterinarians ...			
007132.csv	Screening for Moral Injury: The Moral Injury Symptom Scale-Military Version Short Form	http://t/183/11-12/e659/4934229 2018 2 7 8.90879491899335E+018 None http://scholar.google.com/scholar.com/scholar?cluster=8908794918993354212&hl=en&as_sdt=0,5&as_ylo=2018&as_yhi=2019 None Introducing the Moral Injury Symptom Scale-Military Version (MISS-M) to use when screening for moral injury and monitoring ...			
007132.csv	A CHAF1B-dependent molecular switch in hematopoiesis and leukemia pathogenesis	http://610818304707 2018 1 5 3.61239411341771E+018 None http://scholar.google.com/scholar?cites=3612394113417710872&hl=en&as_sdt=0,5&as_ylo=2018&as_yhi=2019 None CHAF1B is the p60 subunit required for assembly of histones H3.1/H4 heterodimers at the replication fork during S phase. Here we show that ...			
007132.csv	Moral injury, religiosity, and suicide risk in US veterans and active duty military	/advance-article-abstract/doi/10.1093/milmed/usy148/5038451 2018 1 5 1.63403287215519E+019 None Moderating effect of moral injury (MI) is related to greater suicide risk among Veterans and Active Duty Military (VAD). The moderating effect of religiosity on this relationship in ...			
007132.csv	An interferon-driven oxysterol-based defense against tumor-derived extracellular vesicles	e/pii/S1535610818305555 2019 1 4 3.91655570243661E+018 None http://scholar.google.com/scholar.com/scholar?cluster=2016555702436607475&hl=en&as_sdt=0,5&as_ylo=2018&as_yhi=2019 Meng/Tumor-derived ...			

Contextualising Data Mining and Warehousing (4/9)

- **LMS Log Mining**
 - Moodle usage logs

1###0###1###Moodle at University of Zambia###moodle#####0###site##1###3##0###0###0###0##91769##2016-11-22 11:33:45
2###25##4940006##VMM 7802 Health Economics, Policy, Monitoring and Evaluation###VMM 7802##0###0###0###0###0##1479916465##1525439465##0##1##0##1536578236##2016-11-23 17:41:05
3###25##4940009##VMM 7501 Principles of Epidemiology and Biostatistics##VMD 7501##VMD 7501#####1479916418##1530255329##0##1##0##1535391769##2016-11-23 17:53:38
4###25##4940013##Socio-Anthropology##VMM7412##VMM7412#####1##topics##1##2##1479852000##1##0##1535391769##2016-11-23 17:53:39
6###25##4940001##Applied Environmental Health, Water and Sanitation##VMM 7312##VMM 7312##0###0###0##1479916423##1480061951##0##1##0##1535391769##2016-11-23 17:53:43
7###25##4940002##Applied Food Microbiology and Nutritional Toxicology##VMM 7120##VMM 7120##0###0##1479916432##1480060975##0##1##0##1535391769##2016-11-23 17:53:52
9###25##4940010##VMM 8901 Research Methodology##VMM 8901##VMM 8901#####1##topics##1##543951##0##0##0##1535391769##2016-11-23 17:57:44
11##25##4940003##Ethics in Food Safety Practice##VMM 8911##VMM 8911#####1##topics##1##49209321##0##0##1##0##1535391769##2016-11-23 17:59:57
14##25##4940005##Food Safety Managemnt##VMM 7501##VMM 7501#####1##topics##1##4##1543##0##0##1##0##1535391769##2016-11-23 18:04:53
16##25##4940012##VMM 8201 Risk Analysis and Surveillance##VMM 8201##VMM 8201#####1##top917386##1525439576##0##0##0##0##1535550221##2016-11-23 18:09:46
17##220##5120004##VMD 6800 Veterinary Public Health##VMD 6800##VMD 6801#####1##topics##0##1524427683##0##0##0##1535391769##2016-11-23 21:19:13
19##25##4940007##Health Promotion, Education and Communication##VMM8711##VMM8711#####1##79970494##1480062724##0##1##0##1535391769##2016-11-24 08:54:54
20##25##4940014##Zoonotic Diseases and Infections##VMM 7610#####1##topics##1##2##1##8900##0##0##1##0##1535391769##2016-11-24 10:41:59
\###25##4940011##Research Paper##VMM 8900##VMM 8900##<p>
</p><p>\The RESEARCH PROJECT is an important component of these programme and ded the degree of Master of Science in One Health Food Safety. The project is not only importaerves as the final test of students' capability to work independently and think critically. It he sense that researchers attempt to build on and improve upon previous work' <i>(Johnson nsp;that will result in writing a research paper that will be evaluated and graded by your super\ short duration of the time allocated for research, you will have limited time and therefore

