Use Cases in Big Data Software and Analytics

Vol. 1, Fall 2017

Bloomington, Indiana

Editor:
Gregor von Laszewski
Department of Intelligent Systems
Engeneering
Indiana University
`laszewski@gmail.com`

Contents

Chapter 1

Preface

1.1 List of Papers

Name	HID	Title
hid101	Huiyi Chen	Big Data and standardize testing
hid102	Dianprakasa, Arif	This is my paper about xyz
hid104	Jones, Gabriel	What Separates Big Data from Lots of Data?
hid105	Lipe-Melton, Josh	This is my paper about data visualization in sports
hid106	Qiaoyi Liu	Big Data Analytics in Groceries Stores
hid107	Ni,Juan	This is my paper about xyz
hid109	Shiqi Shen	Big Data in Social Media
hid110	Hsiao Yuan Wang	This is my paper about xyz
hid111	Lewis, Derek	Big Data Analytics in Biometric Identity Management
hid201	Arnav, Arnav	Big Data analytics and Edge Compting
hid202	Himani Bhatt	Big data analytics in Weather forecasting
hid203	Chandwani, Nisha	Big Data Analytics using Spark
hid204	Chaturvedi, Dhawal	Big Data Anaytics and High Performance Computing
hid205	Chaudhary, Mrunal L	Applications of Big Data in Fraud Detection in Insurance
hid208	Devineni, Jyothi Pranavi	This is my paper about Big Data and Deep Learning
hid209	Han, Wenxuan	Big Data Application in Web Search and Text Mining
hid210	Hotz, Nicholas	Natual Language Processing of Electronic Health Records
hid211	Ajinkya Khamkar	Distributed environment for neural network
hid212	Kumar, Saurabh	Big Data Analysis using MapReduce
hid213	Liu, Yuchen	Big Data and Speech Recognition
hid214	Lu, Junjie	Big Data and Basketball
hid215	Mallala, Bharat	Big Data and Artificial Neural Networks
hid216	Millard, Mathew	Big Data Analytics in Sports - Track and Field
hid218	Niu, Geng	Big data's influence on e-commerce and lifestyle
hid219	Parampali Sreenath, Syam Sundar Herle	Big Data Analytics Architecture for Real-Time Traffic Control
hid224	Rawat, Neha	Big Data Applications in the Hospitality Sector
hid225	Schwartzer, Matthew	Optimizing Mass Transit Bus Routes with Big Data
hid228	Swargam, Prashanth	Big data applications in Electric Power Distribution
hid229	ZhiCheng Zhu	Big Data and Machine Learning
hid230	YuanMing Huang	Big data with natural language processing
hid231	Vegi, Karthik	Using Big Data for Fact Checking
hid232	Rahul Velayutham	Big Data Analytics in Sports - Soccer

hid233	Wang, Jiaan	Big Data Applications in Media and Entertainment Industry
hid234	Weixuan Wang	Big Data Analytics in Tourism Industry
hid235	Wu, Yujie	Big Data in Recommendation System
hid236	Yang Weipeng	Big Data in MOOC
hid237	Ahmed, Tousif	Big Data Analytics in Cyber Security and Threat Research
hid238	Jeff LaDow	This is my paper about xyz
hid301	Arora, Gagan	Big Data Analytics in Finance Industry
hid302	Sushant Athaley	Big Data Application in Restaurant Industry
hid303	Brunetti Nademlynsky, Lisa	This is my paper about xyz
hid304	Ricky Carmickle	Big Data and Astrophysics
hid305	Andres Castro Benavides	Big Data Analytics for Municipal Waste Management.
hid306	Cheruvu, Murali	Internet of Things Alliance with Big Data
hid308	Pravin Deshmukh	Big Data and Data Visualization
hid309	Dubey, Lokesh	BigData Analytics using Apache Spark in Social Media
hid310	Kevin Duffy	Big Data Applications in Food Insecurity
hid311	Durbin, Matthew	This is my paper about xyz
hid312	Neil Eliason	An Overview of Big Data Applications in Mental Health Treatment
hid313	Tiffany Fabianac	Big Data Platforms as a Service
hid314	Fadnavis, Sarang	Big Data analytics in Media industry
hid315	Garner, Jeffry	This is my paper about xyz
hid316	Robert Gasiewicz	Big Data Analytics in Biometric Identity Management
hid318	Irey, Ryan	This is my paper about xyz
hid319	Mani Kumar Kagita	Big Data Analytics for Municipal Waste Management
hid320	Elena Kirzhner	Big Data Analytics and Applications in Childbirth
hid321	Knapp, William	This is my paper about xyz
hid322	Koslik, Kent	This is my paper about xyz
hid323	Uma M Kugan	This is my paper about NoSQL Databases in support of Big Data Applications and Analytics
hid324	Ashok Kuppuraj	Big data in Blockchain
hid325	J. Robert Langlois	Impact of Big Data on the Privacy of individual with Mental Illeness
hid326	Mahendrakar, Mohan	Bigdadta in Clinical Trails
hid327	Marks, Paul	Waste in Healthcare
hid328	Dhanya Mathew	Big data analysis in Finance Sector
hid329	Ashley Miller	Big Data Analytics in Higher Education Marketing
hid330	Janaki Mudvari Khatiwada	Big data in Improving Patient Care
hid331	Tyler Peterson	Big Data Applications In Population Health Management
hid332	Judy Phillips	Big Data Analytics in Agriculture
hid333	Anil Ravi	Big Data and Artificial Intelligence solutions for In Home, Community and Territory Security
hid334	Peter Russell	AWS in support of Big Data Applications and Analytics
hid335	Sean Shiverick	Big Data Analytics, Data Mining, and Public Health Informatics: Using Data Mining of Social Media to Track Epidemics
hid336	Jordan Simmons	Recommendation Systems on the Web
hid337	Ashok Reddy Singam	Big Data and Artificial Intelligence Solutions for in Home, Community and Territory Security
hid338	Sriramulu, Anand	Docker in support of Big Data Applications and Analytics
hid339	Hady Sylla	Big data application for treatment of breast cancer
hid340	Tim Thompson	Big data analytics for archives and research libraries
hid341	Tibenkana, Jacob	This is my paper about xyz

hid342	Udoyen, Nsikan	Big data analytics in college football (NCAA)
hid343	Usifo, Borga	Big Data Applications in Self-Driving Cars (Approval Waiting)
hid344	Watts, Bradley	This is my paper about xyz
hid345	Wood, Ross	Big data analytics in the entertaiment industry.
hid346	Zachary Meier	Big Data in Oceanography
hid347	Jeramy Townsley	Sociological Applications of Big Data
hid348	Budhaditya Roy	Using Singularity for Big Data

My great Big Dat Paper

Qiaoyi Liu
Indiana University Bloomington
3209 E 10th St
Bloomington, Indiana 47401
ql30@umail.iu.edu

ABSTRACT

KEYWORDS

ACM proceedings, LATEX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data's influence on ecommerce and lifestyle

Geng Niu
Indiana University Bloomington
752 Woodbridge Dr
Bloomington, Indiana 47408
gengniu@iu.edu

ABSTRACT

This paper studies how big data is applied in ecommerce and its influence on lifestyle.

KEYWORDS

big data, ecommerce

1 INTRODUCTION

This is my introduction

1.1 Citations

Citations to articles [**?**]

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

REFERENCES

Big Data Applications in the Hospitality Sector

Neha Rawat
Indiana University
Bloomington, Indiana
nrawat@iu.edu

ABSTRACT

This paper focuses on how big data is used in the hotel industry for better customer satisfaction, marketing effectiveness and yield management using customer data for segmentation and predictive analyses.

1 CONCLUSIONS

This is the conclusion.[1]

ACKNOWLEDGMENTS

Acknowledgements

REFERENCES

[1] Gregor V Laszewski. 2017. test. (2017).

Big Data Analytics in Tourism Industry

Weixuan Wang
Indiana University Bloomington
Bloomington, Indiana 47405
wangweix@indiana.edu

ABSTRACT

This paper focuses on how the tourism industry has been impacted by the development of the Internet and improvements in information and communication technologies and how big data analytic can influence tourism research.

KEYWORDS

Big data analytics, tourism

1 INTRODUCTION

this is my introduction [1].

2 CONCLUSIONS

This my conclusion.

ACKNOWLEDGMENTS

The authors would like to thank I523.

REFERENCES

[1] G. Chareyron, J. Da-Rugna, and T. Raimbault. 2014. Big data: A new challenge for tourism. In *2014 IEEE International Conference on Big Data (Big Data)*. 5–7. https://doi.org/10.1109/BigData.2014.7004475

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Lars Thørväld
The Thørväld Group
1 Thørväld Circle
Hekla, Iceland
larst@affiliation.org

Valerie Béranger
Inria Paris-Rocquencourt
Rocquencourt, France

Aparna Patel
Rajiv Gandhi University
Rono-Hills
Doimukh, Arunachal Pradesh, India

Huifen Chan
Tsinghua University
30 Shuangqing Rd
Haidian Qu, Beijing Shi, China

Charles Palmer
Palmer Research Laboratories
8600 Datapoint Drive
San Antonio, Texas 78229
cpalmer@prl.com

John Smith
The Thørväld Group
jsmith@affiliation.org

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LATEX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data Application in Restaurant Industry

Sushant Athaley
Indiana University
sathaley@iu.edu

ABSTRACT

This paper provides insight into how big data can be used in the restaurant industry. It also explores how big data can be collected and analyzed so that it helps restaurant industry to do better in profit margins and give their customer a great hospitality experience. This paper will try to find out current technologies and solutions available in big data processing for the restaurant industry. It will also focus on various challenges involved in using big data in the restaurant business. This paper is a review/research paper which considers information from various sources like articles, books and web to provide the information.

KEYWORDS

i523, hid302, big data, restaurant, application, analytics

1 INTRODUCTION

Big data is revolutionizing the way business is getting conducted in various industries. The retailer like Amazon uses it to provide personalized buying suggestions and social networking site like LinkedIn uses it to connect more people. Question is, do we have big data available for the restaurant industry and how big data application is going to be beneficial. The restaurant industry is facing challenges like shrinking labor pool, moderate economic growth, costly labor, challenging profit margin, high competition, moderate sales growth and growing expectation from the customer on the dining experience, can big data application help overcome these challenges.[1]

2 BIG DATA FOR RESTAURANT/INGREDIENT

To understand how big data analytics will help, we first need to find out what are the data points present in the restaurant industry which can be considered as big data. As one of the V-variety of big data, the restaurant also has structured and unstructured data. Structured data is something which is getting generated inside the restaurant and unstructured data is something which is outside of the restaurant.

2.1 Structured Data

Structured data is well formatted, easy to understand and analyze. Restaurant POS(point of sale) system shows whatfis selling, where, and at what time[6]. Food and beverage cost, labor cost, product mix, rent cost are obvious data points. Raw material required for preparation, menu, ingredient consideration, meal preparation, product availability from the supplier, prices of products comes from the kitchen of the restaurant. Staffing schedule, table turnover, bar management, wages, salaries, tips, customer feedback is data. The number of time employee coming late, number of times drinks provided as comp due to server error is data.[2]

2.2 Unstructured Data

Unstructured data is un-formatted, difficult to gather and analyze. Data shared from social media like trends, retweets, shares, and comments categorize as unstructured data. Customer promotions, customer profile like age, gender, address, email, taste preference, favorite dish, various milestones like birthdate, anniversary etc, along with family information is also an unstructured data. Weather and traffic information also constitutes as an important data to consider. [2]

3 COLLECT BIG DATA/CONSUME

These various data attributes can be collected from the different system. Most of the data is generated inside the restaurant by the system like POS which captures all sales transactions. POS system can also break down sales by time, size of the party, menu items, and ingredients. The inventory provides information on suppliers, food, beverages, and gas and electricity bill. Payroll provides information on wages, salaries, employee schedule, and time off by the employees. Loyalty program and marketing promotions provides data regarding marketing of the restaurant.

Outside data can be gathered through the various applications like OpenTable, Facebook, Twitter, Yelp, TripAdvisor, Foursquare, Urbanspoon or Instagram, weather and traffic sites. Information can be gathered from customer like his favorite menu/drink item, favorite table, special request, allergies, liking to the presentation, feedback on ambiance, service and food. [2]

4 BIG DATA ANALYTICS/RECIPE FOR SUCCESS

Benjamin Stanley, co-founder of Food Genius, suggests ”A restaurant operator shouldn't just jump into big data unless they have a problem they are trying to solve”[4]. Big data analytics can help with various analysis which can solve different issues but itfis important to know the problem which needs to be solved. Menu analysis can help with deciding the cost of the item, popular menu item, how often items are ordered, the time when menu item ordered, ingredient used and if any ingredient needs to be substituted[4]. Labor cost can be managed better by analyzing overtime pay, absenteeism, costs to sales, costs by department and server, tips, amount of time spent at the table, types of entres sold and whether the server sells the special. This analysis can be used to motivate, train and provide incentives to the servers[2],[4]. Guest check analytics can help determine what sells well, how often somebody orders certain items and detailed pricing analyses[2]. Customer profile analysis gives insight on demographics of the customer, ages, income level, their family information, kind of food they like, allergies, drink habits, places they dine out, special occasions and this analysis can be used to provide the personalized experience to the customer[2]. Servers can use customer profile analysis to suggest menu choices, celebrate birthdays or special occasions, or run specials to drive more business. Reservation system data analysis helps in understanding who all are coming, when they last visited, what they tend to order, are they celebrating any special occasion and accordingly then can decide on the menu[7]. Data mining of data from social media like Facebook, Twitter, Instagram, YouTube can help in understanding sentiments of the customer, social news, training topic, views on self and competitor restaurants, identify brand or restaurant fans[3]. This mining also provides the capability to get feedback real time and respond at the same time. This information can be used to do targeted marketing for the specific audience[3].

5 SOLUTION AND TOOLS AVAILABLE

6 REAL LIFE EXAMPLES/FLAVORFUL IMPLEMENTATION

A quickservice chain monitors its drive-thru lanes to determine which items to display on its digital menu board. When lines are longer, the menu features items that can be prepared quickly. When lines are shorter, the menu features higher-margin items that take a bit longer to prepare. Those subtle changes in the menu board wouldnfit be possible if the company couldnfit tap into a steady stream of data in real time to make instantaneous adjustments.[2]

Haute Dogs and Fries, a two-unit, quickservice restaurant in Alexandria, Va., leverages social media to connect with customers. Being small and community-focused allows the operation to quickly identify market trends and make offers in real-time, says co-owner Lionel Holmes. He monitors social media throughout the day and might post a lunch special at 11 a.m. or a dinner offer at 3 p.m. based on what is trending. Haute Dogs and Fries is on Twitter, Facebook and Instagram and uses email to reach customers and build loyalty.[2]

Fig and Olive, a seven-location New York-based restaurant group, has used guest-management software to track more than 500,000 guests and $17.5 million in checks. The restaurants have been able to customize the dining experience for individual guests and deliver results with targeted email communications. It’s “we miss you campaig“ offered complimentary crostini to guests who hadn’t dined there in 30 days. The result: Almost 300 visits and more than $36,000 in sales, translating into a return of more than seven times the cost of the program. Matthew Joseph, who leads technology and information systems for the company, says linking POS data with online reservations, plus monitoring social media mentions on Facebook, Twitter or TripAdvisor, helped Fig and Olive create its brand identity and build loyalty.[2]

Dickeys Barbecue Pit, which operates 514 restaurants across the U.S., uses Smoke Stack system to provide near real-time feedback on sales and other key performance indicators. All of the data is examined every 20 minutes to enable immediate decisions if the sale is not at certain baseline at a certain store in the region then it enables them to deploy training or operation directly to that store. For example, if there is lower than expected sales one lunchtime, and have an amount of ribs there, then text invitation is sent to people in the local area for ribs special fi?! to both equalize the inventory and catch up on sales.[5]

7 CHALLENGES OF USING BIG DATA

8 CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

A HEADINGS IN APPENDICES

A.1 Introduction

A.2 Big Data for Restaurant/Ingredient

A.2.1 Structured Data.

A.2.2 Unstructured Data.

A.3 Collect Big Data/Consume

A.4 Big Data Analytics/Recipe for Success

A.5 Solution and Tools Available

A.6 Real Life Examples/Flavorful Implementation

A.7 Challenges of Using Big Data

A.8 Conclusions

A.9 References

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

B MORE HELP FOR THE HARDY

Of course, reading the source code is always useful. The file `acmart.pdf` contains both the user guide and the commented code.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts’ Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

[1] 2016. Restaurant industry to navigate continued challenges in 2016. (02 2016). http://www.restaurant.org/News-Research/News/Restaurant-industry-to-navigate-continued-challeng

[2] National Resturant Association. 2014. Big Data and Restaurants: Something to Chew On. Web. (11 2014). https://www.restaurant.org/Downloads/PDFs/BigData

[3] LISA JENNINGS. 2015. Making big data small. *Nation’s Restaurant News* 49, 7 (May 2015), 22–23.

[4] Amanda C. Kooser. 2013. BIG DATA. *Restaurant Business* 112, 9 (September 2013), 24–31.

[5] Bernard Marr. 2015. Big Data At Dickey’s Barbecue Pit: How Analytics Drives Restaurant Performance. (Jun

2015). https://www.forbes.com/sites/bernardmarr/2015/06/02/big-data-at-dickeys-barbecue-pit-how-analytics-drives-restaurant-performance/ Forbes Article.

[6] John Morell. 2013. Get a Grip on Big Data. (may 2013). https://www.qsrmagazine.com/operations/get-grip-big-data

[7] Nicole Torres. 2016. How restaurants know what you want to eat before you do. FOOD and DRINK INC. — MAGAZINE. (May 2016). https://www.bostonglobe.com/magazine/2016/05/26/how-restaurants-know-what-you-want-eat-before-you/hnZHM3xCkL1BhX0PKL3tmM/story.html

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big data analysis in Finance Sector

Dhanya Mathew
Indiana University
711 N Park Ave
Bloomington, Indiana 47408
dhmathew@iu.edu

ABSTRACT

In order to understand what drives customer profit, we want to be able to predict what profit group (extremely unprofitable, average, extremely profitable etc.) a set of customers falls into based on their data at any given time.

KEYWORDS

Random Forest, R, standard deviation

1 INTRODUCTION

Big data as it's name implies, refers to large and complex data which continues to grow enormously day by day. There are huge number of sectors or applications including government, business, technology, universities, health-care, finance, manufacturing etc who make use of big data by obtaining meaningful information using big data technologies. This paper investigates how big data is helpful in financial firms in terms of predictive analysis and profitable growth. The finance sector is generating huge amounts of data on a daily basis from products and marketing, banking, business, to share market. Finance is a very sensitive field and any useful insight can make a positive impact on the overall turnover. Historic data analysis and real time data analysis are equally important in terms of finance sector. The key idea behind is how to retrieve the ”signal” of relevant information form the bulk of data. Let us explore the wide range of possibilities of big data analysis that finance sector can come up with including decision making, discovery of new business opportunities, enhanced productivity and efficiency, risk management, fraud detection, innovation possibilities, efficiency and growth and a detailed view of customer segmentation in banking sector.

1.1 Efficient decision making

1.2 Increased productivity and growth

1.3 Identify business priorities

1.4 Risk Management

1.5 Understand new business opportunities

1.6 Discovery of innovation possibilities

1.7 Fraud detection

1.8 Cost effective information gathering

1.9 Customer Segmentation and personalized marketing

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

Big Data Analytics and Edge Computing

Arnav Arnav
Indiana University, Bloomington
Bloomington, Indiana, USA
aarnav@iu.edu

ABSTRACT

With the exponential increase in the number of connected IoT devices, the data generated by these devices has grown enormously. Sending this data to a centralized server or cloud results in enormous network traffic and may lead to failures and increased latency. One solution of this problem is to do some processing on the edge devices. This is extremely helpful in providing responsive and real time analytics.