Contextualising Data Mining and Warehousing (5/9)

- **Mwabu Tablet Usage Analysis**
 - Android app usage and interaction logs
 - Interaction patterns for learners and educators



Contextualising Data Mining and Warehousing (6/9)

- Effectiveness of FISP Programme Using 'Triple Effect' Method
 - Collaboration with two economists

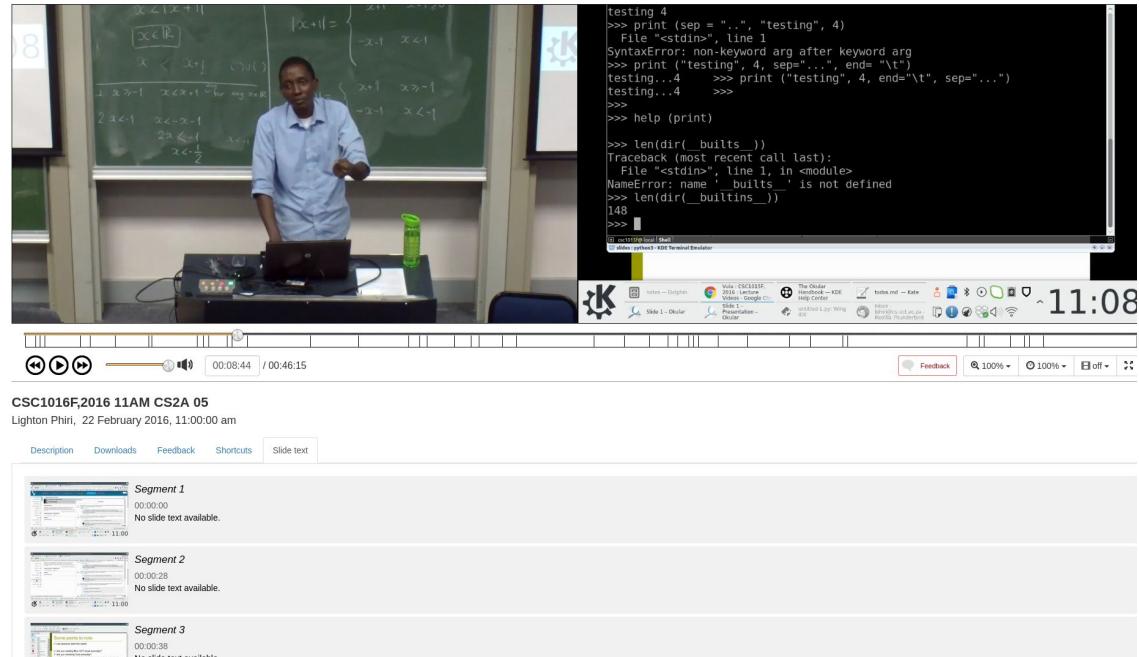
ENTER THE EXAMINATION PERIOD & EXIT

```
> colnames(dataset_cfs0405_crop)
[1] "PROV"   "DIST"   "CONST"  "WARD"   "REGION" "CSA"
[7] "SEA"    "HHNUM"  "CROP"   "ID009"  "S1AFIELD" "S1ACF01"
[13] "S1ACF02" "S1ACF03" "S1ACF04" "S1ACF05" "S1ACF06" "S1ACF07"
[19] "S1ACF08" "S1ACF09" "S1ACF10" "S1ACF11" "S1ACF12" "S1ACF13"
[25] "S1ACF14" "S1ACF15" "S1ACF16" "TOTHARV" "WEIGHT"  "HA_HARV"
[31] "convert" "HA_PLANT"
> head(dataset_cfs0405_crop)
      PROV DIST CONST WARD REGION CSA SEA HHNUM          CROP ID009
1 Central Chibombo 1 1 1 14 1 55 Maize 1
2 Central Chibombo 1 3 1 2 2 77 Maize 1
3 Central Chibombo 1 3 1 2 2 33 Other crops (specify) 2
4 Central Chibombo 2 12 1 2 3 96 Maize 3
5 Central Chibombo 2 12 1 2 3 96 Groundnuts 3
```

```
> colnames(dataset_cfs2004_2005_weight)
[1] "ID001" "ID002" "ID003" "ID004" "ID005" "ID006" "ID007" "ID009"
[9] "WEIGHT"
> head(dataset_cfs2004_2005_weight)
      ID001 ID002 ID003 ID004 ID005 ID006 ID007 ID009 WEIGHT
1 Central Chibombo 1 1 1 14 1 1 388.84409
```