1 INTRODUCTION

With the rapid increase in the acceptanceof Internet of Things (IoT) devices across various fields in the world, ranging from industrial sensors to lifestyle and sports products, and the consequent increase in the data generated by such devices, there is a pressing demand for devices and processes that can analyze this data and provide responsive analytics.[1]. With increase in the number of sucn devices, it gets increasingly dificult to perform all analytics on a server in a traditional manner. Thus, more recent approaches aim to push a part of this computation closer to the end user of the device, or closer to the edge.

REFERENCES

[1] Yogesh Simmhan. 2017. IoT Analytics Across Edge and Cloud Platforms. IEEE IOT Newsletter. (May 2017). https://iot.ieee.org/newsletter/may-2017/iot-analytics-across-edge-and-cloud-platforms.html

Internet of Things (IoT) alliance with Big Data

Murali Cheruvu
Indiana University
3209 E 10th St
Bloomington, Indiana 47408
mcheruvu@iu.edu

ABSTRACT

This paper provides an introduction to Internet of Things (IoT) and how Big Data can effectively improve IoT process.

KEYWORDS

i523, hid306, Internet of Things, IoT, Big Data, Sensors, Actuators, Analytics, Data Science

1 INTRODUCTION

The Internet of things (*IoT*) is the network of physical devices, vehicles, and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to collect and exchange data[5]. Devices of all types - cars, manufacturing equipment, medical devices and more - have become smarter, opening up the need for their connectivity with the internet. Today, over 50% of IoT activity is centered in manufacturing, transportation, smart city, consumer applications like home automation and wearable gadgets, but within five years all industries will have rolled out IoT initiatives. *Gartner, Inc.* forecasts that 8.4 billion connected things will be in use worldwide in 2017 and will reach 20.4 billion by 2020[4].

2 IOT INTUITION

The rise of IoT changes everything by enabling *smart* things. Products and environments are becoming smarter. Broadly speaking, two kinds of IoT are emerging: *Consumer IoT* and *Industrial IoT*. Products such as Apple Watch, Fitbit and Home Automation - TV, thermostats, alarm system, etc. are considered Consumer IoT. Industrial IoT are: manufacturing equipment and medical devices.

A few more examples of IoT include:

Track your activity levels - Using your smart-phone's range of sensors (accelerometer, gyro, video, proximity, compass, GPS, etc) and connectivity options (Cell, Wi-Fi, Bluetooth, etc.) you have a well equipped IoT device in your pocket that can automatically monitor your movements, location, and workouts throughout the day.

Get most out of your medication - The Proteus ingestible pill sensor is powered by contact with your stomach fluid and communicates a signal that determines the timing of when you took your medication and the identity of the pill. This information is transferred to a patch worn on the skin to be logged for you and your doctor's reference. Heart rate, body position and activity can also be detected.

Rolls-Royce is using Azure Stream Analytics and Power BI to link up sensor data from its engines with more contextual information like air traffic control, route data, weather and fuel usage to get a fuller picture of the health of its aircraft engines.

Smart Homes - A smart home is one in which devices have the capability to communicate with each other, as well as with their environment and the Internet. Smart homes enable owners to customize and control their home environments for increased security and efficient energy management. There are already hundreds of IoT technologies available to monitor and build smart homes.

Smart Cities - Smart surveillance, safer and automated transportation, smarter energy management systems and environmental monitoring are all examples of IoT applications for smart cities.

3 NEED OF BIG DATA

The true value of IoT is not in the internet connected devices themselves; the value lies in making context-aware and relevant data and turning the result into enterprise-grade, tangible, actionable business insights. The IoT and big data are clearly intimately connected: billions of internet-connected things will, by definition, generate massive amounts of data. As the Things turn more digital, IoT will analyze - variety sources (structured and unstructured), types of data (text, audio, video, image and binary) and respond intelligently in real time.

Big data, meanwhile, is characterized by *four Vs* - volume, variety, velocity and veracity[3]. That is, big data comes in large amounts (volume), with a mixture of structured, semi-structured and unstructured data (variety), arrives at (often real-time) speed (velocity) and can be of uncertain provenance (veracity). Such information is unsuitable for processing using traditional SQL-queried relational database management systems (RDBMSs), which is why a constellation of alternative tools – notably Apache's open-source *Hadoop* distributed data processing system, various *NoSQL* databases and a range of business intelligence platforms - have evolved to serve such a complex process.

Big Data is being generated at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Much of this data is coming to us in an unstructured form, making it difficult to put into structured tables with rows and columns. To extract insights from this complex data, Big Data projects often rely on cutting edge analytics involving data science and machine learning. Computers running sophisticated algorithms can help enhance the veracity of information by sifting through the noise created by Big Data's massive volume, variety, and velocity.

4 ALLIANCE WITH BIG DATA

To scale the needs of IoT, the strategy should include infrastructure and applications that process machine and sensor data, and leverage it accordingly. At the moment, IoT platforms are often custom-built functional architecture. Enterprises that take the first step into this new market should look for interoperability between existing systems and a new IoT operating environment.

The building blocks of the IoT platform must include:

Things - A major part of the IoT is not so much about smart things (devices), but about sensors and actuators. *Sensors* measure physical inputs and transform them into raw data; *actuators* act on the signal from the sensors and convert it into output, which is then digitally storable for access and analysis. These tiny innovations can measure anything from temperature, force, flow and position, to light intensity and then can be attached to everything from smart phones to the medical devices and then record & send data back into the cloud. Smart-phone would not have been smart if it does not have an array of sensors embedded in it[6]. A typical smart-phone is equipped with five to nine sensors, depending on the model.

Network Connectivity in the devices is achieved through: wireless/wired, Wi-Fi, Bluetooth, ZigBee, VPN and Cellular 2G/3G/LTE/4G. Thread as an alternative for home automation applications and Whitespace TV technologies being implemented in major cities for wider area IoT-based use cases. Depending on the application, factors such as range, data requirements, security, power demands and battery life will dictate the choice of one or some form of combination of the technologies. In March 2015, the Internet Architecture Board - a group within the Internet Society that oversees the technical evolution of the internet - released a guide to IoT networking. This outlined four common communication models used by IoT smart objects: Device-to-Device, Device-to-Cloud, Device-to-Gateway, and Back-End Data-Sharing[1].

Collaboration and Security - Human and organizational behavior is critical in realizing the value of new approaches, and it is particularly important in shifting an organization to demonstrate clearly what will change, how it affects people, and what they stand to gain from IoT applications. Tons of collected IoT data could easily contain sensitive information about people and operations, and can even lose the control of critical systems. Beyond protecting personal privacy and business secrets, as more systems become automated, the risk of attacks becomes both more likely and more impactful.

Devices themselves should be secured, as should operating systems, networks, and every other exposed piece of technology along the way. The roles of users, administrators, and managers should be individually defined with appropriate access and strong authentication embedded in the design. A multi-layered approach to security is essential, and it should have checks and balances to reinforce protection and, if necessary, diagnose any breaches. For the IoT to work effectively, all the challenges around regulatory, legal, privacy and cybersecurity must be addressed; there needs to be a framework within which things (devices) and applications can exchange data securely over wired or wireless networks. To address these challenges, one key player, *OneM2M* issued Release 1, a set of 10 specifications covering requirements, architecture, Application Programming Interfaces (API) specifications, security solutions and mapping to common industry protocols[2].

Cloud - The cloud brings needed agility, scalability, storage, processing, global reach and reliability to an IoT platform. Needed scalability can be achieved by using (1) Cloud Centric IoT - Good choice for low-cost things where data can easily be moved, with few ramifications (2) Edge Analytics - Ideal for things producing large volumes of data that are difficult, costly or sensitive to move (3) Distributed Mesh Computing - *Future-ready* multi-party things automatically collaborate with privacy intact.

Big Data Analytics - The resulting flood of IoT sensor data must be understood and made actionable in the moment and over the long term. These data points will include structured data, unstructured data, and structured time series data, as well as a variety of analytical methods. Structured data might come from ERP systems and relational databases, such as supply chain and parts listings for automobile manufacturing. Exact specifications of each component are captured as transactional updates in tightly defined fields (part number, production lot, factory, etc.). Later the data may be extracted and joined with looser, unstructured text data like service records notes from car dealerships. And a time element may come in also as the service dates for oil changes and other periodic maintenance occur. Each data type introduces more information, but combined together will yield the secret of when a part failure is happening, help diagnose the origin of the problem, and suggest a preventative maintenance fix.

Big data, in the context of the IoT, refers to analog inputs being converted to digital data and analyzed, and resulting in a response going back the other direction. Unlike some big data applications, the inputs should be at least semi-structured, but the sheer quantities and immediateness will raise other hurdles. Some analytics may need to be performed at the edge, some in the data center, and some in a cloud environment, depending on the trade-off of speed versus depth.

5 CONCLUSION

IoT is becoming disruptive yet inevitable for companies to welcome it. Creating a connected IoT ecosystem that maximizes business value, we need to collaborate: technologies, data, process, insight, action and people. The *T* of IoT is clearly important, but too often, it is the only area of focus when examining IoT in business. The Things are only the mean to an end as entities that can capture data measuring physical conditions or sometimes as actuators to affect the system. Rest of the systems need to be instrumented to leverage the data: communicating it to the right place for action - whether the cloud, data center, or edge - and then using analytics to understand data patterns and craft a response to fix or optimize. The goal of a connected IoT ecosystem is to get the most out of the Internet of Your Things in Your Context. Innovative organizations are starting to put this to use today.

ACKNOWLEDGMENTS

The author would like to thank

REFERENCES

[1] 2015. IoT: An Overview. Whitepaper. (Oct. 2015). https://www.internetsociety.org/resources/doc/2015/iot-overview

[2] 2015. IoT Interoperability. Whitepaper. (Jan. 2015). http://www.onem2m.org/images/files/oneM2M-whitepaper-January-2015.pdf

[3] 2017. Big Data. Web page. (Sept. 2017). https://en.wikipedia.org/wiki/Big_data

[4] 2017. Garner Press Release. Web page. (Feb. 2017). http://www.gartner.com/newsroom/id/3598917

[5] 2017. Internet of things. Web page. (Sept. 2017). https://en.wikipedia.org/wiki/Internet_of_things

[6] Hakim Cassimally Adrian McEwen. 2014. *Designing the Internet of Things.* Wiley.

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Lars Thørväld
The Thørväld Group
1 Thørväld Circle
Hekla, Iceland
larst@affiliation.org

Valerie Béranger
Inria Paris-Rocquencourt
Rocquencourt, France

Aparna Patel
Rajiv Gandhi University
Rono-Hills
Doimukh, Arunachal Pradesh, India

Huifen Chan
Tsinghua University
30 Shuangqing Rd
Haidian Qu, Beijing Shi, China

Charles Palmer
Palmer Research Laboratories
8600 Datapoint Drive
San Antonio, Texas 78229
cpalmer@prl.com

John Smith
The Thørväld Group
jsmith@affiliation.org

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data in Higher Education Marketing

Ashley Miller
Indiana University
admille@iu.edu

ABSTRACT

While the collection of vast amounts of data in the world higher education has happened for decades, the use of big data applications and analytics is fairly new to this environment. There is a need to understand how the use of big data can help higher education further understand student needs as well as stay relevant in a digital and evolving age of technological advances, tools, and skills. Further, the population of students going to college is on the decline which increases competition and the need for institutions to be more strategic in their efforts for attracting students to their institutions. This purpose of this paper is to explore at a very high level how higher education could utilize big data to inform marketing initiatives in recruiting and enrolling students.

KEYWORDS

Big Data, Higher Education, Marketing, Analytics LaTeX

1 INTRODUCTION

Todayfis colleges and universities are drowning in data. With the emergence of big data, institutions are now faced with providing useful analysis and reports to a variety of stakeholders including administrators, professors, as well as to the students themselves. A variety of challenges lie in the path of institutions using big data effectively such as finding the necessary skill set for staff, technology tools and resources, as well as understanding then what to do with the data collected to better inform decision-making.

While there is literature that addresses utilizing big data for learning analytics and even course enrollment and development, as Daniel states, there is still filimited research into big data in higher educationfi (2015). This paper seeks to explore ways in which higher education could utilize big data in their marketing efforts for recruiting and enrolling students as well as what gaps may still exist in the quest to understand todayfis college search as they make their choice on which university to attend.

2 CURRENT ENVIRONMENT

While high school graduation rates have increased over time, the number of those who go on to pursue higher education has been on the decline for the past four years (The Atlantic, 2016). Meanwhile, the number of four-year institutions in the United States has increased with there now being more than 3,000 college options (NCES, 2014). Increased competition and fewer students have made the higher education marketplace crowded and convoluted. There are a variety of factors that go into a studentfis decision on where to attend and ultimately what area to study. In their 2013 trends report, the Lawlor group identified a number of aspects that will impact the higher education landscape, among those included are:

- Demographics of todayfis college student is changing with more women attending college than men in addition to an increase in ethnic and socio-economic diversity as well as first-generation students.
- The college search process today happens primarily in the digital space which include third-party websites, email, social media, and digital advertising. This fiGeneration Zfi of student grew up in a technology rich and connected environment which means that colleges have to also be ficonstantly onfi in their effort to recruit and enroll students.
- The need to showcase the fivaluefi of going to college, not only through the quality of education received relative to the price paid but also through outcomes-level data including placement rates and starting salaries of recent graduates.

3 BIG DATA TO SEGMENT STUDENTS BY DEMOGRAPHICS

With these trends in mind, there is a need for institutions to more effectively target and recruit students. Big data can be one way to better inform these efforts and also help with the return-on-investment (ROI) for advertising and marketing related efforts. Other universities have capitalized on utilizing big data in attracting students. For instance, St. Louis University described a process of retroactively looking at demographics of students who succeeded at the university and had high satisfaction scores (The Atlantic 2017). This information coupled with nearly fi120 data pointsfi gave insight to the admissions team when exploring new markets as well as identified clusters of students that may be interested in attending St. Louis University. The university was then able to develop a targeted digital campaign in these areas that they believed include students who would be a figood fitfi. With the reliance on big data, the university was able to reduce costs as the need to mass market went away and ultimately increased enrollment as a result.

4 BIG DATA TO UNDERSTAND STUDENT BEHAVIOR ONLINE

The web environment is common tool in college exploration as a report by Ruffalo Noel Levitz shows that three out of four high school students state that the institutionfis website is their most used resource when exploring colleges (2016). Web analytics provides a wealth of information on users such as how much time is spent on certain pages, bounce rate, paths in website exploration and ultimately conversion rates when various goals are completed such as scheduling a tour or filling out an application for college. Google Analytics is one tool used to track and evaluate efforts on websites. Higher education institutions could take advantage of this tool by tracking top pages viewed, geography and age of visitors, as well as areas where they may be filosingfi students in the information search process. With this data, institutions can identify opportunities for improvement in ensuring students are finding the information they need in a timely and efficient way as well as develop customized marketing efforts to invite students back into the experience to complete various calls-to-action.

5 BIG DATA TO CONVEY VALUE

Utilizing big data to understand outcomes of students can help tell the value story to prospective students. By tracking the experiences among currently students during their four (or more) year college career, predictive analytics could be implemented to determine which combination set of experiences best contribute to the fisuccessfi of a student. One university to showcase the impact of this data on outcomes is American University with their fiWe Know Successfi tool (CITE SOURCE). By collecting data from graduates over time, the university can further showcase to others

6 CONCLUSIONS

Competition for todayfis student will only increase with changing educational needs and offerings, including development of emerging degree programs as well as delivery, including online classes. In order for the use of big data in higher education marketing to be successful, there are basic measures that have to be met. RAND outlines some key considerations when using big data for effective decision-making which include: accessibility, quality, timeliness, and motivation to use.

For marketers in higher education, they need to have access to necessary data about current as well as prospective students to better tailor messaging and marketing efforts appropriately. With this, the validity of available data is key as making decisions based on fibadfi or incomplete data can be problematic and costly for an institution. Given the nature of the web environment that is constantly changing, obtaining data in a timely manner is crucial so action can be taken at the right time. Further, there has to be a culture within an institution that motivates others to make data-driven decisions.

REFERENCES

Big Data Applications in Electric Power Distribution

Swargam, Prashanth
Indiana University Bloomington
107 S Indiana Ave
Bloomington, Indiana 47408
pswargam@iu.edu

ABSTRACT

Now-a-days, the process of storing the power measurements have changed. Conventional meters are replaced by the smart meters. New distribution management systems like SCADA and AMI are implemented to monitor power distribution. These smart meters record the readings and communicate the data to the server. However, these systems are designed to generate the readings very frequently i.e., 15 minutes to an hour. Upon that, smart meters are being deployed at every possible location to improve the accuracy of the data. This advancements in electric power distribution system results in enormous amounts of data which requires advance analytics to process, analyse and store data. This paper discusses about the implementation of Big Data technologies, challenges of implementing Big Data in Electric Power Distribution Systems. [1]

KEYWORDS

Big Data, Power Distribution,Smart Power

1 INTRODUCTION

Volume of data is increasing. According to forbes, it is said that, worldfis data utilization will increase to 44 zettabytes from the current utilization of 4.4 zettabytes. To process this data, Big Data analytics will be useful. But, instantiating a big data architecture is not easy task.

In electrical Power Distribution industry, data deluge is picking its pace. The data which was recorded for month, is now being noted for very small intervals. This quadruples the amount of data that should be process. There is a lot of potential work to be put in for designing a good Big Data architecture to process and analyse this data. Most of the power generation units are developing their infrastructure to support these designs.

1.1 Data Sources

Smart meters which are placed at customerfis vicinity will record the consumption of a specific group of customersfi. This data can be used to analyse the behaviour of customer for certain circumstances of weather and environment.

Distribution systems which manage the distribution of power, generate large amount of data related to voltages and currents at various levels of distribution. This data is very important in analysing the load level and demand for the distribution circle.

Power measuring units at generation. This data is used to analyse the behaviour of generator and amount of power generation that will be required to supply enough power. This data will be used to decide the functioning of generators.

Old market data will be used to analyse the pricing and marketing strategies. These data is more focused on users and their behaviour.

1.2 4 v's in Big Data in Power Distribution System

Volume: The data is periodically generated by many data sources like smart meters, machines and other appliances. Variety: Each data source in electric power distribution system is explicit to each other. Each source has its own frequency of data generation and its own method of data generation. Thus, the data is heterogenous. Velocity: is the speed at which the data is available for the end user. Veracity: It deals with the correctness of the data. As all the data collected by sensors, meter tend to have various losses, correction algorithms should be defined to find the accurate data. Their might be chances for data transfer losses.

REFERENCES

[1] Amr A. Munshi and Yasser A.-R.I.Mohamed. 2017. Big data framework for analytics in smart grid. *Electric Power Systems* (2017).

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data Analytics for Municipal Waste Management

Andres Castro Benavides
Indiana University
107 S. Indiana Avenue
Bloomington, Indiana 43017-6221
acastrob@iu.edu

Mani Kumar Kagita
Indiana University
107 S. Indiana Avenue
Bloomington, Indiana 43017-6221
mkagita@iu.edu

ABSTRACT

As waste management becomes a greater concern for cities and municipalities around the world, big data analysis has the potential to not only help assess the current waste management strategies but also provide information that can be used to optimize the systems used in various institutions, local government, companies, etc.

KEYWORDS

Waste Management, Big Data, Local Government

1 INTRODUCTION

In the current fast paced society, as production of goods increases and new distribution chains constantly change, the production of disposed materials and goods, from now on called solid waste, has increased over the past ten years, going from around 0.64 kg per person per day of solid waste to approximately 1.2 kg per person per day, and it is expected to increase to about 1.42 kg. [?] this causes the problem of waste management to increase in complexity and magnitude.

Because of this, the different local governments and organizations have seen the need to develop regulations to control the different features, segments, processes? Of the action of disposal. From the moment the material is discarded until the moment the material reaches itfis ultimate destination: recycling plant or landfill. This set of systematic regulations is called solid waste management [?]

2 WASTE MANAGEMENT

The amounts of disposed material and itfis composition vary depending on the country, place and activity that is performed at the site where the waste is generated. [?] There are also important differences between the general composition of the waste generated in rural area and what is produced in urban area, the waste produced in the later is highlyu influenced by the culture and the practices of our modern society. [?] p47 to 63

For this reason, every process related to waste management-transportation, storing and final disposition, among others- must be engineered and tailored to fit the specific needs of each case.

In general, decisions can be classified as optimal, good, or fortuitous. [?] and this can be applied to Waste Management.

Having that Good decision-making is mostly based on experience, comparison of elements and trial and error, and that fortuitous decision-making have no scientific base; one must always try to solve the problem -in this case waste management related- with Optimal Decision making, that requires techniques and technologies provided by other fields. [?]

3 BIG DATA AND WASTE MANAGEMENT

By collecting and storing large volumes of data related to types of waste, quantities, periodicity, and composition; usually from independent sources. Big data can be interpreted in a way that allows the different actors that intervene in Waste Management to make Optimal Decisions. [?]

3.1 Opportunities for Waste Management Optimization

The process of solving a math program requires a large number of calculations and is, therefore, best performed by a computer program. [?]

4 OPPORTUNITIES FOR WASTE MANAGEMENT OPTIMIZATION

4.1 Statistics and Waste Management

There are many data analysis methods that are used when studying waste management, but the two most popular are PCA and PLS1. [?]

Lingo is a mathematical modeling language designed particularly for formulating and solving a wide variety of optimization problems including linear programing. Lingo optimization software uses branch and bound methods to solve problems of this type. [?]

4.2 GIS Analytics

When it comes to Geographical Information Systems (From now on GIS) There are multiple software and hardware options in the market. From paid software like ArcGIS to Open and free software like GVSIG, there are solutions that can help interpret large data sets, apply statistics and algorithms of different kinds and display them in a way that make reference to a geographical space.

/cite shahrokni2014big

The second category of studies focuses on minimizing transportation of waste collection through optimal routing algorithms. For example Kim et al [18] use two methods to calculate an optimal set of routes, the ffirst being Solomonfis insertion algorithm, the second being a clustering algorithm. Their aim was to minimize the driven distance, as well as to balance the workload. At the same time, the constraint of legally prescribed lunch breaks (so called time-window problem) had to be satisffied. McLeod and Cherrett [19] suggestedarouteoptimizationforthreeareasandconnectedwaste companies in North Hampshire (UK). By applying simple rerouting, sharing of routes between the 3 areas and adding vehicle depots at the waste disposal sites, they estimated annual savings as large as 10,000 km for the studied routes (this covers one ffifth of all routes in North Hampshire). Another study performed by Wy, Kim and Kim [20] studied

a routing algorithm for waste collection using roll-on/roll-off containers, again while factoring in the time windows. Buhrkal, Larsen and Ropke [21] were one of the ffirst to suggest the environmental importance of optimizing waste collection itineraries. They utilized an adaptive large neighborhood search algorithm, and a clustering method and their scope was residential waste collection. Depending on the computation time, using the actual collection points and lunch time windows, the savings amounted to 13 percent average. With larger time windows and better starting conditions, heuristics with a distance reduction of up to 45

/cite shahrokni2014big

Many data analysis methods are used when studying waste management, but the two most popular are PCA and PLS1. [**?**]

5 CONCLUSIONS

There are different tools to optimize the different waste management practices and to improve the information available for decision makers...

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

A MORE HELP FOR THE HARDY

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

Big Data Analytics for Municipal Waste Management

Andres Castro Benavides
Indiana University
107 S. Indiana Avenue
Bloomington, Indiana 43017-6221
acastrob@iu.edu

Mani Kumar Kagita
Indiana University
107 S. Indiana Avenue
Bloomington, Indiana 47405
mkagita@iu.edu

ABSTRACT

As waste management becomes a greater concern for cities and municipalities around the world, big data analysis has the potential to not only help assess the current waste management strategies, but also provide information that can be used to optimize the systems used in various institutions, local government, companies, etc.

KEYWORDS

Waste Management, Big Data, Local Government

1 INTRODUCTION

Concept of waste managementfi

Solid Waste Management (SWM) is a set of consistent and systematic regulations related to control generation, storage, collection, transportation, processing and land filling of wastes according to the best public health principles, economy, preservation of resources, aesthetics, other environmental requirements and what the public attends to [1]

Managing solid waste is one of the most essential services which often fails due to rapid urbanization along with changes in the waste quantity and composition. Quantity and composition vary from country to country making them difficult to adopt for waste management system which may be successful at other places. Quantity and composition of solid waste vary from place to place [3]

2 OPPORTUNITIES FOR WASTE MANAGEMENT OPTIMIZATION

By collecting and storing data related to types of waste, quantities, periodicity and composition.

2.1 GIS Analytics

3 STATISTICS AND WASTE MANAGEMENT

While rural area usually generates organic and biodegradable, urban area produces waste influenced by culture and practices of society. [3] p47 to 63

There are many data analysis methods that are used when studying waste management, but the two most popular are PCA and PLS1. [2]

decision makers should distinguish between optimal, good, and fortuitous decision-making. In the optimal decision making, one can solve the optimal problem using the techniques available in other fields. In this solution method, generally some constraints (criteria) are consid- ered, where the function(s) is to be optimized through applying some methods. Good decision-making is done based on experience, trial and error or comparison between different options of the integrated SWM. Although it is possible to choose decisions close to the optimal state using this decision-making method, today these methods are not applicable due to increased number of different combinations in the decision-making process. In the fortuitous decision-making, since decisions are made with no scientific base, so the results are not acceptable [1]

The process of solving a math program requires a large number of calculations and is, therefore, best performed by a computer program. Lingo is a mathematical model- ing language designed particularly for formulating and solving a wide variety of optimization problems including linear programing. Lingo optimization software uses branch and bound methods to solve problems of this type. [1]

4 CONCLUSIONS

Working on this

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

A MORE HELP FOR THE HARDY

Of course, reading the source code is always useful. The file `acmart.pdf` contains both the user guide and the commented code.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

[1] Mohsen Akbarpour Shirazi, Reza Samieifard, Mohammad Ali Abduli, and Babak Omidvar. 2016. Mathematical modeling in municipal solid waste management: case study of Tehran. *Journal of Environmental Health Science and Engineering* 14, 1 (18 May 2016), 8. https://doi.org/10.1186/s40201-016-0250-2

[2] K. Bofkhm, E. Smidt, and J. Tintner. 2013. Application of Multivariate Data Analyses in Waste Management. In *Multivariate Analysis in Management, Engineering and the Sciences*, Leandro Valim de Freitas and Ana Paula Barbosa Rodrigues de Freitas (Eds.). InTech, Rijeka, Chapter 02, 24. https://doi.org/10.5772/53975

[3] R. Chandrappa and J. Brown. 2012. *Solid Waste Management: Principles and Practice*. Springer Berlin Heidelberg, Berlin. https://books.google.com/books?id=kUOwuAAACAAJ

Automated Information Extraction in Electronic Health Records

Nicholas J Hotz
Indiana University
nhotz@iu.edu

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size. [1]

ACKNOWLEDGMENTS

The authors would like to

REFERENCES

[1] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY.

An Overview of Big Data Applications in Mental Health Treatment

Neil Eliason
Indiana University
107 S. Indiana Ave.
Bloomington, Indiana 47405
nreliaso@iu.edu

ABSTRACT

Mental health treatment presents with complex informational challenges, which could be effectively tackled with big data techniques. However, as researchers and treatment providers explore these applications, they find a lack of infrastructure and ethical concerns hamper their progress. [? ? ? ? ? ?].

KEYWORDS

Mental Health Treatment

1 INTRODUCTION

Big Idea: Mental Illness is a big societal problem, which could benefit from a big data solution.

Mental health difficulties are a common problem across the United States, and worldwide. Mental illness of some kind was prevalent among 17.9 % of Americans in 2015, and of that number 4% experienced serious functional impairment as a result [?]. A 2014 meta-analysis study estimated that the worldwide prevalence of mental illness was 17.6% and that 29.2% of the world population would experience mental illness at some point during their life [?].

The effects of these disorders on individuals and societies is costly. The US Center for Disease Control and Prevention estimated that 36,035 people died during a suicide attempt in 2008, and that 666,000 sought emergency room care for self harming behavior [?]. In 2013, the Social Security Administration reported that 1,947,775 persons received social security/disability benefits for either a mood or psychotic disorder, which is around 19% of all recipients [?]. It is estimated that mental health issues had a $100 billion cost on the US economy in 2002 [?] (more recent stats), and in 2015 there were over 12,000 mental health treatment facilities in the US [?].

1.1 The State of Mental Health Treatment

Brief summary of Mental Health Treatment Big picture impact on population and economy:

Goal of mental health treatment Techniques

Brief summary of Big Data Big picture impact on everyday life and economy Goal of Big data analytics Techniques

Thesis

2 BIG DATA APPLICATIONS IN MENTAL HEALTH TREATMENT

Introduce concept of different levels of maturity

2.1 Mature Applications

2.2 Developing Applications

2.3 Initiatives

3 CONCLUSIONS

3.1 Barriers

3.2 Future Directions

form:

.1 Introduction

.2 The Body of the Paper

.2.1 Type Changes and Special Characters.

.2.2 Math Equations.

Inline (In-text) Equations.

Display Equations.

.2.3 Citations.

.2.4 Tables.

.2.5 Figures.

.2.6 Theorem-like Constructs.

A Caveat for the TEX Expert.

.3 Conclusions

.4 References

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

A MORE HELP FOR THE HARDY

Of course, reading the source code is always useful. The file `acmart.pdf` contains both the user guide and the commented code.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

2

Impact of Big Data on the Privacy of Mental Health Patients

J. Robert Langlois
Indiana University Bloomington, School of Informatics and Computing
langloir@umail.iu.edu

ABSTRACT

Society has experienced a lot of benefits with the introduction of technology. Today, one of the essential functions of technology is the collection, storage, processing, and transmission of data. The healthcare industry, including mental health services, are huge benefactors of these advances in technology. From birth, medical facilities start collecting information about all individuals; they do so even up to the point of death and all points in between. Over a lifetime, that is an abundance of information about an individual. The question that must be answered is, "How is that data protected to ensure patients' privacy rights?" The more information collected on individuals, the more responsibility is assumed by those who collect data; methods for how the data is collected, used and shared must ensure the protection of patients' privacy rights. This challenge is one that needs to be navigated and addressed by medical professionals and facilities, policymakers, and the individuals whose data is collected. Specifically in the mental health field, by resolving patients' privacy concerns, policymakers and researchers can transform the field by introducing more cost effective strategies, ensuring patients' sense of security, and establishing new and more appropriate norms to communicate sensitive health information.

1 INTRODUCTION

We live in an era where data is constantly being produced; data exists everywhere in large quantities. The advances in technology have opened the door for businesses to collect inconceivable amounts of information on individuals via emails, smart-phones, sensors, and other technology devices. The 21st century has witnessed a data explosion; many fields have experienced a data deluge that can contribute to boast the economy via data analysis, make new discoveries based on existing data, respond to health problems in a quicker manner, and so forth. While it is worth celebrating the rapid innovations in technology and the presence of huge amounts of data, it is also crucial to consider the number of barriers and risks that come with the increased availability of data; often refers to as big data. One of the barriers that big data faces is privacy. In the healthcare industry, for example, there are protocols to accessing data that can cause financial burdens and can be time-consuming. The cost of collecting, disseminating, and organizing patient information, along with the time it takes to handle the information are some of the challenges. There are also very serious concerns regarding who can have access to what kind of patient information. Policymakers have a very important role in establishing more up-to-date policies and parameters that address the massive amounts of information available and the appropriate ways to collect, share, and house the data. "When considering the risks that big data poses to individual privacy, policymakers should be mindful of its sizable benefits"[5]. While it is important to address the numerous advantages of big data, it remains relevant to figure out ways to prevent data leakage, and to protect the privacy of individuals. This paper showcases the advantages of big data and the ways to overcome the individual privacy concerns. [3]

2 ADVANTAGES OF BIG DATA

Big data analysis presents numerous advantages. For instance, it helps businesses to increase their productivity. This done through a process of analyzing raw data that produces information that identifies trends and patterns that will help businesses make cost effective decisions. It is also helpful in aiding government agencies to improve public sector administration, and assists global organizations in analyzing information that has wide-reaching impact on the world. The information produced by big data can help medical professionals to detect diseases in earlier stages. Some other advantages of big data analysis is present in many different areas, such as: smart grids, which monitor and control electricity use; traffic management systems, which provide information about transportation infrastructure likes roads and highways, mass transit, construction, and traffic congestion; retail by studying customer purchasing behavior to improve store layout and marketing; payment processing by helping to detect fraudulent activity, etc.[5].

Certain research studies have supported the idea that big data allows for real time tracking of diseases and the development, prediction of outbreaks, and facilitates the development of personalized healthcare. Big data can also be used to maximize profits in many disciplines, including healthcare if harnessed properly.[6]. As indicates in [2] "by harnessing big data, businesses gain many advantages, including increased operational efficiency, informed strategic direction, improved customer service, new products, and new customers and markets." While data exists in huge quantities in many fields, including the health care field, individual privacy concerns remain a big problem that policymakers have to tackle to meet current trends in data collection. Improved methods of protecting very personal, private and sensitive health information is needed order to allow for safe, necessary and adequate access to protected health information within the health care industry.

3 BARRIERS TO BIG DATA IN HEALTH-CARE

One of the barriers faced by big data analysts in health care, including mental health services, is privacy. Regardless of the efforts policymakers try to establish, the different strategies in place to protect individual health information can pose serious challenges that scientists have to wrestle with when it comes to big data analytics. One of the most notable efforts that policymakers have introduced to secure health information, is the creation of the Health Insurance Portability and Accountability Act (HIPAA) in 1996. HIPAA has established norms for data privacy and has mandated security provisions for safeguarding medical and mental health information. Every provider in the healthcare industry must comply with HIPAA privacy laws if they want their practices to remain up and running.

The HIPAA laws prohibit providers from sharing patients' information without their consent. The challenge for big data analysts is that a lot of times, patients refuse to share their personal information for research purposes due to fears that the health issue will be the cause of being ostracized, discriminated against, marginalized, etc. "The unintended release of a person's health information into the public realm has huge potential to undermine personal dignity and cause embarrassment and financial harm"[6]. While the healthcare field is faced with a huge increase in health information, individual privacy concern remains a huge conundrum for big data analysis. What can policymakers do to overcome individual privacy concerns, but still allow for the sharing of information that would be for the better good of society at large?

4 WAYS TO OVERCOME PRIVACY CONCERN

4.0.1 Data Anonymization. One way policymakers can protect individual privacy is by making the data anonymous. Researchers have identified three types of data: personal and proprietary data that is controlled by individuals; government-controlled data, which government agencies can restrict access to; and, open data commons, which means that the data is centrally located and available to all. Big data analysts and researchers have advocated for linking data together that can help to improve health care planning at both the patient and population levels. They also argued for an increase in the amount of information that is available in open data commons. Although the anonymization of data appears to be a great technique that policymakers could espouse to address privacy concerns, other studies have indicated that some data can be traced back to their respective individual; thus, destroying the argument for anonymity.[6]. " Every copy of data increases the risk of unintended disclosure. To reduce this risk, data should be anonymized before transfer; upon receipt, the recipient will have no choice but anonymize it at rest...And re-identification is by design, in order to ensure accountability, reconciliation and audit." If proper norms are established for data analysis, this can potentially contribute to improvements in the health care industry.

Still, there are others that have advocated for data de-identification and data minimization. The term de-identification is the process by which the data is made anonymous. The proponents of this process explain that this protective measure is valid under security and accountability principles, but admonish that policymakers should think about other ways to protect patients' privacy. The term data minimization, describes the extent to which organizations can limit the collection of personal data. It is worth noting that data minimization is contrary to big data analysis because data minimization encourages deleting data that is no longer in use in order to protect privacy; whereas, big data analysts would prefer to archive the data for ulterior usage. While this technique can help protect privacy, it is antithetical to big data analysis because it contributes to reducing the amount of data collection that could be used in data analysis to make new discoveries, respond to crises, and maximize profits [5]. As found in [1], privacy principles should be introduced during the process of data architecture; privacy should be incorporated into the design and operational procedures. In so doing, personal health care data will be protected against malicious hackers who try to access individuals' personal health information for the purposes of stealing individuals' identity. Another type of data that has been introduced to the healthcare industry is the concept quantified self data. It can be understood as the data produced by individuals that engage in self-tracking of personal health information, such as heart rate, weight, energy levels, sleep quality, cognitive performance, etc. These individuals use devices like smart-phones, watches, and wearable technology sensors in the collection of their personal data and biometrics. It has been shown that 60 percent of U.S. adults are tracking their weight, diet or exercise routines, while 33 percent are monitoring their blood sugar, blood pressure, sleep patterns, etc. This indicates that there is a vast amount of health information that has been produced by individuals. What is done with all of this data? This massive supply demonstrates the need to develop policies and protocols that involve individual patient consent to share their collected data; this data can be critical to the advancement of health-care with the support of data analysis. Before that can be done, however, we must first establish the proper norm to use this type of data so that the privacy of individuals can be protected; this ought to be primary action to take. [4].

5 CONCLUSION

We have seen that healthcare data exists in large quantities; however, privacy concerns are one of the biggest barriers that scientists face when it comes to utilization of healthcare data. Certain researchers have proposed data anonymization as a solution to privacy concerns, while others have proposed a minimization of the amount of data collected on individual patients. "Privacy concerns exist wherever personally identifiable information or other sensitive information is collected and stored in any form"[2]. This indicates that scientists will always have to wrestle with privacy concern whenever they are dealing with personal health information.

A HEADINGS IN APPENDICES

the body of this document in Appendix-appropriate form:

A.1 Introduction

A.2 The Body of the Paper

A.2.1 Type Changes and Special Characters.

A.2.2 Math Equations.

Inline (In-text) Equations.

Display Equations.

A.2.3 Citations.

A.2.4 Tables.

A.2.5 Figures.

A.2.6 Theorem-like Constructs.

A Caveat for the TEX Expert.

A.3 Conclusions

A.4 References

`.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

REFERENCES

[1] Ann Cavoukian and Jeff Jonas. 2012. *Privacy by design in the age of big data.* Information and Privacy Commissioner of Ontario, Canada.

[2] Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani. 2014. Big data: survey, technologies, opportunities, and challenges. *The Scientific World Journal* 2014 (2014).

[3] Joachim Roski, George W Bo-Linn, and Timothy A Andrews. 2014. Creating value in health care through big data: opportunities and policy implications. *Health affairs* 33, 7 (2014), 1115–1122.

[4] Melanie Swan. 2013. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data* 1, 2 (2013), 85–99.

[5] Omer Tene and Jules Polonetsky. 2012. Big data for all: Privacy and user control in the age of analytics. *Nw. J. Tech. & Intell. Prop.* 11 (2012), xxvii.

[6] J Van Den Bos, K Rustagi, T Gray, M Halford, E Zeimkiewicz, and J Shreve. 2011. Health affairs: At the intersection of health, health care and policy. *Health Affairs* 30 (2011), 596–603.

Bigdata in Clinical Trails

Mohan Mahendrakar
Indiana University
P.O. Box 1212
Raleigh, North Carolina 43017-6221
mohan1.data@gmail.com

ABSTRACT

The vast volumes of data collected across the clinical trials process offers pharma and biotech the opportunity to leverage the information. ACM SIG Proceedings.