Contextualising Data Mining and Warehousing (7/9)

- Open Matterhorn Video Segmentation Analysis
 - Seeking to points of interest



Contextualising Data Mining and Warehousing (8/9)

- Automatic Content Generation
 - Underrepresentation on platforms like Wikipedia
 - We have VERY few Textbooks!!!



Contextualising Data Mining and Warehousing (9/9)

- There is more out there [...]
 - Parliament TV? Video and audio analysis
 - Tollgates! Automatic detection of vehicles
 - Automatic prediction of learning outcomes
 - :
 - Sentiment analysis: Popular Zambian Facebook pages, Twitter
 - What if we harvested articles written in mainstream newspaper articles

Introduction (1/2)

- Identify the key processes of data mining, data warehousing and knowledge discovery process
- Describe the basic principles and algorithms used in practical data mining and understand their strengths and weaknesses
- Apply data mining techniques to solve problems in other disciplines in a mathematical way
- Apply data mining methodologies with information systems and generate results which can be immediately used for decision making in well-defined business problems

Introduction (2/2)

- Data Mining and Data Pre-processing
- Data Warehousing
- Classification
- Associative Rule Mining
- Clustering Analysis

Theme #1: Data Mining and Data Pre-processing

- Data Mining vs. Statistics
- Knowledge discovery process
- Machine learning
- Pattern recognition
- Date cleaning
- Data integration
- Data selection
- Data transformation
- Pattern evaluation
- Knowledge presentation

Theme #2: Data Warehousing

- Decision support system
- Data warehouse architecture
- Online transaction processing
- Online analytical processing
- Star schema, Snowflake schema
- Fact constellation
- Dimension Tables and Fact tables
- Data Granularity
- Data cube
- Pivot, slice and dice, roll-up and drill down

Theme #3: Classification

- Decision Tree; Hunt's Algorithm; C4.5; Tree Induction; Binary split and Multi-way
- split; Measures of Impurity: Gini Index, Entropy and Misclassification error; Rule-
- Based Classifier; Coverage and Accuracy; Mutually exclusive and exhaustive rules;
- Ripper; Rule Pruning; Instance-Based Classifiers; Nearest neighbour classification;
- Probabilistic classifier; Naïve Bayes classifier.