KEYWORDS

Bigdata, Clinical, Trails, Healthcare, Phase I, Phase II, Phase III, Phase IV

1 INTRODUCTION

After transforming customer-facing functions such as sales and marketing, big data is extending its reach to other parts of the enterprise. In research and development, for example, big data and analytics are being adopted across industries, including pharmaceuticals.

2 INTEGRATE ALL DATA

Having data that are consistent, reliable, and well linked is one of the biggest challenges facing pharmaceutical clinical Trails. The ability to manage and integrate data generated at all phases of the value chain, from discovery to real-world use after regulatory approval, is a fundamental requirement to allow companies to derive maximum benefit from the technology trends. Data are the foundation upon which the value-adding analytics are built. Effective end-to-end data integration establishes an authoritative source for all pieces of information and accurately links disparate data regardless of the sourcefi!?be it internal or external, proprietary or publicly available. Data integration also enables comprehensive searches for subsets of data based on the linkages established rather than on the information itself. fiSmartfi algorithms linking laboratory and clinical data, for example, could create automatic reports that identify related applications or compounds and raise red flags concerning safety or efficacy.

Implementing end-to-end data integration requires a number of capabilities, including trusted sources of data and documents, the ability to establish cross-linkages between elements, robust quality assurance, workflow management, and role-based access to ensure that specific data elements are visible only to those who are authorized to see it. Pharmaceutical companies generally avoid overhauling their entire data-integration system at once because of the logistical challenges and costs involved, although at least one global pharmaceutical enterprise has employed a fibig bangfi approach to remaking its clinical IT systems.

3 CHALLENGE

Big pharma companies typically keep their cards close to the vest because it costs so much to develop a drug throughout its lifetime. From discovery to prescription pad, a typical medication can take twelve years and $4 billion to shepherd through its lifecycle, a significant investment that would be hard to recoup if everyone had the secret to the newest blockbuster pill, especially since only ten percent of drugs ever make it to market.

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

Waste in Healthcare Using Big Data to help optimize our Healthcare spend

Paul Marks
Indiana University
107 S Indiana Ave
Bloomington, Indiana 47405
pcmarks@iu.edu

ABSTRACT

The cost of healthcare includes the cost of inefficient services. While this could include topics such as misdiagnosis, less effective treatment plans, and more efficient use of types of services (emergency room vs. immediate care vs. telemedicine) the purpose of this paper is the cost of Fraud, Waste, and Abuse (FWA) within the system. The estimate for what percentage of cost are attributable to FWA can vary from insurer to insurer. Medicare estimates that 11 percent of its payments for Original Medicare are improper primarily due to FWA. (Reference 2016 Financial Report). The question is "How can we use big data analysis to help minimize these costs and thus optimize the money spent on healthcare."

KEYWORDS

i523, hid327, Fraud, Waste, Abuse, Healthcare, Medicare, Medicaid, FWA

1 INTRODUCTION

The definition of verb waste includes "to spend or use carelessly" (Reference webster). When money spent on healthcare goes to FWA, it is money being spent carelessly. We, FWA is all of our issue, are not doing enough to ensure the money is used for the goods or services provided. FWA are varying degrees of culpability of waste. The Centers for Medicare and Medicaid Services (CMS) in part defines fraud as "is knowingly and willfully executing, or attempting to execute, a scheme or artifice to defraud any health care benefit program", Waste as "overusing services, or other practices that, directly or indirectly, result in unnecessary costs", and Abuse as "involves payment for items or services when there is not legal entitlement to that payment and the provider has not knowingly and/or intentionally misrepresented facts". (Reference training doc) In combination these cost the United States healthcare system 80 billion dollars (Reference Vinil Menon doc) annually. Advances in big data technology can help reduce these losses. Big data can offer the ability to look at data in real time to determine if a claim is legitimate or not. Historically, due to the amount of data involved, this type of analysis would have to happen after the claims have been paid. Specific models targeting specific schemes would identify FWA. Big data can help lower the cost of health-care in the United States by identifying FWA claims and stopping payment before it occurs.

Gregor includes a citation [1].

2 FWA IN HEALTHCARE

It is easy to understand the problem FWA poses. Healthcare funds are of limited quantity. Insurance helps to spread the cost among groups of people, but does not provide limitless funds. As costs increase, so do premiums or direct payments for health-care. Reducing costs by eliminating as much FWA as possible is one solution. In order to fully utilize advances in technology, the sources of information must be brought together. Sources include claims (current and historic), clinical, provider, geospatial, and other sources of information. The problem is the deluge of information and how to process it fast enough. Payments are generally made in so many days depending on the insurer and their agreement with providers. Using CMS as an example, being a government entity much of their data is available publicly, it is easy to get an idea of the amount of data. Medicare processed 1.2 billion claims in 2014, covering 53.8 million beneficiaries, with 6,142 million hospitals, and 1,173,802 non-institutional providers. (Reference 2015 Stats doc)

2.1 Ideas for Big Data

So how can big data be used to approach this issue? The theme could be divide and conquer. Leveraging big data tools such as Hadoop they could divide the different sources of information into data lakes, looking at each source separately, and then combining the results. Figure 1 (Reference Dallas Thornton) on page 5 shows sources of information and what level of FWA they are generally related to. The highest level combines sets of data. "Level 7 combines all previous data views and concerns all fraud that is part of criminal networks which involve many different beneficiaries and/or providers. This much larger data view, spanning billions of claims in the case of Medicaid, is the most rich, delivering the ability to perform complex network analysis that could detect intricate conspiracies. However, performance of analysis here will be much lower than in previous levels." (Reference Dallas Thornton)

2.2 Big Data Techniques for FWA

Traditionally programs are written to look for specific sets of circumstances. Leveraging existing knowledge about the data and using it to look for specific patterns is known as supervised in big data terms. "There are several supervised fraud detection methods such as: Bayesian Networks, Neural Networks (NNs), Decision Trees, and Fuzzy Logic. NNs and decision trees are the most popular fraud detection methods because of their high tolerance of noisy data and huge data set handling." There are also unsupervised methods in which data is fed into the system without preexisting notions of what to look for. (Reference Namrata Ghuse) Unsupervised methods sort through data and find relationships and groupings of related information, find clusters of what could be considered normal, and determine where the outliers are. Applying unsupervised methods to healthcare data will identify patterns that will then have to be verified as FWA or acceptable patterns. This greatly increases the ability to fight FWA by having the machine pinpoint where to look in all the data available. Suddenly the task of finding fraud is not as daunting. By leveraging both of these techniques FWA can be discovered at an accelerated pace. The number of models the system knows will grow over time as more data is fed into it and more patters are discovered and verified.

2.3 The Future

Currently there is still a certain amount of honor built into healthcare. If a claim is submitted by a valid entity, using the correct process, and everything is in order then it is most likely paid. This is done without any specific proof of the services being provided. With more and more healthcare information being digitized this may not be the case in the future. X-rays, lab tests, clinical notes, etc. are all being stored digitally. Computers are now able to interpret images and unstructured text very accurately. By linking this data to claims data the clinical information could be required as part of claims payment. An x-ray of broken bone, notes which support a diagnosis, Magnetic Resonance Imaging files, could all be interpreted automatically. Not only would the data be used to compare to the claims information, but to other images/notes on file to ensure that the same files were not being submitted with multiple claims. It could know what one individual medical history looks like compared to another similar to how facial recognition is able to match like images. This would not be possible without the ability to process massive amounts of data quickly.

3 CONCLUSIONS

While there may be disagreement on aspects of healthcare in America, everyone should agree that eliminating Fraud, Waste, and Abuse within the system is good for everyone. FWA costs billions of dollars annually. Just a 1 percent reduction in the estimated 80 billion dollars annually would result in 800 million dollars in savings. With this amount of money at stake significant investments should continue to be made in leveraging advanced big data technologies into solving this problem. Because of the continued rise in the amount of data collected traditional programming cannot keep up with the pace. Advanced techniques must be leveraged which can learn in an unsupervised manner. While there will inevitably be privacy concerns, new sources of information must be brought into the fight against FWA. Historically payers of healthcare claims, insurers, have not had the ability to require actual evidence that a service has taken place. By leveraging advances in big data and combining data stores such as electronic health records into the payment process a difference can be made in the amount of money spent on healthcare in America.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

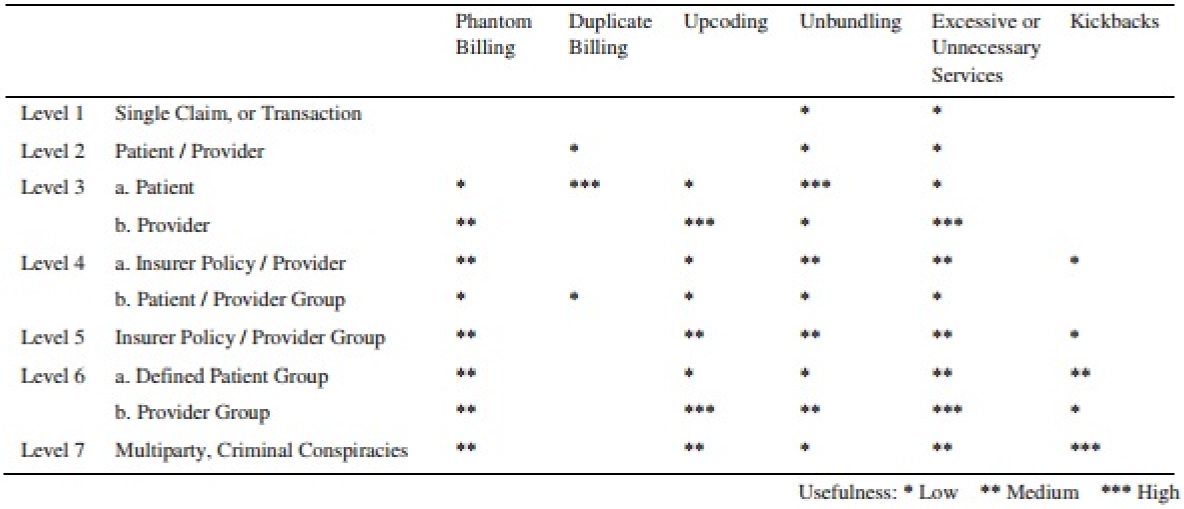
REFERENCES

[1] 2017. Waste. Online. (09 2017). https://www.merriam-webster.com/dictionary/waste

[Table 1 about here.]

List of Tables

Table 1: Types of Fraud and their related Sources



		Phantom Billing	Duplicate Billing	Upcoding	Unbundling	Excessive or Unnecessary Services	Kickbacks
Level 1	Single Claim, or Transaction				*	*	
Level 2	Patient / Provider		*		*	*	
Level 3	a. Patient	*	***	*	***	*	
	b. Provider	**		***	*	***	
Level 4	a. Insurer Policy / Provider	**		*	**	**	*
	b. Patient / Provider Group	*	*	*	*	*	
Level 5	Insurer Policy / Provider Group	**		**	**	**	*
Level 6	a. Defined Patient Group	**		*	*	**	**
	b. Provider Group	**		***	**	***	*
Level 7	Multiparty, Criminal Conspiracies	**		**	*	**	***

Usefulness: * Low ** Medium *** High

Big Data Applications in Improving Patient Care

Janaki Mudvari Khatiwada
University of Indiana
Bloomington, Indiana 47408
jmudvari@iu.edu

ABSTRACT

This paper will explore how service providers in health-care industries use data generated when patients provide information about their family history, medical history, food habit, exercise habit.

1 INTRODUCTION

Health service providers collect high volume of information from the consumers every time they visit the facilities. These informations or big data provides helpful insights for diagnostic purpose and treatment options. These data can range from Clinical or pathological category to food and exercise habits, family history or personal body mass index. Clinical practitioners require data to make their medical diagnosis, treatment recommendation, and prognosis. A richer set of near-real-time information can greatly help physicians determine the best course of action for their patients, discover new treatment options, and potentially save lives [**?**]. So to speak fields big data applications in health care for the purpose of improving patient care is wide; disease prevention and management, health education, research and development, prognosis information sharing, public and individual health management, medical optimization.

Health data are stored as electronic medical records(EMR) which are analyzed and shared among clinicians. These data are near real time data. One of the trending example is application of big data in tackling opioid crisis in US. Data scientists at Blue Cross Blue Shield have started working with big data experts at Fuzzy Logix to tackle the problem. Using years of insurance and pharmacy data, Fuzzy Logix analysts have been able to identify 742 risk factors that predict with a high degree of accuracy whether someone is at risk for abusing opioids[**?**].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

Big Data Applications In Population Health Management

Tyler Peterson
Indiana University - School of Informatics, Computing, and Engineering
711 N. Park Avenue
Bloomington, Indiana 47408
typeter@iu.edu

ABSTRACT

My abstract will go here

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

My introduction will go here [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data Analytics, Data Mining, and Public Health Informatics: Using Data Mining of Social Media to Track Epidemics

Sean M. Shiverick
Indiana University-Bloomington
smshiver@indiana.edu

ABSTRACT

Data mining of internet search queries and social media for influenza related keywords has been used to track seasonal influenza and correlates highly with official reports of 'infuenza-like-illness' (ILI). Efforts to monitor epidemics using big data analytics can provide early detection that supplements existing systems of disease surveillance. A review of the literature shows that data extracted from social media has applications for public health informatics. Prediction models based on social media work best in areas with a high degree of internet access.

KEYWORDS

i523, HID335, Data Mining, Social Media, Public Health Informatics

1 INTRODUCTION

In the information age, *Big Data* offers great promise to fuel innovation, generate new revenue streams, and transform society [9]. Can the potential of big data be harnessed for the greater good, to prevent disease and improve health? Seasonal influenza epidemics are a major public health concern, that each year result in an estimated 250,000 to 500,000 deaths worldwide [21]. This paper explores big data in public health informatics, specifically reviewing research on data mining to track epidemics and the spread of contagious disease [10]. Can these approaches be extended to monitor other epidemics such as the opioid crisis in North America? [19] Epidemic spreading is a complex phenomenon based on contact networks between individuals and distributed by transportation networks [4]. Some question remains as to whether prediction models based on social networking platforms can be generalized to other epidemics.

1.1 Public Health Informatics

The field of Health Informatics is generating huge amounts of data at a rapid pace, from MRI imaging data, electronic medical records (EMRs), clinical research data, to population-level data. This review focuses on population data from search queries and social media to provide insights about epidemics and pandemics [10, 11]. Big data is an ambiguous term that lacks a single unified definition, but is often described in terms of *Volume, Velocity, Variety, Veracity, and Value* [5]. Trying to track an epidemic in real-time from multitudes of incoming web searches and posts involves a high volume of data coming in at high velocity [13, 16]. In order to be of any use, diverse and often messy raw data has to be sifted through and effectively organized for further analysis. The issue of Veracity raises the questions of how reliable social media data are for predicting real life events. What is the relationship between social media data to biological events such as the spreading of contagion and disease? The question of Value evaluates the quality of the data as it pertains to intended outcomes. There are legitimate concerns about the quality of data obtained from the internet; however, the literature suggests that mining information from social media can produce valuable data. An important challenge for making sense of big data is developing analytic tools adequate to handle large volumes of data in real time.

1.2 Data Mining Social Media

Health Informatics research is considered from two levels: where the data is collected, and the research questions being addressed. Research on social media can yield data on a range of issues related to public health, including: spatiotemporal information of disease outbreaks, real-time tracking of infectious diseases, global distributions of various diseases, and search queries on medical questions that people might have [11]. The questions of interest in the current review are: *Can search query data be used to accurately track epidemics in real-time?* and, *can Twitter data be used to monitor epidemics across different regions??*. The general idea is that increasing search query or social media activity is associated with an increasing interest in a given health topic. A limitation of social media data is that, although it has high Volume, Velocity, and Variety, it can be unreliable, resulting in both low Veracity and Value [10, 14]. A review of the literature shows how useful data can be extracted by data mining and analytic techniques.

1.3 Using Search Queries to Track Epidemics

1.3.1 Tracking Epidemics Using Google Search Terms in the U.S.. Seasonal influenza is an acute viral infection that spreads easily from person to person, circulates across regions, affecting people of every age. Traditional flu monitoring estimates from the U.S. Center of Disease Control and Prevention (CDC) based on physician reports of patients with "*influenza-like illness*" (ILI) are released weekly [6], but generally with a one to two week delay. In an effort to improve on early detection of season influenza, a team of researchers developed an automated method to analyze Google search queries to track ILI terms from historical logs between 2003 and 2008, using 50 million most popular searches, and CDC historical data [7]. The *Google Flu Trends* (GFT; https://www.google.org/flutrends) model sought to find the probability that a given search query is related to an ILI of a patient visiting a physician in the same region. GFT used a feature selection method to narrow the 50 million most popular search queries down to 45. These top queries showed connections to influenza symptoms, complications, remedies, that were consistent with searches by individuals with influenza. The researchers based their estimates of the current level of weekly influenza based on the correlation of the relative frequency of search queries and the percentage of physician visits with patients presenting influenza-like symptoms. Thus, analyzing high volume Google search queries was used to predict ILI epidemics in real time for areas with a large population of web users, and provide information to the public in a matter of days to help physicians and hospitals prepare and respond to the outbreak.

1.3.2 Tracking H1N1 Epidemic Using Baidu Queries in China. A research study in China monitored influenza activity by comparing internet search query data from *Baidu* (https://www.baidu.com) to influenza case counts from the Chinese Ministry of Health (MOH) between 2009 to 2012 during the H1N1 epidemic [22]. This study consisted of four parts: (i) Selecting keyword terms related to influenza, (ii) Filtering keywords unrelated to flu epidemics, (iii) Defining weights and composite search index, and (iv) Fitting a regression model with keyword index to influenza case data. In the process of filtering, only 40 of 94 keywords were correlated with the case data, and only 8 of these 40 keywords were used as the optimal set in the composite search index. As expected, the search index captured seasonal variation of influenza epidemics in the Winter and Spring, indicating a good predictor for tracking influenza activity in China. The regression model accounted for 95 percent of the variability in influenza case data (ICD), and the model was validated for a test period in 2012. The mean absolute percent error rate was less than 11 percent for the validation test data. This research yields additional evidence that novel approaches using big data can provide early indicators of epidemic activity that supplement official public health information sources, rather than replacing them. A limitation acknowledged by the authors is the relatively small initial number of keyword search terms used compared to the Google Flu Trends (GFT) project [7]. Another limitation of using search query data is that, although the keywords selected in this model performed well at capturing temporal trends in the H1N1 epidemic, the same keywords may not reflect the trend of an influenza epidemic at a future time. The authors also noted the lack of internet access in rural areas, which underscores the fact that effective tracking of epidemics based on search queries relies on internet access. Furthermore, caution should be used when evaluating correlational data, as causation cannot be inferred from correlation.

1.4 Using *Twitter* API Data to Track Epidemics

Twitter is a free online social networking and micro-blogging service, where users can send and read messages of 140 characters (i.e., *"tweets"*). As of 2017, Twitter has more than 320 million monthly active users (67 million in U.S.), with an estimated 500 million tweets posted per day (https://about.twitter.com). Twitter users share their perspectives and reactions on a wide range of topics, approximately 80 percent from handheld mobile devices, acting as "sensors" of events in real time [1]. The Twitter stream provides a rich data source for tracking or forecasting general sentiment, political attitudes, linguistic variation, detecting earthquakes, and disease surveillance. The large volume of users provides a high likelihood that ILI epidemic information is posted; however, Twitter post data is noisy and perhaps unreliable insofar as it can be difficult to differentiate posts about the flu based on instances of concerned awareness (*"I am worried about the swine flu epidemic!"*) versus actual infection, (*"Robbie might have swine flu. I am worried."*)[13]. Despite the noise in Twitter data from much useless chatter, useful information be obtained from mining data in the Twitter stream.