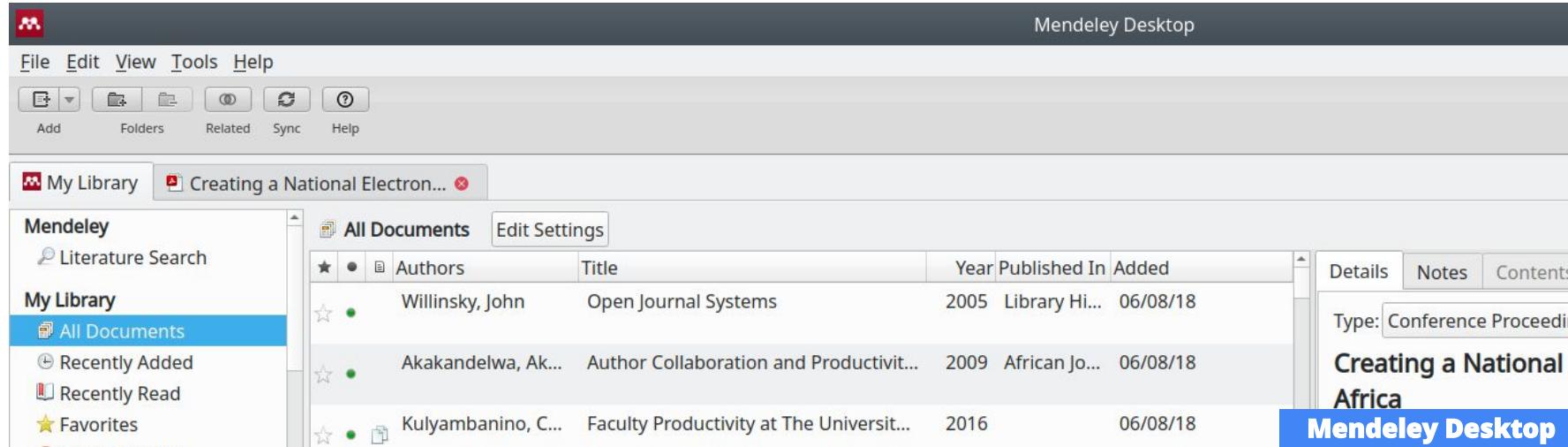
Theme #4: Associative Rule Mining

- Rule Evaluation Metrics: Support and confidence
- Frequent Itemsets, Maximal
- Frequent Itemset, Closed Frequent Itemsets
- Brute-force approach
- Apriori principle
- Frequent-Pattern Tree
- Prefix paths, Conditional FP-Tree
- Rule Generation

Theme #5: Clustering Analysis

- Intra-cluster distances, Inter-cluster distances
- Partitional clustering
- K-means
- Centroid; Sum of Squared Error
- Hierarchical clustering
- Agglomerative and divisive
- Dendrogram
- Single linkage, complete linkage and group average
- Ward's Method.

Readings and Paper Summaries (1/5)



The screenshot shows the Mendeley Desktop application window. The menu bar includes File, Edit, View, Tools, and Help. Below the menu is a toolbar with icons for Add, Folders, Related, Sync, and Help. The main interface has a sidebar on the left with options: Mendeley (Literature Search), My Library (selected), All Documents (highlighted in blue), Recently Added, Recently Read, and Favorites. The central area displays a grid of documents with columns for Authors, Title, Year, Published In, and Added. The first three documents are visible:

Authors	Title	Year	Published In	Added
Willinsky, John	Open Journal Systems	2005	Library Hi...	06/08/18
Akakandelwa, Ak...	Author Collaboration and Productivit...	2009	African Jo...	06/08/18
Kulyambanino, C...	Faculty Productivity at The Universit...	2016		06/08/18

On the right side, there are tabs for Details, Notes, and Contents. A tooltip "Creating a National Africa" is shown over the Details tab. A blue box highlights the "Mendeley Desktop" text at the bottom right.

- You **MUST** develop the habit of using a bibliographic manager
 - Mendeley, Zotero, JabRef, RefWorks

Readings and Paper Summaries (2/5)

≡ Google Scholar data mining and machine learning 

 Articles About 2,040,000 results (0.12 sec)

Any time [CITATION] Data Mining: Practical **machine learning** tools and techniques [PDF] fue.edu.eg
Since 2019 I H Witten, E Frank, M A Hall, C J Pal - 2016 - Morgan Kaufmann
Since 2018 ☆ 99 Cited by 34827 Related articles All 38 versions

Since 2015 Distributed GraphLab: a framework for **machine learning** and **data mining** in the cloud [PDF] arxiv.org
Custom range... Y Low, D Bickson, J Gonzalez, C Guestrin... - Proceedings of the ..., 2012 - dl.acm.org
[] - []
Search
Any time While high-level **data** parallel frameworks, like MapReduce, simplify the design and implementation of large-scale **data** processing systems, they do not naturally or efficiently support many important **data mining** and **machine learning** algorithms and can lead to ...
Since 2019 ☆ 99 Cited by 1554 Related articles All 29 versions

Sort by relevance Business **data mining**—a **machine learning** perspective
Sort by date I Bose, R K Mahapatra - Information & management, 2001 - Elsevier
[] include patents The objective of this paper is to inform the information systems (IS) manager and business analyst about the role of **machine learning** techniques in business **data mining**. **Data mining** is a fast growing application area in business. **Machine learning** techniques are used for ...
[] include citations
[] Create alert ☆ 99 Cited by 367 Related articles All 5 versions

The WEKA **data mining** software: an update [PDF] researchgate.net
M Hall, E Frank, G Holmes, B Pfahringer... - ACM SIGKDD ..., 2009 - dl.acm.org
Machine learning WEKA is a collection of machine learning algorithms for data mining and machine learning tasks.

<https://scholar.google.com>

Readings and Paper Summaries (3/5)



The screenshot shows the AMiner search interface. At the top left are the KEST logo and the AMiner logo. To the right is a search bar with the placeholder "Whatever comes to your mind". Below the search bar are two blue buttons: a magnifying glass icon for search and a refresh/circular arrow icon. Further right are icons for mobile, English language selection, a user profile, and a navigation arrow.

Computer Science	Rank	Conference (Full Name)	Short Name	H5-Index
All	1	International World Wide Web Conferences	WWW	66.00
High Performance Computing	2	Information Sciences	Inf. Sci.	62.00
Computer Network	3	ACM Knowledge Discovery and Data Mining	KDD	56.00
Network and Information Security	4	IEEE Transactions on Knowledge and Data Engineering	TKDE	53.00
Software Engineering	5	ACM International Conference on Web Search and Data Mining	WSDM	50.00
Database and Data Mining	6	International Conference on Research an Development in Information Retrieval	SIGIR	47.00
Theoretical Computer Science	7	Journal of the American Society for Information Science and Technology	JASIST	42.00
Computer Graphics and	8	IEEE International Conference on Data Engineering	ICDE	40.00
	9	ACM International Conference on Information and Knowledge Management	CIKM	38.00
	10	IEEE International Conference on DataMining	ICDM	33.00
	11	Journal of Web Semantics	J. Web Sem.	33.00
	12	Knowledge and Information Systems	KAIS	31.00
	13	International Journal of Geographical Information Science	IJGIS	

<https://aminer.org>

Readings and Paper Summaries (4/5)



Zambia ICT Journal

Announcements

Current

Archives

About ▾

The **Zambia ICT Journal (ISSN: 2616-2156)** is published four times a year by the ICT Association of Zambia (ICTAZ) with technical support from the University of Zambia, Copperbelt University and Mulungushi University. The objective of Journal is to support and stimulate active productive research which could strengthen the technical foundations of engineers and scientists in the African continent, develop strong technical foundations and skills and lead to new small to medium enterprises within the African sub-continent. We also seek to encourage the emergence of functionally skilled technocrats within the continent on publishing research results and studies in Computer Science and Information Technology through a scholarly publication. The Zambia ICT journal is double blind peer reviewed.