1.4.1 Using Twitter to Track Disease Activity and Public Concern in the U.S. during the H1N1 Pandemic. In a 2011 study, researchers searched through post data from Twitter's streaming API during the H1N1 epidemic (October 2009 to May 2010) across spatiotemporal areas of the U.S. to predict weekly ILI levels [17]. Tweets were sifted according to keywords related to H1N1 (e.g., *"flu", "swine", "influenza"*) and additional terms about vaccines, side effects, and/or vaccine shortages. The first data set consisted of 951,697 tweets containing influenza related keywords from 334,840,972 tweets extracted between April to June 2009 (results were reported as a percentage of observed tweets). These tweets represent just over 1 percent of the sample tweet volume, and this percentage declined rapidly over time as the number of reported H1N1 cases increased. In the U.S. surveillance programs track reported influenza-like illness (ILI) seasonally, from October to May, monitoring the total number of patients seen along with the number with ILIs reported. Quantitative estimates of ILI values based on the Twitter stream were analyzed using support vector regression (SVR) and leave-one-out cross-validation to test model accuracy. Weekly ILI values were estimated using a model trained on roughly 1 million influenza-related tweets obtained between October, 2009 to May 2010. Point estimates of national ILI values produced by the system were good with an average error of 0.28 percent. A regional model, based on significantly fewer tweets, approximated the epidemic curve for CDC region 2 (New York, New Jersey) as reported by the ILI data, but the estimate was less precise with an average error of 0.37 percent. In terms of public interest, Twitter users' interest in antiviral drugs dropped, as official disease reports indicated most influenza cases were relatively mild, even as the number of cases was increasing. In addition, interest in hand hygiene and face masks was associated with public health messages from CDC. A limitation of the study is that only a limited number of search terms and one prediction method was used. An important question is whether the results could be improved using broader search terms and other prediction models.

1.4.2 Twitter Improves Seasonal Influenza Prediction. In a 2012 study, researchers implemented a system using an online social network (OSN) Crawler bot to retrieve tweets by keywords (e.g., *"flu", "H1N1", "swine flu"*), geospatial location, relative keyword frequency , and CDC ILI reports [1]. The *Social Network Enabled Flu Trends* (SNEFT) network continuously monitored tweets and profile details of the Twitter users who commented on flu keywords (starting October 2009), to detect and track the spread of ILI epidemics. The correlation between flu related tweets and ILI was very high between 2009 to 2010 (r=0.98) during the H1N1 outbreak, but the correlation dropped substantially for 2010-2011 (r=0.47) after the epidemic, suggesting that noisy tweets became more prominent as H1N1 was less of an issue. To reduce noise, text classification using support vector machines (*SVMs*) was trained on a dataset of 25,000 tweets to determine whether a tweet was related to a flu event or not; data cleansing was conducted to remove multiple tweets posted by the same user during a single bout with the same illness. These methods improved the correlation between the Twitter data and ILI rates from the CDC from October 2010 to May 2011 in the U.S. (r=0.89), and Twitter data was correlated with ILI rates across subregions. The authors reported that Twitter data alone had higher prediction rates toward the beginning and end of the flu season, and during an epidemic. In addition, age analysis suggested Twitter data best fit the age groups of 5-24 years and 25-49 years, for most regions in the U.S. The results showed Twitter data can be used to detect and possibly predict ongoing ILI epidemics in real time with relatively low error, up to 1-2 weeks earlier than the CDC reportings. It would also be interesting to determine whether these results are generalizable outside the U.S. with different populations in other countries [22].

1.5 *Limitations* of Using Search Queries and Social Media Data to Track Epidemics

The research reviewed above shows how data mining search queries and twitter posts for ILI related information provides an early detection signal to supplement existing epidemic monitoring systems and may help improve public health responses and prevention. There is some evidence that influenza forecasting models based on Twitter data performed better than general search query data [16]; the Google Flu Trends (GFT) algorithms underestimated ILI in the U.S. at the start of the H1N1 (i.e. *swine flu*) pandemic in 2009 [2], and over-predicted seasonal influenza in January 2013 compared to the CDC ILI by almost double [14]. As described above, there are important limitations in using social media data for predicting epidemics: First, internet access and Twitter usage is not uniform by geographical region. Urban areas have higher density of internet connections than rural areas [22], and coastal regions of the U.S. (CA, NY) produced more tweets per person than Midwestern U.S. states (or Europe) [1]. Thus, performance of seasonal influenza predictions models may be limited to regions with high internet access and where tweets are more frequent. Second, exact demographic information about the Twitter population is not easy to estimate (or unknown) and the demographic of internet users does not represent characteristics of the general population. If we consider that outbreaks such as swine flu or avian flu originated at points of contact between humans and domesticated animals in agricultural areas, then internet searches or Twitter posts would provide limited information to predict epidemic spreading in the larger population. Third, though promising, the results of this research are based on correlations between often noisy internet search queries or Twitter posts and physician reports of ILI compiled by official governmental sources. We should be cautious in evaluating predictions about serious health concerns such as epidemics or pandemics based on correlational evidence as the data do not support causal inferences .

2 CONCLUSION

Big data mining of social media has tremendous potential to detect trends and confirm observations based on real time events, providing opportunities to monitor infectious disease on a global level [10]. Can these methods be extended to survey other types of epidemics such as the opioid crisis in North America?[18] Epidemics are described in terms of the proportion of the population infected, those yet to be infected, and the rate of transmission [12]. The dynamics of epidemic spreading is a complex phenomenon, based on contact networks of person-to-person interaction, indirect exposure, and transmission highways such as the *airline transportation network* (ATN) [4]. In addition, the structure of the contact network can influence epidemic spreading [15]. For example, in the case of simple contagion, weak ties among acquaintances or infrequent associations provide shortcuts between distant nodes that reduce distance within the network [8] and can facilitate the spread of disease. Furthermore, networks with "small world" properties have many nodes with few connections, but a small number of highly connected nodes that can rapidly transmit contagion throughout the network [20]. Analyzing the correlation between Twitter posts and rate of ILI reports does not capture the complexity underlying disease epidemics and pandemics. By analyzing the structure of social media networks, future research may help to identify how points of connection online is associated with epidemic spreading in the external world [23]. The emergence of new technologies, such as wearable biosensors [3] may help improve geospatial mapping seasonal influenza and other epidemics. A combination of approaches may prove to be more effective than any individual approach.

REFERENCES

[1] H. Achrekar, A. Gandhe, R. Lazarus, S. H. Yu, and B. Liu. 2012. Twitter improves seasonal influenza prediction.. In *International Conference on Health Informatics.*

[2] D. Butler. 2013. When Google got flu wrong. *Nature* (2013).

[3] Stephanie Carreiro, David Smelson, Megan Ranney, Keith J. Horvath, R. W. Picard, Edwin D. Boudreaux, Rashelle Hayes, and Edward W. Boyer. 2015. Real-Time Mobile Detection of Drug Use with Wearable Biosensors: A Pilot Study. *Journal of Medical Toxicology.* (2015). DOI:http://dx.doi.org/10.1007/s13181-014-0439-7

[4] Vittoria Colizza, Alain Barrat, Marc Barthlemy, and Alessandro Vespignani. 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America* 103, 7 (2006), 2015–2020. DOI:http://dx.doi.org/10.1073/pnas.0510525103 arXiv:http://www.pnas.org/content/103/7/2015.full.pdf

[5] Y. Demchenko, Z. Zhao, P. Grosso, A. Wibisono, and C. De Laat. 2012. Addressing big data challenges for scientific data infrastructure. In *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on.* IEEE, 614–617.

[6] U.S. Centers for Disease Control and Prevention (CDC). 2017. Weekly U.S. Influenza Surveillance Report. https://www.cdc.gov/flu/weekly/index.htm

[7] J. Ginsberg, M.H. Mohebbi, R. S. Patel, L. Brammer, M.S. Smolinski, and L. Brilliant. 2009. Detecting influenza epidemics using search engine query data. *Nature* (2009). http://dx.doi.org/10.1038/nature07634]10.1038/nature07634

[8] M. S. Granovetter. 1973. The strength of weak ties. *Amer. J. Sociology* 78 (1973), 1360fi?!1379.

[9] S. Gupta. 2015. Big Data: Big Deal or Big Hype? *European Business Review* (2015).

[10] Simon I Hay, Dylan B George, Catherine L Moyes, and John S Brownstein. 2013. Big data opportunities for global infectious disease surveillance. *PLoS medicine* 10, 4 (2013), e1001413.

[11] M. Herland, T. M. Khoshgoftaar, and R. Wald. 2014. A review of data mining using big data in health informatics. *Journal Of Big Data* (2014).

[12] Herbert W. Hethcote. 2000. The Mathematics of Infectious Disease. *Society for Industrial and Applied Mathematics (SIAM) Review* 42(4)] (2000), 599fi?!653. DOI: http://dx.doi.org/https://doi.org/10.1137/S0036144500371907

[13] Alex Lamb, Michael J Paul, and Mark Dredze. 2013. Separating Fact from Fear: Tracking Flu Infections on Twitter.. In *HLT-NAACL.* 789–795.

[14] D. Lazer, R. Kennedy, G. King, and A. Vespignani. 2014. The Parable of Google Flu: Traps in Big Data Analysis. *Science* (2014).

[15] Romualdo Pastor-Satorras and Alessandro Vespignani. 2001. Epidemic spreading in scale-free networks. *Physical review letters* 86, 14 (2001), 3200.

[16] M. J. Paul, M. Dredze, and D. Broniatowski. 2014. Twitter Improves Influenza Forecasting. *PLOS Currents: Outbreaks* (2014).

[17] A. Signorini, A. M. Segre, and P. M. Polgreen. 2011. The use of twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLOS ONE* (2011).

[18] Maya Smith. 2016. Can social media help prevent opioid abuse? *Science — News* (2016). http://www.sciencemag.org/news/2016/07/can-social-media-help-prevent-opioid-abuse

[19] N.D. Volkow, T.R. Frieden, P.S. Hyde, and S.S. Cha. 2014. Medication-Assisted Therapies fi!? Tackling the Opioid-Overdose Epidemic. *New England Journal of Medicine* 22 (2014). DOI:http://dx.doi.org/10.1056/NEJMp1402780 PMID:

24758595.

[20] D. J. Watts and S .H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* 393 (1998), 440.

[21] World Health Organization 2016. *Influenza (Seasonal),*. World Health Organization. http://www.who.int/mediacentre/factsheets/fs211/en/

[22] Qingyu Yuan, Elaine O Nsoesie, Benfu Lv, Geng Peng, Rumi Chunara, and John S Brownstein. 2013. Monitoring influenza epidemics in china with search query from baidu. *PloS one* 8, 5 (2013), e64323.

[23] Yu-Xiao Zhu, Wei Wang, Ming Tang, and Yong-Yeol Ahn. 2017. Social contagions on weighted networks. *Phys. Rev. E* 96 (Jul 2017), 012306. Issue 1. DOI: http://dx.doi.org/10.1103/PhysRevE.96.012306

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data and Deep Learning

Jyothi Pranavi Devineni
Indiana University Bloomington
Bloomington, Indiana
jyodevin@umail.iu.edu

ABSTRACT

This paper provides a sample of a LATEX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LATEX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

Distributed Environment For Parallel Neural Networks

Ajinkya Khamkar
Indiana University
Bloomington, Indiana 47408
adkhamka@iu.edu

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size.

REFERENCES

Big Data and Artificial Neural Networks

Bharat Mallala
Indiana University
Smith Research Center
2805 E. 10th St, Suite 150
Bloomington, IN 47408, USA
bmallala@iu.edu

ABSTRACT

This is my abstract.

KEYWORDS

ACM proceedings, LATEX, text tagging

1 INTRODUCTION

This is my Introduction

2 CONCLUSIONS

This is my Conlusion

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for all the help he has provided for this paper.

REFERENCES

My First paper

ZhiCheng Zhu
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221

ABSTRACT

This paper edit by zzc

KEYWORDS

info523 big data

1 INTRODUCTION

this is the introduction

2 THE BODY OF THE PAPER

this is the body of the paper

3 CONCLUSIONS

This is the conclusion

ACKNOWLEDGMENTS

this is the acknow of th para

REFERENCES

My great Big Dat Paper

Shiqi Shen
Indiana University Bloomington
3209 E 10th St
Bloomington, Indiana 47408
trovato@corporation.com

ABSTRACT

This paper

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the [1]

2 THE BODY OF THE PAPER

Typically, the body of a

3 CONCLUSIONS

This paragraph wi [**?**]

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

REFERENCES

[1] Matthew Van Gundy, Davide Balzarotti, and Giovanni Vigna. 2009. Catch me, if you can: Evading network signatures with web-based polymorphic worms. In *Proceedings of the first USENIX workshop on Offensive Technologies (WOOT '09)*. USENIX Association, Berkley, CA, 90–100.

Big Data Application in Web Search and Text Mining

Wenxuan Han
Indiana University Bloomingtonn
1150 S Clarizz Blvd
Bloomington, Indiana 47401-4294
wenxhan@iu.edu

ABSTRACT

Because of the rapid development of social media, there are gigantic amount of data generated in every second on the web. And those data could be stored in any forms like text, videos, images or their combinations. The more complicated forms of data, the more space it will take up and will cost more time to read it. Although most of today's personal computers have a very high performance, it is extremely difficult to process and analyze useful text information from those huge amount of unstructured data by using traditional single computer methods without the help of big data tools or text mining techniques. Fortunately, the improvements in big data application are also increasing fast in order to support those difficult works on web search and text mining. In this paper, we first study the data analytic steps in web search, then analyze some of the popular approaches or algorithms (e.g. Hubs, PageRank, etc), and at last, we discuss their applications in this field of big data.

KEYWORDS

I523, HID209, Big Data, Social Media, Web Search, Text Mining, PageRank, Hubs

1 INTRODUCTION

In recent years, social media has become more and more popular as a new way of communication and knowledge transfer. People could use it to create, share, exchange information and create their own network. Social media usage has been boosted from 2005 to 2015. Users between 18 and 29 ages are the mainly part of social media users [2]. Today 90% of young adults are active on social media. This proportion was 12% in 2005 [1]. And since the development of mobile products, social media has also been offered a better platform for users to share data faster and more convenient. Thus, this proportion could be keep stable or still increase during the next few years.

Nowadays, a growing number of people prefer to express their opinion and feelings through tweeting, sharing images, commenting on social sites [2]. Since the amount of such data become extremely large, it is significant to extract and analyze useful information through them by using text analysis methods. Therefore, some applications which based on these information have been developed, such as recommendation system and search engine. To implement thoses application, web search and text mining technologies play the important role. Text mining aims to find information, meaningful contents, topics, word relations and patterns from the text data.

As the complexity of user searching contents and the amount of data are increasing, short length of search queries may not support the high precision requirement of searching result which cause the one and two-word search engine became slightly less popular. Longer search queries, averaging searches of 5+ words in length, have increased 10 percents comparing January 2009 to January 2008. But as the search queries become longer, we must face the other problems:

2 DATA ANALYTIC STEPS

This part in in in

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Perrin A. 2015. Social Networking Usage: 2005-2015. (Octobe 2015).
[2] Mehmet U. and Secren G. 2016. Text Mining Analysis in Turkish Language Using Big Data Tools. *IEEE Computer Society* (2016).

Using Big Data for Fact Checking

Karthik Vegi
Indiana University
2619 E. 2nd St, Apt 11
Bloomington, IN 47401, USA
kvegi@iu.edu

ABSTRACT

This paper intends to discuss how Big Data can be used to spot fake news, bad data used by politicians, advertisers, and scientists.

KEYWORDS

Big Data, Fact checking

1 INTRODUCTION

Big Data can be used to spot fake news, bad data used by politicians, advertisers, and scientists.

2 CONCLUSIONS

Add a conclusion here

3 REFERENCES

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

ACKNOWLEDGMENTS

I thank all the people who made this possible

REFERENCES

Big Data Applications in Media and Entertainment Industry

Jiaan Wang
Indiana University Bloomington
3209 E 10th St
Bloomington, Indiana 47408
jervwang@indiana.edu

ABSTRACT

This paper demonstrates the growth of big data and its various applications in media and entertainment industry. We showcases the rapid surge of big data and the increasing need for big data technologies. We also describes the problems that come with big data and its challenges in the industry. We then present various utilization of big data and why big data is important in the advancement of media and entertainment industry.

KEYWORDS

Big Data, Media, Entertainment Industry, Technology

1 INTRODUCTION

“2013 is the first year known as the beginning of big data, the world officially enter the era of big data. But big data is not clearly defined, until now, except for large enterprise data also have different definitions, such as Wanda defines the big data as DIKW hierarchical model, that is, Data, Knowledge and wisdom” [7].

“The era of big data is not coming; it is here. The birth and growth of big data was the defining characteristic of the 2000s. As obvious and ordinary as this might sound to us today, we are still unraveling the practical and inspirational potential of this new era. Google processes over 20 petabytes of data a day (a little less than half the entire written works of mankind from the beginning of recorded history in all languages). In addition to collecting and searching for more information, the technologies that allow us to capture and interpret that data are improving every time we blink. Something as simple as a snapshot has become a data collection event” [4].

“Big Data is about the growing challenge that organizations face as they deal with large and fast-growing sources of data or information that also present a complex range of analysis and use problems. Big Data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis” [5].

“IDC, International Data Corporation, believes that organizations that are best able to make real-time business decisions using Big Data streams will thrive, while those that are unable to embrace and make use of this shift will increasingly find themselves at a competitive disadvantage in the market and face potential failure. This will be particularly true in industries experiencing high rates of business change and aggressive consolidation” [5].

“New data sources for Big Data include industries that just recently began to digitize their content. In virtually all of these cases, data growth rates in the past five years have been near infinite, since in most cases it started from zero. The media and entertainment industry moved to digital recording, production, and delivery in the past five years and is now collecting large amounts of rich content and user viewing behaviors” [5].

“The problem with the massive data collection and distribution system we have created is: big data is a big mess. Most of the data we capture in our daily lives just sits around, cluttering up storage space on our devices and slowing down our connections” [4].

“Under the era of big data, the traditional TV media are facing opportunities and challenges, how to deal with challenges and to seize the opportunity is the traditional TV media should concern. Comparison to the Traditional TV media, network TV and new media, the biggest advantage is that the traditional TV media have high-quality TV content, and the strong support of national policy. Traditional TV media itself has a lot of data, but traditional media did not make good use of these data that has been the impact of new media” [7].

2 APPLICATIONS IN MEDIA AND ENTERTAINMENT INDUSTRY

“Social media solutions such as Facebook, Foursquare, and Twitter are the newest new data sources. A number of new businesses are now building Big Data environments, based on scale-out clusters using power-efficient multicore processors like the AMD Opteron 4000 and 6000 Series platforms, that leverage consumers’ (conscious or unconscious) nearly continuous streams of data about themselves (e.g., likes, locations, opinions). Thanks to the ”network effect” of successful sites, the total data generated can expand at an exponential rate. One company IDC spoke with collected and analyzed over 4 billion data points (Web site cut-and-paste operations) in its first year of operation and is approaching 20 billion data points less than a year later” [5].

“Some of the most interesting, but also most challenged, industries when it comes to Big Data adoption will be utilities and content service providers (e.g., cable TV, mobile carriers). These communities (with assists from related companies such as video gaming system and appliance manufacturers) are building out Big Data generating fabrics. Their opportunity now is to figure out how to handle and then do something with all that data, despite the fact that from a cultural standpoint data guardianship and use were much less in the past” [5].

“An additional hurdle for these industries is that it isn’t enough to just get the ”answers” from Big Data platforms. They also need to implement automated response systems (e.g., automated power management or ”in game” ad placement) that will ultimately be the foundation of their business models” [5].

“We would like to offer a set of rules for the new data world: 1) big is not enough, and 2) it is neither necessary nor practical to fix every piece of data we have collected as a species into some particular order" [4].