Announcements

Call for paper for Volume 3 Issue 2 (June 2019)

□ 2019-03-08

The Zambia ICT Journal wishes to call for original research papers containing new research findings which have not been submitted concurrently to any other publication to be published in **Volume 3 Issue 2 (June 2019)** in any of the

[Make a Submission](#)

[View Pending Submissions](#)

[Author Guidelines](#)

[Article Template](#)

[Publications Fees](#)

[Editorial Team](#)

[Special Issue](#)

<http://ictjournal.icict.org.zm>

CSC 5741 L01 - 46

Readings and Paper Summaries (5/5)



University of Cape Town

[My Author Page](#) [My Binders](#) [SIGN OUT: Lighton Phiri](#)

How to read a paper

Full Text: [PDF](#)

Author: [S. Keshav](#) [University of Waterloo](#)

Published in:



- Newsletter
- ACM SIGCOMM Computer Communication Review archive



[2007 Article](#)



[Bibliometrics](#)

Reading a computer science research paper

Full Text: [PDF](#)

Author: [Philip W.L. Fong](#) [University of Calgary, Calgary, Alberta, Canada](#)

Published in:



- Newsletter
- ACM SIGCSE Bulletin [archive](#)
- Volume 41 Issue 2, June 2009
- Pages 138-140



[2009 Article](#)



[Bibliometrics](#)

- Citation Count: 4
- Downloads (cumulative): 2,025

Tools and Resources

[Buy this Article
\(PRINT\)](#)

[Recommend the ACM DL
to your organization](#)

[TOC Service:](#)
 [Email](#) [RSS](#)

Tools and Resources

[Buy this Article
\(PRINT\)](#)

[Recommend the ACM DL
to your organization](#)

[TOC Service:](#)
 [Email](#) [RSS](#)

<https://dl.acm.org>

Closing CSC 5710 Remarks

- **Beyond CSC 5741**
 - Research focus
 - Vision 2030
- **About assessments**
 - Ensure all assessments are attempted
- **Academic dishonesty**
 - NOTE: Any form of academic dishonesty (plagiarism, copying, cheating etc) will result in a ZERO mark for the entire continuous assessment score.

Q & A Session

- **Comments, concerns and complaints?**

Lecture Series Outline

- Part I: Administrivia
- Part II: Course Introduction
- Part III: On Academic Activities
 - Public Talks
 - Public Oral Examinations
 - DRGS Organised Events
- Part IV: About Next Week

Public Talks

- **Make time to attend public academic talks irrespective of whether it is computing related**
 - Inspiration for potential topics next year
 - Potential collaboration

2nd Seminar in the Colloquium Series for 2019 [Inbox ×](#)  

Public Relations UNZA <public@unza.zm>
to unza ▾

Wed, Feb 27, 3:30 PM   

The Department of Media and Communication Studies will hold the 2nd seminar in the Colloquium Series for 2019 on **Friday, 1st March 2019 in the Senate Chamber at 15:00 hours.**

Please see attachment for details.

You are all welcome to attend.

Regards.

Damaseke Chibale
Manager, Public Relations



March 15 2019

Public Oral Examinations

- Make time to attend public oral examinations so you have an idea what to expect.



SCHOOL OF AGRICULTURAL SCIENCES SEMINAR SERIES

PhD Public Defence

*“Assessment of the impact of climate change on maize (*Zea mays L.*) yield using crop simulation and statistical downscaling models in a subtropical environment of Zambia”*

By: Charles Bwalya Chisanga
(PhD Candidate – Integrated Soil Fertility Management)

All students to attend

DATE: Thursday, 7th March, 2019

TIME: 12:00-13:00 hrs.

VENUE: VET LT

DRGS Organised Events

- You want to attend important postgraduate events in order to gain a sense of what is expected
 - Announcements are sent through to your official UNZA-assigned email addresses.