"We are already capturing massive quantities of data about our entertainment. Take, for example, Supernatural, an American horror series, created by Eric Kripke in 2005.1 Now in its seventh season, it has generated roughly 112 hours of footage. So we have a lot of pixels, yes, but we also have much more. We have every action of every character; every line of dialogue; a history of when, where, and how often everyone dies. Because all of that information is data, what we actually have, in and around those 112 hours of pixels, is a map to the world of Supernatural, and the characters inside it" [4].

"Today, all of that footage and all of that information is locked away in old style data collections: fixed and unwieldy. But if we can store all that information in a system, modeled more on biology than books, and apply our significant and increasing processing power to analyze and respond to the world, rather than just move it around mechanically, then we have the possibility of generating and interacting with the world and the characters of Supernatural (or possibly even a story you like). This requires computational intelligence, not a Google search. It is not the ability to hunt down a single piece of data in the massive haystack of global information but rather the ability to make something new and interesting emerge out of that data" [4].

"In the era of big data, mass user behavior data is used to model predictions. Where big data are the personal recommendation system in a typical application of radio and television, The traditional approach is based on the user's clicking behavior, to analyze the user's preferences, then recommend related programs. But now in order to recommend more accurate, use not just the set-top box data for statistical analysis, but also dig out the sharing behavior on the user network along with the comment feature behavior and other behaviors, in order to better characterize user portrait. In the era of big data, television media should be the depth of excavation and analysis of user information on the user's viewing behavior , the initiative to understand what users really want to see, in order to provide better services for television users. In other countries, the television media successful application of large data typical case is the "house of cards", which analyzes the form of selection and decision-making with actors play using the big data" [7].

"Technically, the first to take in consideration is television media are capable of producing large amounts of data every day, how to integrate their data, define combing their data assets to create a connection between the television media and their users, effective analysis of audience preferences to realize customization. secondly the traditional TV media have with respect to network operators, the biggest advantage is that they have high-quality TV content, but how to use these high quality content effectively disseminated to users. in addition to drawing telecommunications powerful content communication technologies and outside network framework, also taking into account the characteristics of the television media itself" [7].

"The most important point that TV media can use big data technology is that television is the media itself has data, the data are the main source of set-top boxes, network management systems. To collect the data more widely, some companies such as Nielsen TV media can also take the technology to provide brain waves, using 32 sensors, acquisition frequency of 500 times / Sec, measurable indicators are mainly about emotional investment, triggering memories and attention. Therefore, data collection is more mature as it showed" [7].

"But the TV media data from multiple data sources and scattered, besides the internal data such as set-top box data, network management systems data, BOSS system, etc., as well as external data, such as online user behavior data, data integration is the primary challenge in television media big data applications. How TV media make internal data and external data streams to achieve mutual exchange, how to create their own big data, sort out their own data assets, which need the support by big data technology. And television media use Big Data technologies to meet the individual needs of the "precise communication", which can improve service quality, protection of cultural rights and interests of the public, the media TV plays disseminating information, building culture, guide public opinion, the responsibility to resist foreign cultural erosion at the same time, therefore more need to focus on high-tech applications" [7].

3 CONCLUSION

Put here an conclusion. Conclusions and abstracts must not have any citations in the section. [1] [2] [3] [6]

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper as well as Lee Yang for her proofreading on this paper.

REFERENCES

[1] Richard Abel. 2013. The Pleasures and Perils of Big Data in Digitized Newspapers. *Film History* 25, 1-2 (2013), 1–10. https://doi.org/10.2979/filmhistory.25.1-2.1 HID: 233, Accessed: 2017-09-28.

[2] Kenneth Cukier and Viktor Mayer-Schoenberger. 2013. The Rise of Big Data: How It's Changing the Way We Think About the World. *Foreign Affairs* 92, 3 (May 2013), 28–40. https://www-jstor-org.proxyiub.uits.iu.edu/stable/pdf/23526834.pdf?refreqid=excelsior%3Aab262e886bf5da6437092dbbcebb14de HID: 233, Accessed: 2017-09-28.

[3] Frank McCoy. 2013. BIG DATA Equals BIG BUCKS for STEM Grads: BUT TO REALLY SUCCEED, LEARN HOW TO COMMUNICATE. *US Black Engineer and Information Technology* 37, 3 (2013), 17–19. https://www-jstor-org.proxyiub.uits.iu.edu/stable/pdf/43772972.pdf?refreqid=excelsior%3A4ab0dc0263a5a0e302c511b9c34c7a7e HID: 233, Accessed: 2017-09-28.

[4] Tawny Schlieski and Brian David Johnson. 2012. Entertainment in the Age of Big Data. *Proc. IEEE* 100, Special Centennial Issue (May 2012), 1404–1408. https://doi.org/10.1109/JPROC.2012.2189918 HID: 233, Accessed: 2017-09-20.

[5] Richard L. Villars, Carl W. Olofson, and Matthew Eastwood. 2011. Big data: What it is and why you should care. *White Paper, IDC* 14 (June 2011). www.tracemyflows.com/uploads/big_data/idc_amd_big_data_whitepaper.pdf HID: 233, Accessed: 2017-09-20.

[6] Patrick J. Wolfe. 2013. Making sense of big data. *Proceedings of the National Academy of Sciences of the United States of America* 110, 45 (Nov. 2013), 18031–18032. https://www-jstor-org.proxyiub.uits.iu.edu/stable/pdf/23754685.pdf?refreqid=excelsior%3A1e139d650259ac78c3c111865d39492c HID: 233, Accessed: 2017-09-28.

[7] Chunjie Zhang, Wenqian Shang, Weiguo Lin, Yongan Li, and Rui Tan. 2017. Opportunities and challenges of TV media in the big data era. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. IEEE, Wuhan, China, 551–553. https://doi.org/10.1109/ICIS.2017.7960053 HID: 233, Accessed: 2017-09-20.

Big Data Analytics: Recommendation Systems on the Web

Jordan Simmons
Indiana University Bloomington
jomsimm@iu.edu

ABSTRACT

This paper is an overview of Recommendation Systems being used on the web. It will go over some popular techniques that are being used in modern systems. It will briefly discuss a couple state of the art systems. Then if will finally touch on some of the limitations and challenges that there are to overvome in the field.

KEYWORDS

i523, hid336, Recommendation Systems, Big Data

1 INTRODUCTION

Recommendation systems (RS) leverage big data in ways that create value for both businesses and customers."The goal of a recommender system is to generate meaningful recommendations to a collection of users for items or products that might interest them" [6]. In this sense, an item can range from a product in a store, a news article on a site, or even a search query. RS is beneficial to businesses and customers by increasing metrics such as revenue and customer satisfaction [2]. Many online platforms are starting to use RS to analyze their data. General analysis of modern techniques, companies currently using RS, and challenges and limitations will give a better understanding of RS.

2 RECOMMENDATION TECHNIQUES

Three common RS techniques would include content-based, collaborative, and hybrid recommendations [1]. Other techniques exist, but these three are the most widely used today. The best technique depends on what recommendations need to be made, and the data used to make them. Many times, the hybrid approach is used because there can be limitations with other approaches [1]. It is best to understand a little bit about each technique before choosing which is best.

2.1 Content-Based

Content-Based RS recommend items to users by using descriptions of items and how the user is profiled based on their interest [7]. Items are classified by different characteristics,attributes, or variables [7]. Once items are classified, they can be grouped together based on the classifications. Users are classified by information they provide to the system, and/or data collected by interacting with the system.

Content-Based RS are commonly seen on web applications and E-commerce sites. These types of systems can easily track and monitor almost all user activities. Usually a user has an account with the system, where information was voluntarily provided. With this data, users can be classified easier compared to a customer walking into a brick and mortar business.

2.2 Collaborative Filtering

"Collaborative Filtering is the process of filtering or evaluating items using the opinions of other people" [9]. This type of RS is commonly seen on systems where an item can be rated by a user. With this technique, user rating are collected and store from a user for an item that they have used or purchased. The ratings from users are then compared to other users that have rated the same item. For example, person A buys items 1 and 2 and rates each item highly. Then, person B buys item 1 and rates it highly. Since person A and B both bought and rated item 1 highly, the system would likely recommend item 2 to person B. On the contrary, if person B gave item 1 a low rating, the system would not likely recommend item 2 to person B. This concept uses the assumption that "people with similar tastes will rate things similarly" [9]. This assumption may not be true in all cases, but it is a good base for RS to start learning users interest, and recommend items based on those interest. With this technique, the more ratings that the systems has collected per item, and the more ratings given by the user, the easier it is for that system to make recommendations to that specific user.

2.3 Hybrid

Hybrid RS takes two or more techniques and combines them to improve performance and reduce limitations that a single technique might have [3]. In most cases, collaborative filtering is used with one or more of the other techniques to improve performance. Other techniques that are used and not discussed include Demographic, Utility-Based, and Knowledge-based recommendations [3].The hybrid approach narrows down items with one technique, and then uses another technique on that subset of items to make a more accurate recommendation. The best hybrid system really depends on the specific business case, and the data being used to make the recommendation. In some cases, a set of techniques may produce better recommendations than any of the other set of techniques.

An example of a hybrid approach would use collaborative filtering and the content-based methods described above. Say that User A is interested in baseball. The system would use the content-based approach to narrow down all items that are classified as baseball items. From this subset of baseball items, the system could then use the collaborative-filtering approach to find the items with ratings from other users which will be user group B. The system would then find all item ratings from user group B and compare those item ratings to person A. If there are any users in group B that have similar likes to person A, the system would likely recommend the baseball items to person A that person B has previously rated highly. This is a generic example of how a hybrid RS would work. Real world examples are more complex than this example, and use large amounts of data.

3 MODERN SYSTEMS

Two well known companies that are currently using RS are Netflix and Amazon. These two companies have huge customer bases, in which they collect data on. The data is what drives their state of the are RS to make predictions to their users, and they are doing this very well.

3.1 Netflix

Netflix is an internet based company that offers a variety of movies and television shows. Netflix had a problem of customers sorting through its large selection of movies and shows, and eventually losing interest which resulted in abandonment of their services [5]. Over the years, Netflix has created and continually developed new RS algorithms which they claim saves them more than one billion dollars per year and a monthly turnover in the low double digits [5].

Netflix does very well at recommending movies and shows to its users. They have incorporated different strategies to collect data from users whcih is the base of their RS. Data is collected in the form of customized search, video ratings, continue watching feature, amount of time spent watching and other user activities [5]. From the data collected from these features, Netflix can recommend top rated, now trending, and videos based on user interest, which is very appealing to the user when there are so many selections to choose from.

3.2 Amazon

Amazon is an online store that sell a large variety of products. Amazons RS provides recommendations for millions of customers from a catalog that has millions of products. [10]. Instead of comparing customers to customers, amazon uses an item-based collaborative filtering approach. This process finds items that were bought together with unusually high frequencies, and uses these relationships to recommend products to customers based on what they have purchased in the past [10]. With this algorithm, Amazon is providing a unique experience to every user and helping them find products they may not have found. Since the initial launch of this algorithm, it has ”been tweaked to help people find videos to watch or news to read, been challenged by other algorithms and other techniques, and been adapted to improve diversity and discovery, recency, time-sensitive or sequential items, and many other problems. ” [10]

4 CHALLENGES AND LIMITATIONS

As with most technologies, RS has its challenges and limitations. It is hard to speak of this topic without speaking about the questions ”more data usually beats better algorithms” [8]. This quote has raised controversy to which of the two actually produce better results. In most cases, there are many diferent variables to consider when answering this question.

4.1 Limitations

With complex systems, there can be many variables that cause issues that limit full capabilities of that system. Specifically, in RS, some of these limitations include cold start problems,data sparsity, limited content analysis, and latency problems [**?**]. These limitations seem to be more data related rather than the actual techniques and approaches of the technology being used to analyze that data. When there is no data for a new user, it is hard for RS to recommend anything to this user. The system has no data on the users activities or what interests that user has. When a new item is added to a system, there are no reviews and no data collected with the interaction of user for this particular item. On the other hand, too much data can become redundant. At this point gathering more data will have limited gains.

4.2 Cross-Domain Recommendations

Cross-Domain recommendations aim to ”leverage all the available user data provided in various systems and domains, in order to generate more encompassing user models and better recommendations” [4]. Every day the amount of data being collected increases. This data is being collected from different sources. Cross-Domain RS could use data from different sources to perhaps makes up for some of the data caused problems. An example of a Cross-Domain recommendation would be Netflix using data from facebook to help recommend movies to a new user. Using data from various systems like this would bring up new issues like privacy and security, but if systems started working together and sharing data there could be benefits for both systems.

Cross-Domain Recommendations help with domain specific data issues. Two different systems may have different ways of collecting and organizing data. If system 1 collects variables A ,B and C, and system 2 collects variables A, B, and D, each system has information that the other system does not have. This is where sharing the data between systems could have benefits for both systems. In doing this, each system is not only benefiting from more data, but different and perhaps better data. This would also require using better algorithms to analyze the different sets of data. Depending on the system, more data can be more beneficial than better algorithms. In terms of scaleability, gathering more data that is different from what is currently being collected, and using better algorithms along with the different data could potentially maximize recommendations for that system.

5 CONCLUSION

With a base understanding of RS, it is easy to see how this technology can be very beneficial in online platforms. RS has different techniques that can be used in a variety of online systems. Many large companies are creating custom RS and are benefiting greatly from them. As the massive amount of data grows from day to day, the ways in which RS is used will continue to evolve. It will be interesting to see how Cross-Domain Recommendations are used in the future, and if companies start to adopt this concept of sharing data. Data being analyzed from various systems could unlock hidden information that a single system may not be capable of producing.

ACKNOWLEDGMENTS

The author would like to thank course instructors for organizing setup of the latex format used in this paper.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Trans. on Knowl. and Data Eng.* 17, 6 (June 2005), 734–749. https://doi.org/10.1109/TKDE.2005.99

[2] Xavier Amatriain and Justin Basilico. 2016. Past, Present, and Future of Recommender Systems: An Industry Perspective. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 211–214. https://doi.org/10.1145/2959100.2959144

[3] Robin Burke. 2002. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction* 12, 4 (01 Nov 2002), 331–370. https://doi.org/10.1023/A:1021240730564

[4] Iván Cantador, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Paolo Cremonesi. 2015. *Cross-Domain Recommender Systems*. Springer US, Boston, MA, 919–959. https://doi.org/10.1007/978-1-4899-7637-6_27

[5] Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Trans. Manage. Inf. Syst.* 6, 4, Article 13 (Dec. 2015), 19 pages. https://doi.org/10.1145/2843948

[6] Prem Melville and Vikas Sindhwani. 2010. *Recommender Systems*. Springer US, Boston, MA, 829–838. https://doi.org/10.1007/978-0-387-30164-8_705

[7] Michael J. Pazzani and Daniel Billsus. 2007. *Content-Based Recommendation Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 325–341. https://doi.org/10.1007/978-3-540-72079-9_10

[8] Anand Rajaraman. 2008. More Data Usually Beats Better Algorithms. (03 2008). http://anand.typepad.com/datawocky/2008/03/more-data-usual.html

[9] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. 2007. *Collaborative Filtering Recommender Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 291–324. https://doi.org/10.1007/978-3-540-72079-9_9

[10] Brent Smith and Greg Linden. 2017. Two Decades of Recommender Systems at Amazon.com. *IEEE Internet Computing* 21, 3 (2017), 12–18. https://doi.org/doi.ieeecomputersociety.org/10.1109/MIC.2017.72

Big Data Analytics for Research Libraries and Archives

Timothy A. Thompson
Indiana University Bloomington
School of Informatics, Computing, and Engineering
Bloomington, Indiana 47408
timathom@indiana.edu

ABSTRACT

Research libraries and archives have played a longstanding role in information management and access. In the second half of the twentieth century, libraries were at the forefront of automation and networked access to information. Since the advent of the internet, however, they have failed to keep pace with technological advances and now face serious challenges in serving the evolving needs of researchers, which are increasingly focused on solutions for preserving and processing large amounts of data. To remain relevant in the current information landscape, libraries and archives must implement new strategies for converting legacy data to formats that can add value to the research lifecycle.

KEYWORDS

Libraries, Archives, Data Management, Data Integration, ETL

1 INTRODUCTION

Examples of big data analytics in research libraries and archives are still scarce. In the library domain, the leading data hub is the Online Computer Library Center (OCLC)[1].

2 CONCLUSION

Conclusions and abstracts must not have any citations in the section.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions in writing this paper.

REFERENCES

[1] M. Teets and M. Goldner. 2013. Libraries' Role in Curating and Exposing Big Data. *Future Internet* 5 (2013), 429–438. https://doi.org/10.3390/fi5030429

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Lars Thørväld
The Thørväld Group
1 Thørväld Circle
Hekla, Iceland
larst@affiliation.org

Valerie Béranger
Inria Paris-Rocquencourt
Rocquencourt, France

Aparna Patel
Rajiv Gandhi University
Rono-Hills
Doimukh, Arunachal Pradesh, India

Huifen Chan
Tsinghua University
30 Shuangqing Rd
Haidian Qu, Beijing Shi, China

Charles Palmer
Palmer Research Laboratories
8600 Datapoint Drive
San Antonio, Texas 78229
cpalmer@prl.com

John Smith
The Thørväld Group
jsmith@affiliation.org

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

2 CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

A HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are different in the appendices. In the **appendix** environment, the command **section** is used to indicate the start of each Appendix, with alphabetic order designation (i.e., the first is A, the second B, etc.) and a title (if you include one). So, if you need hierarchical structure *within* an Appendix, start with **subsection** as the highest level. Here is an outline of the body of this document in Appendix-appropriate form:

A.1 Introduction

A.2 The Body of the Paper

A.2.1 Type Changes and Special Characters.

A.2.2 Math Equations.

Inline (In-text) Equations.

Display Equations.

A.2.3 Citations.

A.2.4 Tables.

A.2.5 Figures.

A.2.6 Theorem-like Constructs.

A Caveat for the TeX Expert.

A.3 Conclusions

A.4 References

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

B MORE HELP FOR THE HARDY

Of course, reading the source code is always useful. The file `acmart.pdf` contains both the user guide and the commented code.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Lars Thørväld
The Thørväld Group
1 Thørväld Circle
Hekla, Iceland
larst@affiliation.org

Valerie Béranger
Inria Paris-Rocquencourt
Rocquencourt, France

Aparna Patel
Rajiv Gandhi University
Rono-Hills
Doimukh, Arunachal Pradesh, India

Huifen Chan
Tsinghua University
30 Shuangqing Rd
Haidian Qu, Beijing Shi, China

Charles Palmer
Palmer Research Laboratories
8600 Datapoint Drive
San Antonio, Texas 78229
cpalmer@prl.com

John Smith
The Thørväld Group
jsmith@affiliation.org

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data Analytics in Biometric Identity Management

Robert W. Gasiewicz
Indiana University
711 N. Park Avenue
Bloomington, IN 47408
rgasiewi@iu.edu

ABSTRACT

An understanding how biometrics is changing rapidly, and with it, both the size and scope of data being collected. From 2-print to 10-print, iris to facial recognition, the demand for both data intensive processes and rapid matching have grown exponentially, a case study in big data if there ever was one.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data and Artificial Intelligence Solutions for in Home, Community and Territory Security

Ashok Reddy Singam
Indiana University
711 N Park Ave
Bloomington, Indiana 47408
asingam@iu.edu

Anil Ravi
Indiana University
711 N Park Ave
Bloomington, Indiana 47408
anilravi@iu.edu

ABSTRACT

Having an intelligent ear-and-eye monitoring at the home to constantly observe the surroundings both inside and outside can protect the house and personnel much more safer way. By extending this capability to the neighborhood and city through collaboration would create safe cities across the world.Smart Security systems equipped with Video and Audio sensors produce huge amounts unstructured data continuously. This paper talks about high level architecture of an intelligent security system with video surveillance, audio monitoring/recording, video and Audio analytics, alerting homeowners/authorities/agencies as needed. Big Data analytics becomes critical in supporting these modern applications.

The key differentiating capability of this system is to use micro drone with voice and video sensors to process the audio and video data with machine learning algorithms.

KEYWORDS

i523,hid337,hid333,Big Data,AI,Deep Learning

1 INTRODUCTION

In the today's technology world, drones are becoming more familiar in the main stream life activities such as recreational, spy cameras by authorities, home delivery of goods etc. The present security systems used by households are static cameras used at a fixed location inside or outside the house. They are connected to network and provide alerts when any event occurred. However, they are not intelligent enough to understand the context, recognizing the people faces, and aware of family members behaviors, house needs etc. By making cameras movable across the house and outside and process the data as humans do and take decisions of alerting or informing to the right people is the key concept of this paper.This system to be functional, the following technologies needed:

Micro drones with audio and video sensors

Facial recognition and mapping through video analytics to recognize people

Natural language processing (NLP) to recognize family members, friends and strangers

Machine learning algorithms to understand family members habits and behaviors

Interfacing with email servers, phone, text message servers

2 BIG DATA:VIDEO ANALYTICS

Video analytics plays a key role in modernizing video surveillance systems.Deep Learning the fastest growing field of Artificial Intelligence, enabling computers to interpret huge amounts of data like videos.The Graphics Processing Units (GPUs) provided by vendors like Nvidia enabling the deep learning infrastructure to cameras and recorders.

2.1 Big Data:Audio Analytics

We have already seen several typeface changes in this sample. You can indicate italicized words or phrases in your text with the command `\textit`; emboldening with the command `\textbf` and typewriter-style (for instance, for computer code) with `\texttt`. But remember, you do not have to indicate typestyle changes when such changes are part of the *structural* elements of your article; for instance, the heading of this subsection will be in a sans serif[1] typeface, but that is handled by the document class file. Take care with the use of the curly braces in typeface changes; they mark the beginning and end of the text that is to be in the different typeface.

You can use whatever symbols, accented characters, or non-English characters you need anywhere in your document; you can find a complete list of what is available in the *LaTeX User's Guide* [**?**].

2.2 Yet to define

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections.

2.2.1 Inline (In-text) Equations. A formula that appears in the running text is called an inline or in-text formula. It is produced by the **math** environment, which can be invoked with the usual `\begin . . . \end` construction or with the short form `$ . . . $`. You can use any of the symbols and structures, from α to ω, available in LaTeX [**?**]; this section will simply show a few examples of in-text equations in context. Notice how this equation:

$\lim_{n\rightarrow \infty} x = 0$,

set here in in-line math style, looks slightly different when set in display style. (See next section).

2.2.2 Display Equations. A numbered display equation—one set off by vertical space from the text and centered horizontally—is produced by the **equation** environment. An unnumbered display equation is produced by the **displaymath** environment.

Again, in either environment, you can use any of the symbols and structures available in LaTeX; this section will just give a couple of examples of display equations in context. First, consider the equation, shown as an inline equation above:

[1] Another footnote here. Let's make this a rather long one to see how it looks. Footnotes must be avoided.

2.3 Citations

Citations to articles [? ? ? ?], conference proceedings [?] or maybe books [? ?] listed in the Bibliography section of your article will occur throughout the text of your article. You should use BibTeX to automatically produce this bibliography; you simply need to insert one of several citation commands with a key of the item cited in the proper location in the `.tex` file [?]. The key is a short reference you invent to uniquely identify each work; in this sample document, the key is the first author's surname and a word from the title. This identifying key is included with each item in the `.bib` file for your article.

The details of the construction of the `.bib` file are beyond the scope of this sample document, but more information can be found in the *Author's Guide*, and exhaustive details in the *LATEX User's Guide* by Lamport [?].

This article shows only the plainest form of the citation command, using `\cite`.

Some examples. A paginated journal article [?], an enumerated journal article [?], a reference to an entire issue [?], a monograph (whole book) [?], a monograph/whole book in a series (see 2a in spec. document) [?], a divisible-book such as an anthology or compilation [?] followed by the same example, however we only output the series if the volume number is given [?] (so Editor00a's series should NOT be present since it has no vol. no.), a chapter in a divisible book [?], a chapter in a divisible book in a series [?], a multi-volume work as book [?], an article in a proceedings (of a conference, symposium, workshop for example) (paginated proceedings article) [?], a proceedings article with all possible elements [?], an example of an enumerated proceedings article [?], an informally published work [?], a doctoral dissertation [?], a master's thesis: [?], an online document / world wide web resource [? ? ?], a video game (Case 1) [?] and (Case 2) [?] and [?] and (Case 3) a patent [?], work accepted for publication [?], 'YYYYb'-test for prolific author [?] and [?]. Other cites might contain 'duplicate' DOI and URLs (some SIAM articles) [?]. Boris / Barbara Beeton: multi-volume works as books [?] and [?].

A couple of citations with DOIs: [? ?].

Online citations: [? ? ?].

We use jabref to manage all citations. A paper without managing a bib file will be returned without review. in the bibtex file all urls are added to rfernces with the *url* filed. They are not to be included in the *howpublished* or *note* field.

3 CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LATEX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

Big Data Analytics in Sports - Track and Field

Mathew Millard
Indiana University Bloomington
938 N Walnut St. Apt. G
Bloomington, Indiana 47404
mdmillar@indiana.edu

ABSTRACT

This paper covers the impact that Big Data has and could have on the sport of track and field.

KEYWORDS

i523

1 INTRODUCTION

This is my introduction

2 THE BODY OF THE PAPER

This is the body of my paper

3 CONCLUSIONS

This is my conclusion

ACKNOWLEDGMENTS

Acknowledgments

REFERENCES

Big Data Analytics in Sports - Soccer

Rahul Velayutham
Indiana University Bloomington
2661 E 7th Street Apt H
Bloomington, Indiana 47408
rahul.vela@gmail.com.com

ABSTRACT

The aim of this paper is to provide an understanding as to how big data is playing a huge role in Football clubs helping them scout players.

KEYWORDS

Big Data, Soccer , Scouting

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size.

REFERENCES

Big Data in NCAA Football

Nsikan Udoyen
School of Informatics and Computing, Indiana University
P.O. Box 1212
Dublin, Indiana 43017-6221
nudoyen@iu.edu

ABSTRACT

This paper provides an overview of applications of big data in NCAA football.

KEYWORDS

i523

1 INTRODUCTION

National Collegiate Athletics Association (NCAA) football is one of the most widely watched sports in the United States. The size of the fan base and the profits that can be derived from televised games incentivizes universities and other interested parties to invest in the application of big data analytics and data science methods in general to improve on-field outcomes by enabling better management of player well-being and performance. The purpose of this paper is to provide an overview of the use of data science in National Collegiate Athletics Association (NCAA) football. Recent research on the use of data science to improve various aspects of NCAA football will be surveyed, while current trends and their implications will be discussed.

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

Big Data Analytics using Spark

Nisha Chandwani
Indiana University Bloomington
107 S Indiana Ave
Bloomington, Indiana 47405
nchandwa@iu.edu

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size.

2 THE BODY OF THE PAPER

Typically, the body of a paper is organized into a hierarchical structure, with numbered or unnumbered headings for sections, subsections, sub-subsections, and even smaller sections. The command `\section` that precedes this paragraph is part of such a hierarchy. LaTeX handles the numbering and placement of these headings for you, when you use the appropriate heading commands around the titles of the headings. If you want a sub-subsection or smaller part to be unnumbered in your output, simply append an asterisk to the command name. Examples of both numbered and unnumbered headings will appear throughout the balance of this sample document.

Because the entire article is contained in the **document** environment, you can indicate the start of a new paragraph with a blank line in your input file; that is why this sentence forms a separate paragraph.

2.1 Type Changes and *Special* Characters

We have already seen several typeface changes in this sample. You can indicate italicized words or phrases in your text with the command `\textit`; emboldening with the command `\textbf` and typewriter-style (for instance, for computer code) with `\texttt`. But remember, you do not have to indicate typestyle changes when such changes are part of the *structural* elements of your article; for instance, the heading of this subsection will be in a sans serif[1] typeface, but that is handled by the document class file. Take care with the use of the curly braces in typeface changes; they mark the beginning and end of the text that is to be in the different typeface.

You can use whatever symbols, accented characters, or non-English characters you need anywhere in your document; you can find a complete list of what is available in the *LaTeX User's Guide* [26].

2.2 Math Equations

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections.

2.2.1 Inline (In-text) Equations. A formula that appears in the running text is called an inline or in-text formula. It is produced by the **math** environment, which can be invoked with the usual `\begin . . . \end` construction or with the short form `$ . . . $`. You can use any of the symbols and structures, from α to ω, available in LaTeX [26]; this section will simply show a few examples of in-text equations in context. Notice how this equation:

$\lim_{n\to\infty} x = 0$,

set here in in-line math style, looks slightly different when set in display style. (See next section).

2.2.2 Display Equations. A numbered display equation—one set off by vertical space from the text and centered horizontally—is produced by the **equation** environment. An unnumbered display equation is produced by the **displaymath** environment.

Again, in either environment, you can use any of the symbols and structures available in LaTeX; this section will just give a couple of examples of display equations in context. First, consider the equation, shown as an inline equation above:

$$\lim_{n\to\infty} x = 0 \tag{1}$$

Notice how it is formatted somewhat differently in the **displaymath** environment. Now, we'll enter an unnumbered equation:

$$\sum_{i=0}^{\infty} x + 1$$

and follow it with another numbered equation:

$$\sum_{i=0}^{\infty} x_i = \int_0^{\pi+2} f \tag{2}$$

just to demonstrate LaTeX's able handling of numbering.

2.3 Citations

Citations to articles [6–8, 19], conference proceedings [8] or maybe books [26, 34] listed in the Bibliography section of your article will

[1] Another footnote here. Let's make this a rather long one to see how it looks. Footnotes must be avoided.

occur throughout the text of your article. You should use BibTeX to automatically produce this bibliography; you simply need to insert one of several citation commands with a key of the item cited in the proper location in the `.tex` file [26]. The key is a short reference you invent to uniquely identify each work; in this sample document, the key is the first author's surname and a word from the title. This identifying key is included with each item in the `.bib` file for your article.

The details of the construction of the `.bib` file are beyond the scope of this sample document, but more information can be found in the *Author's Guide*, and exhaustive details in the *LaTeX User's Guide* by Lamport [26].

This article shows only the plainest form of the citation command, using `\cite`.

Some examples. A paginated journal article [2], an enumerated journal article [11], a reference to an entire issue [10], a monograph (whole book) [25], a monograph/whole book in a series (see 2a in spec. document) [18], a divisible-book such as an anthology or compilation [13] followed by the same example, however we only output the series if the volume number is given [14] (so Editor00a's series should NOT be present since it has no vol. no.), a chapter in a divisible book [37], a chapter in a divisible book in a series [12], a multi-volume work as book [24], an article in a proceedings (of a conference, symposium, workshop for example) (paginated proceedings article) [4], a proceedings article with all possible elements [36], an example of an enumerated proceedings article [16], an informally published work [17], a doctoral dissertation [9], a master's thesis: [5], an online document / world wide web resource [1, 30, 38], a video game (Case 1) [29] and (Case 2) [28] and [27] and (Case 3) a patent [35], work accepted for publication [31], 'YYYYb'-test for prolific author [32] and [33]. Other cites might contain 'duplicate' DOI and URLs (some SIAM articles) [23]. Boris / Barbara Beeton: multi-volume works as books [21] and [20].

A couple of citations with DOIs: [22, 23].

Online citations: [38–40].

We use jabref to manage all citations. A paper without managing a bib file will be returned without review. in the bibtex file all urls are added to rfernces with the *url* filed. They are not to be included in the *howpublished* or *note* field.

2.4 Tables

Because tables cannot be split across pages, the best placement for them is typically the top of the page nearest their initial cite. To ensure this proper "floating" placement of tables, use the environment **table** to enclose the table's contents and the table caption. The contents of the table itself must go in the **tabular** environment, to be aligned properly in rows and columns, with the desired horizontal and vertical rules. Again, detailed instructions on **tabular** material are found in the *LaTeX User's Guide*.

Immediately following this sentence is the point at which Table 1 is included in the input file; compare the placement of the table here with the table in the printed output of this document.

[Table 1 about here.]

To set a wider table, which takes up the whole width of the page's live area, use the environment **table*** to enclose the table's contents and the table caption. As with a single-column table, this wide table will "float" to a location deemed more desirable. Immediately following this sentence is the point at which Table 2 is included in the input file; again, it is instructive to compare the placement of the table here with the table in the printed output of this document.

[Table 2 about here.]

It is strongly recommended to use the package booktabs [15] and follow its main principles of typography with respect to tables:

(1) Never, ever use vertical rules.
(2) Never use double rules.

It is also a good idea not to overuse horizontal rules.

2.5 Figures

Like tables, figures cannot be split across pages; the best placement for them is typically the top or the bottom of the page nearest their initial cite. To ensure this proper "floating" placement of figures, use the environment **figure** to enclose the figure and its caption.

This sample document contains examples of `.eps` files to be displayable with LaTeX. If you work with pdfLaTeX, use files in the `.pdf` format. Note that most modern TeX systems will convert `.eps` to `.pdf` for you on the fly. More details on each of these are found in the *Author's Guide*.

[Figure 1 about here.]

[Figure 2 about here.]

As was the case with tables, you may want a figure that spans two columns. To do this, and still to ensure proper "floating" placement of tables, use the environment **figure*** to enclose the figure and its caption. And don't forget to end the environment with **figure***, not **figure**!

[Figure 3 about here.]

[Figure 4 about here.]

2.6 Theorem-like Constructs

Other common constructs that may occur in your article are the forms for logical constructs like theorems, axioms, corollaries and proofs. ACM uses two types of these constructs: theorem-like and definition-like.

Here is a theorem:

THEOREM 2.1. *Let f be continuous on $[a, b]$. If G is an antiderivative for f on $[a, b]$, then*

$$\int_a^b f(t)\,dt = G(b) - G(a).$$

Here is a definition:

Definition 2.2. If z is irrational, then by e^z we mean the unique number that has logarithm z:

$$\log e^z = z.$$

The pre-defined theorem-like constructs are **theorem**, **conjecture**, **proposition**, **lemma** and **corollary**. The pre-defined definition-like constructs are **example** and **definition**. You can add your own constructs using the *amsthm* interface [3]. The styles used in the `\theoremstyle` command are **acmplain** and **acmdefinition**.

Another construct is **proof**, for example, Proof. Suppose on the contrary there exists a real number L such that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L.$$

Then

$$l = \lim_{x\to c} f(x) = \lim_{x\to c} \left[gx \cdot \frac{f(x)}{g(x)} \right] = \lim_{x\to c} g(x) \cdot \lim_{x\to c} \frac{f(x)}{g(x)} = 0 \cdot L = 0,$$

which contradicts our assumption that $l \neq 0$. □

3 CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

A HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are different in the appendices. In the **appendix** environment, the command **section** is used to indicate the start of each Appendix, with alphabetic order designation (i.e., the first is A, the second B, etc.) and a title (if you include one). So, if you need hierarchical structure *within* an Appendix, start with **subsection** as the highest level. Here is an outline of the body of this document in Appendix-appropriate form:

A.1 Introduction

A.2 The Body of the Paper

A.2.1 Type Changes and Special Characters.

A.2.2 Math Equations.

Inline (In-text) Equations.

Display Equations.

A.2.3 Citations.

A.2.4 Tables.

A.2.5 Figures.

A.2.6 Theorem-like Constructs.

A Caveat for the TeX Expert.

A.3 Conclusions

A.4 References

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

B MORE HELP FOR THE HARDY

Of course, reading the source code is always useful. The file `acmart.pdf` contains both the user guide and the commented code.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

[1] Rafal Ablamowicz and Bertfried Fauser. 2007. CLIFFORD: a Maple 11 Package for Clifford Algebra Computations, version 11. (2007). Retrieved February 28, 2008 from http://math.tntech.edu/rafal/cliff11/index.html

[2] Patricia S. Abril and Robert Plant. 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan. 2007), 36–44. https://doi.org/10.1145/1188913.1188915

[3] American Mathematical Society 2015. *Using the amsthm Package.* American Mathematical Society. http://www.ctan.org/pkg/amsthm

[4] Sten Andler. 1979. Predicate Path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN symposium on Principles of Programming Languages (POPL '79).* ACM Press, New York, NY, 226–236. https://doi.org/10.1145/567752.567774

[5] David A. Anisi. 2003. *Optimal Motion Control of a Ground Vehicle.* Master's thesis. Royal Institute of Technology (KTH), Stockholm, Sweden.

[6] Mic Bowman, Saumya K. Debray, and Larry L. Peterson. 1993. Reasoning About Naming Systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (November 1993), 795–825. https://doi.org/10.1145/161468.161471

[7] Johannes Braams. 1991. Babel, a Multilingual Style-Option System for Use with LaTeX's Standard Document Styles. *TUGboat* 12, 2 (June 1991), 291–301.

[8] Malcolm Clark. 1991. Post Congress Tristesse. In *TeX90 Conference Proceedings.* TeX Users Group, 84–89.

[9] Kenneth L. Clarkson. 1985. *Algorithms for Closest-Point Problems (Computational Geometry).* Ph.D. Dissertation. Stanford University, Palo Alto, CA. UMI Order Number: AAT 8506171.

[10] Jacques Cohen (Ed.). 1996. Special issue: Digital Libraries. *Commun. ACM* 39, 11 (Nov. 1996).

[11] Sarah Cohen, Werner Nutt, and Yehoshua Sagic. 2007. Deciding equivalances among conjunctive aggregate queries. *J. ACM* 54, 2, Article 5 (April 2007), 50 pages. https://doi.org/10.1145/1219092.1219093

[12] Bruce P. Douglass, David Harel, and Mark B. Trakhtenbrot. 1998. Statecarts in use: structured analysis and object-orientation. In *Lectures on Embedded Systems*, Grzegorz Rozenberg and Frits W. Vaandrager (Eds.). Lecture Notes in Computer Science, Vol. 1494. Springer-Verlag, London, 368–394. https://doi.org/10.1007/3-540-65193-4_29

[13] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

[14] Ian Editor (Ed.). 2008. *The title of book two* (2nd. ed.). University of Chicago Press, Chicago, Chapter 100. https://doi.org/10.1007/3-540-09237-4

[15] Simon Fear. 2005. *Publication quality tables in LaTeX.* http://www.ctan.org/pkg/booktabs

[16] Matthew Van Gundy, Davide Balzarotti, and Giovanni Vigna. 2007. Catch me, if you can: Evading network signatures with web-based polymorphic worms. In *Proceedings of the first USENIX workshop on Offensive Technologies (WOOT '07).* USENIX Association, Berkley, CA, Article 7, 9 pages.

[17] David Harel. 1978. *LOGICS of Programs: AXIOMATICS and DESCRIPTIVE POWER.* MIT Research Lab Technical Report TR-200. Massachusetts Institute of Technology, Cambridge, MA.

[18] David Harel. 1979. *First-Order Dynamic Logic.* Lecture Notes in Computer Science, Vol. 68. Springer-Verlag, New York, NY. https://doi.org/10.1007/3-540-09237-4

[19] Maurice Herlihy. 1993. A Methodology for Implementing Highly Concurrent Data Objects. *ACM Trans. Program. Lang. Syst.* 15, 5 (November 1993), 745–770. https://doi.org/10.1145/161468.161469

[20] Lars Hörmander. 1985. *The analysis of linear partial differential operators. III.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 275. Springer-Verlag, Berlin, Germany. viii+525 pages. Pseudodifferential operators.

[21] Lars Hörmander. 1985. *The analysis of linear partial differential operators. IV.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 275. Springer-Verlag, Berlin, Germany. vii+352 pages. Fourier integral operators.