Monday 13 th May, 2019 to Friday 26 th July, 2019	IDE Students School Experience (11 Weeks)
Monday 20 th May, 2019 to Thursday 24 th May, 2019	Graduation Week (Second Graduation Ceremony)
Monday 3 rd June, 2019 to Friday 7 th June, 2019	Study Break and Post Graduate Seminar Week
Wednesday 2 nd October, 2019	Senate Curriculum and Examinations Committee Meeting (Consideration IDE Results)
Monday 14 th October, 2019 to Friday 18 th October, 2019	Study Break and Post Graduate Seminar Week
Monday 21 st October, 2019 to Friday 15 th November, 2019	Final Examinations (15 Days)
Saturday 16 th November, 2019	Vacation for Regular Students Starts
Monday 24 th November, 2019 to Friday 29 th November, 2019	Deferred examination (5 Days)
Friday 29 th November, 2019	Senate Examination and Irregularities Committee

Lecture Series Outline

- **Part I: Administrivia**
- **Part II: Course Introduction**
- **Part III: On Academic Activities**
- **Part IV: About Next Week**
 - Data Mining and Data Processing
 - Getting Started: Python, SciKit-learn, Pandas
 - Mini Projects Overview
 - Paper Reading List [Homework]
 - Academic Talk: L. Phiri

Theme #1: Data Mining and Data Processing

- Data Mining vs. Statistics
- Knowledge discovery process
- Machine learning
- Pattern recognition
- Date cleaning
- Data integration
- Data selection
- Data transformation
- Pattern evaluation
- Knowledge presentation

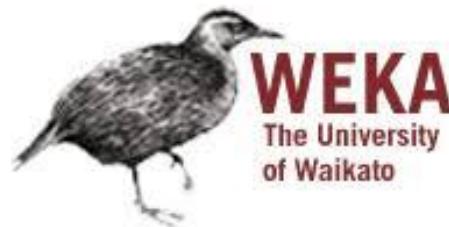
Getting Started with Python, SciKit-learn & Pandas

- Tools installation
- Common commands
- SciKit-learn
- Pandas
- Sample datasets



Mini Project Overview

- Ideally, everyone will work on an individual project, however, the themes will be similar
 - Mini Project descriptions
 - Grade distribution
 - Open date and deadline
- Deliverables
 - Properly curated datasets
 - Well-organised scripting code (Github/Bitbucket)
 - Technical report



Paper Reading List [Homework]

- [1] **S. Keshav (2007) "How to Read a Research Paper"**
<https://doi.org/10.1145/1273445.1273458>
- [2] **P. W. L. Fong (2004) "How to Read a CS Research Paper?"**
<https://doi.org/10.1145/1595453.1595493>
- [3] **L. Phiri (2018) "Research Visibility in the Global South: Towards Increased Online Visibility of Scholarly Research Output in Zambia"**
http://lis.unza.zm/~lightonphiri/papers/paper-icict18-online_visibility.pdf

Academic Talk: L. Phiri

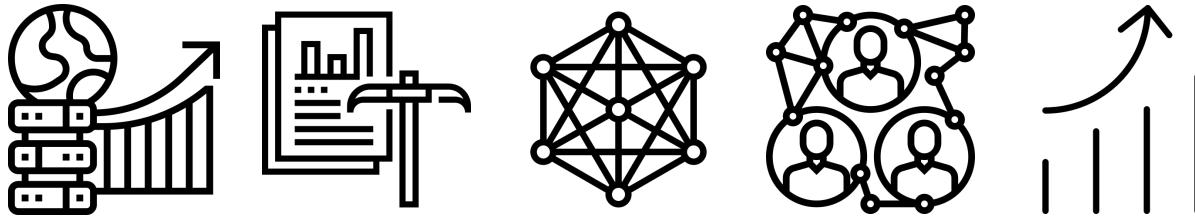
- [1] **L. Phiri (2018) “Research Visibility in the Global South: Towards Increased Online Visibility of Scholarly Research Output in Zambia”**

http://lis.unza.zm/~lightonphiri/papers/paper-icict18-online_visibility.pdf

Bibliography

[1] CSC 5741 Syllabus

<http://lis.unza.zm/~lightonphiri/teaching/unza/2019/csc5741>



CSC 5741

Lecture 1: Administrivia and Course Overview

Lighton Phiri <lighton.phiri@unza.zm>
Department of Library and Information Science
University of Zambia