[22] IEEE 2004. IEEE TCSC Executive Committee. In *Proceedings of the IEEE International Conference on Web Services (ICWS '04).* IEEE Computer Society, Washington, DC, USA, 21–22. https://doi.org/10.1109/ICWS.2004.64

[23] Markus Kirschmer and John Voight. 2010. Algorithmic Enumeration of Ideal Classes for Quaternion Orders. *SIAM J. Comput.* 39, 5 (Jan. 2010), 1714–1747. https://doi.org/10.1137/080734467

[24] Donald E. Knuth. 1997. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms (3rd. ed.).* Addison Wesley Longman Publishing Co., Inc.

[25] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY.

[26] Leslie Lamport. 1986. *LaTeX: A Document Preparation System.* Addison-Wesley, Reading, MA.

[27] Newton Lee. 2005. Interview with Bill Kinder: January 13, 2005. Video. *Comput. Entertain.* 3, 1, Article 4 (Jan.-March 2005). https://doi.org/10.1145/1057270.1057278

[28] Dave Novak. 2003. Solder man. Video. In *ACM SIGGRAPH 2003 Video Review on Animation theater Program: Part I - Vol. 145 (July 27–27, 2003).* ACM Press, New York, NY, 4. https://doi.org/99.9999/woot07-S422

[29] Barack Obama. 2008. A more perfect union. Video. (5 March 2008). Retrieved March 21, 2008 from http://video.google.com/videoplay?docid=6528042696351994555

[30] Poker-Edge.Com. 2006. Stats and Analysis. (March 2006). Retrieved June 7, 2006 from http://www.poker-edge.com/stats.php

[31] Bernard Rous. 2008. The Enabling of Digital Libraries. *Digital Libraries* 12, 3, Article 5 (July 2008). To appear.

[32] Mehdi Saeedi, Morteza Saheb Zamani, and Mehdi Sedighi. 2010. A library-based synthesis methodology for reversible logic. *Microelectron. J.* 41, 4 (April 2010), 185–194.

[33] Mehdi Saeedi, Morteza Saheb Zamani, Mehdi Sedighi, and Zahra Sasanian. 2010. Synthesis of Reversible Circuit Using Cycle-Based Approach. *J. Emerg. Technol. Comput. Syst.* 6, 4 (Dec. 2010).

[34] S.L. Salas and Einar Hille. 1978. *Calculus: One and Several Variable.* John Wiley and Sons, New York.

[35] Joseph Scientist. 2009. The fountain of youth. (Aug. 2009). Patent No. 12345, Filed July 1st., 2008, Issued Aug. 9th., 2009.

[36] Stan W. Smith. 2010. An experiment in bibliographic mark-up: Parsing metadata for XML export. In *Proceedings of the 3rd. annual workshop on Librarians and Computers (LAC '10)*, Reginald N. Smythe and Alexander Noble (Eds.), Vol. 3. Paparazzi Press, Milan Italy, 422–431. https://doi.org/99.9999/woot07-S422

[37] Asad Z. Spector. 1990. Achieving application requirements. In *Distributed Systems* (2nd. ed.), Sape Mullender (Ed.). ACM Press, New York, NY, 19–33. https://doi.org/10.1145/90417.90738

[38] Harry Thornburg. 2001. Introduction to Bayesian Statistics. (March 2001). Retrieved March 2, 2005 from http://ccrma.stanford.edu/~jos/bayes/bayes.html

[39] TUG 2017. Institutional members of the TeX Users Group. (2017). Retrieved May 27, 2017 from http://wwtug.org/instmem.html

[40] Boris Veytsman. [n. d.]. acmart—Class for typesetting publications of ACM. ([n. d.]). Retrieved May 27, 2017 from http://www.ctan.org/pkg/acmart

List of Figures

Figure 1: A sample black and white graphic.

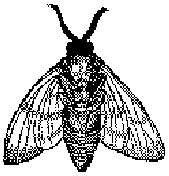


Figure 2: A sample black and white graphic that has been resized with the `includegraphics` command.

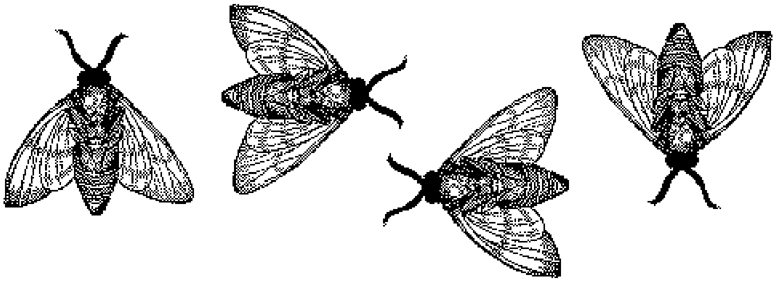


Figure 3: A sample black and white graphic that needs to span two columns of text.

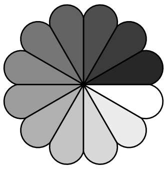


Figure 4: A sample black and white graphic that has been resized with the `includegraphics` command.

List of Tables

Table 1: Frequency of Special Characters

Non-English or Math	Frequency	Comments
Ø	1 in 1,000	For Swedish names
π	1 in 5	Common in math
$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage

Table 2: Some Typical Commands

Command	A Number	Comments
`\author`	100	Author
`\table`	300	For tables
`\table*`	400	For wider tables

Big Data Analytics and High Performance Computing

Dhawal Chaturvedi
Indiana University
2679 E. 7th St, Apt. C
Bloomington, IN 47408, USA
dhchat@iu.edu

ABSTRACT

This paper provides an introduction to Big Data and High Performance Computing and tries to find how they are related to each other.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating and information privacy.

2 THE BODY OF THE PAPER

3 CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

Big Data Analysis using MapReduce

Saurabh Kumar
Indiana University
Bloomington, Indiana 47408
kumarsau@iu.edu

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size.

REFERENCES

Big Data and Data Visualization

Pravin Deshmukh
Indiana University
P.O. Box 1212
Bloomington, Indiana 43017-6221
praadesh@iu.edu

ABSTRACT

This paper will provide an overview on how analytical findings of Big Data solutions can be visualized using various visualization technologies

KEYWORDS

i523

1 INTRODUCTION

Big data is widely used technology to consume huge amount of data. While there are various technologies available to process this data it is very important to have interactive, intuitive, user friendly data visualizations in place so that decision makers, business users will have clear understanding of findings of big data solutions. These visualizations will make help us to make informed decision looking at various trends over the period of time.

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data Platforms as a Service

Tiffany Fabianac
Indiana University
Bloomington, Indiana
tifabi@iu.edu

ABSTRACT

Big Data platform solutions allow data producers to use data to the fullest potential by combining processing engines with storage solutions and analytic technologies. Pharmaceutical clients are looking into platform solutions to safely store, analyze, and use clinical trial data, experimental data, drug development studies, drug production, regulation, and a number of other outlets. Just a few of the benefits of using a platform solution to manage these data outlets are not having to change current work processes, that management and other research groups can access and use data without needing special access to systems, and scaleability of storage and analytic components is seamless. The problems faced to implementing big data platform solutions include the selection of a platform vendor, the design of appropriate data architecture, and establishing effective user interfaces.

KEYWORDS

i523, HID313, Big Data, Platform, Cloud Architecture

1 INTRODUCTION

Most pharmaceutical companies have adopted one or many Laboratory Information Management Systems (LIMS) and/or Electronic Laboratory Notebooks (ELN). These systems are often implemented as standalone systems within a single Research and Development (R&D) group or even within a single laboratory. A problem seen in large- or mid-sized pharmaceutical companies is that different research groups within the same organization often implement different LIMS or ELN. This severely restricts data sharing and reuse between groups which leads to many problems including the same experiment being run multiple times between different groups, regulatory inefficiencies in tracking sample use and storage, and bottle necked development cycles due to missing data.

One of the emerging strategies to combat the problems arising from isolated systems is to combine systems using cloud computing. Platform as a Service (PaaS) provides an environment for the development and execution of applications and software tools. The platform is the heart of a cloud computing infrastructure that enables software on-top as well as data created from such software to be accessed and used my a multitude of users[7].

The benefits and challenges of using a PaaS approach to share and regulate R&D data within a large pharmaceutical company that has already implemented numerous laboratory systems will be outlined below.

2 IMPORTANCE OF PLATFORMS

Many organizations struggle with the aim of sharing data and processing tools among researchers. SaaP provides a method of better resource utilization while reducing maintenance costs[6].

3 IMPLEMENTING PLATFORMS

Some of the biggest concerns associated with implementing platforms involve security, selecting the right solution, designing the data architecture and associated relationships, and planning the user interface. All of the large platform providers have invested enormous amounts of resources into assuring the security of their data storage solutions. The right solution might be based on the applications available, the storage solution's design, the cost, the learning curve, or a number of other client based requirements. Data architecture has the overarching purpose to design the data warehouse solution without limitations to growth or analysis tools and query speed. User interface depends mostly on the user requirements, it could be driven by how much visibility is needed and how read and write privileges are designated.

The overarching concern with storing data outside of the organization is security. Numerous methods have been developed to assure cloud security such as integrated stacks used by Google and Microsoft Azure and Service Level Agreements (SLAs)[5]. Cloud companies are required to maintain high security at all levels. Google runs various vulnerability reward programs that pay developers, hackers, and security experts for finding security bugs. In addition to the product bugs, Google also maintains high security at their data centers which includes laser beam intrusion detection, multifactor access control, and biometrics to a limited population of less than 1% of Googlers[3].

Microsoft big data solutions have taken advantage of open source technologies by setting Hadoop as the center of their big data platform. Hadoop is implemented through Hortonworks Data Platform (HDP) which has been developed as a open source solution with Apache and other open source components. Microsoft allows cloud and on-premise implementation, but generally local environments are only used as proof of concept testing. Microsoft platform solutions allow for data to be manipulated and used in Microsoft tools such as Sharepoint and Excel while big data analysis, visualization, and mining can be performed using SQL Server Analysis Services or HDInsight. The Hadoop-based platform has no limitations with structured or unstructured data, a number of additional tools are available for data storage, and efficient queries provide a potential boost to discovery. Microsoft Azure storage runs $40 a month per 1TB and employs a pay for use plan to resource use within the platform's toolbox[4].

Amazon Web Services (AWS) offers data storage solutions in NoSQL and Relational Database models. Interactions with these data engines can be done using Hadoop, Interactive Query Service, or Elasticsearch. Amazon has designed their storage sources in such a way that clients can use any preferred open source application, but Amazon has also developed a toolbox of analytic tools. Amazon offers data warehousing through Amazon Redshift which allows for management, query, and analysis at the petabyte-scale. Amazon storage runs around $80 a month per 1TB. AWS offers Business Intelligence, Artificial Intelligence, Machine Learning, Internet of Things, Serverless Computing, and a number of data interface tools available in a pay-as-you-use billing form[1].

Google Cloud Platform (GCP) offers a complete end-to-end data storage solution which allows the use of GCP developed systems and open source tools. BigQuery is Google's data warehouse tool which is serverless and requires no infrastructure management with the assist of Google Cloud Dataflow. Dataflow eliminates the need for resource management and performance optimization. GCP storage runs $10 a month per 1TB. GCP has a number of applications for data manipulation. Dataproc allows dataset management through Hadoop and Spark, data visualization can be generated through Datalab, Data Studio, and Dataprep which are all Google developed applications[2].

All data storage solutions from relational databases to noSQL data stores to cloud data warehouses have to start with a defined architecture. The data architecture model will illustrate how data components will be organized and connected. The mindset of a data architect should be focused on reducing complexity of the data model while maintaining the highest level on utilization. This can be a fine line to walk as a designer. Complexity can be reduced by breaking user requirements down to the most basic and generalized principles to define the simplest data modules. An example of this might be a system that requires a number of different requests and instead of designing a component for vendor requests, user requests, and management requests the component is designed for request and request type. This generality allows for easy future scaling or additional system requirements not yet defined. Cloud systems maintain high utilization by manipulating data using strategic layering. One layer for storage, one layer for defining storage keys, another for combining query tools, another for consolidating query results and so on. With the more established cloud offerings a lot of these layers have already been supplied, but they transitions and interconnections still have to be outlined by a designer[8].

4 PLATFORMS AND BIG DATA

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor Von Laszewski and Teaching Assistants Saber Sheybani and Miao Jiang.

REFERENCES

[1] 2017. Big Data on AWS. Website. (Oct. 2017). https://aws.amazon.com/big-data/

[2] 2017. Big Data Solutions. Website. (Oct. 2017). https://cloud.google.com/products/big-data/

[3] 2017. Google Security Whitepaper. Website. (Oct. 2017). https://cloud.google.com/security/whitepaper#state-of-the-art_data_centers

[4] 2017. Understanding Microsoft big data solutions. Website. (Oct. 2017). https://msdn.microsoft.com/en-us/library/dn749804.aspx

[5] Valentina Casola, Alessandra De Benedictis, Massimiliano Rak, and Villano Umberto. 2014. Preliminary design of a platform-as-a-service to provide security in cloud. *ResearchGate* (01 2014), 752–757. https://www.researchgate.net/publication/289573602

[6] Sungyoung Oh, Jieun Cha, Myungkyu Ji, Hyekyung Kang, Seok Kim, Eunyoung Heo, Jong Soo Han, Hyunggoo Kang, Hoseok Chae, Hee Hwang, and Sooyoung Yoo. 2015. Architecture Design of Healthcare Software-as-a-Service Platform for Cloud-Based Clinical Decision Support Service. *Healthcare Informatics Research* 21, 2 (April 2015), 102–110. https://doi.org/10.4258/hir.2015.21.2.102

[7] Arto Ojala and Nina Helander. 2014. Value creation and evolution of a value network: A longitudinal case study on a Platform-as-a-Service provider. In *47th Hawaii International Conference on System Science*, Vol. 47. 975–984.

[8] Jerome H. Saltzer and M. Frans Kaashoek. 2009. *Principles of Computer System Design: An Introduction.* Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-374957-4.00010-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Amazon Web Services (AWS) in Support of Big Data and Analytics

Peter Russell
University of Indiana - Bloomington
petrusse@iu.edu

ABSTRACT

This paper will explore the logistics of Amazon Web Services and how companies are currently utilizing the service to process their big data needs.

KEYWORDS

Big Data, Cloud Computing, AWS, Big Data Analytics

1 INTRODUCTION

Amazon Web Services (AWS), the cloud service arm of Amazon, is currently the most dominant company in the cloud computing marketplace. With a market share of 31%, AWS holds a larger share than the next three closest competitors (Google, Microsoft and IBM)[1]. As a $10 billion a year line of business for Amazon, the revenue stream is incredibly diversified across multiple product offerings. One of these categories, which can broadly be described as 'business analytics,' have helped companies gain new insights into their customer experiences and competitive landscape.

REFERENCES

[1] Synergy Research Group. 2016. AWS Remains Dominant Despite Microsoft and Google Growth Surges. Website. (Feb. 2016).

Docker in support of Big Data Applications and Analytics

Anand Sriramulu
Indiana University
107 S Indiana Ave
Bloomington, Indiana, USA 47405
asriram@iu.edu

ABSTRACT

This paper will analyze the processing power of docker with big data use cases

KEYWORDS

i523

1 INTRODUCTION

ACKNOWLEDGMENTS

REFERENCES

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Lars Thørväld
The Thørväld Group
1 Thørväld Circle
Hekla, Iceland
larst@affiliation.org

Valerie Béranger
Inria Paris-Rocquencourt
Rocquencourt, France

Aparna Patel
Rajiv Gandhi University
Rono-Hills
Doimukh, Arunachal Pradesh, India

Huifen Chan
Tsinghua University
30 Shuangqing Rd
Haidian Qu, Beijing Shi, China

Charles Palmer
Palmer Research Laboratories
8600 Datapoint Drive
San Antonio, Texas 78229
cpalmer@prl.com

John Smith
The Thørväld Group
jsmith@affiliation.org

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size.

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

My great Big Dat Paper

YuanMing Huang
Indiana University
800 N Union st
Bloomington, Indiana 47408
huang226@iu.edu

ABSTRACT

THIS IS AN ABSTRACT

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

This is an indtroduction

2 THE BODY OF THE PAPER

this is the body

3 CONCLUSIONS

this is the conclusion

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

What Separates Big Data from Lots of Data

Gabriel Jones
Indiana University
107 S Indiana Ave
Bloomington, Indiana, USA 47405
gabejone@indiana.edu

ABSTRACT

TIn this paper, we will briefly analyze the history of data to show how having *lots of data* stored in large databases hardly differs from data storage and analysis in the early days of SQL, or even before computers. We then explain how *big data* represents a paradigmatic shift from traditional large data storage and analysis. We conclude that organizations that do not understand this paradigmatic shift are more likely to fail in big data projects.

KEYWORDS

i523

1 INTRODUCTION

This is my introduction. [1]

2 CONCLUSIONS

I conclude that...

ACKNOWLEDGMENTS

Generic acknowledgements

REFERENCES

[1] Carl Lagoze. 2014. Big Data, data integrity, and the fracturing of the control zone. *Big Data and Society* (NO 2014). https://doi.org/10.1177/2053951714558281

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Lars Thørväld
The Thørväld Group
1 Thørväld Circle
Hekla, Iceland
larst@affiliation.org

Valerie Béranger
Inria Paris-Rocquencourt
Rocquencourt, France

Aparna Patel
Rajiv Gandhi University
Rono-Hills
Doimukh, Arunachal Pradesh, India

Huifen Chan
Tsinghua University
30 Shuangqing Rd
Haidian Qu, Beijing Shi, China

Charles Palmer
Palmer Research Laboratories
8600 Datapoint Drive
San Antonio, Texas 78229
cpalmer@prl.com

John Smith
The Thørväld Group
jsmith@affiliation.org

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size.

2 THE BODY OF THE PAPER

Typically, the body of a paper is organized into a hierarchical structure, with numbered or unnumbered headings for sections, subsections, sub-subsections, and even smaller sections. The command `\section` that precedes this paragraph is part of such a hierarchy. LaTeX handles the numbering and placement of these headings for you, when you use the appropriate heading commands around the titles of the headings. If you want a sub-subsection or smaller part to be unnumbered in your output, simply append an asterisk to the command name. Examples of both numbered and unnumbered headings will appear throughout the balance of this sample document.

Because the entire article is contained in the **document** environment, you can indicate the start of a new paragraph with a blank line in your input file; that is why this sentence forms a separate paragraph.

2.1 Type Changes and *Special* Characters

We have already seen several typeface changes in this sample. You can indicate italicized words or phrases in your text with the command `\textit`; emboldening with the command `\textbf` and typewriter-style (for instance, for computer code) with `\texttt`. But remember, you do not have to indicate typestyle changes when such changes are part of the *structural* elements of your article; for instance, the heading of this subsection will be in a sans serif[1] typeface, but that is handled by the document class file. Take care with the use of the curly braces in typeface changes; they mark the beginning and end of the text that is to be in the different typeface.

You can use whatever symbols, accented characters, or non-English characters you need anywhere in your document; you can find a complete list of what is available in the *LaTeX User's Guide* [**?**].

2.2 Math Equations

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections.

2.2.1 Inline (In-text) Equations. A formula that appears in the running text is called an inline or in-text formula. It is produced by the **math** environment, which can be invoked with the usual `\begin . . . \end` construction or with the short form `$ . . . $`. You can use any of the symbols and structures, from α to ω, available in LaTeX [**?**]; this section will simply show a few examples of in-text equations in context. Notice how this equation:
$\lim_{n\rightarrow\infty} x = 0$,
set here in in-line math style, looks slightly different when set in display style. (See next section).

2.2.2 Display Equations. A numbered display equation—one set off by vertical space from the text and centered horizontally—is

[1] Another footnote here. Let's make this a rather long one to see how it looks. Footnotes must be avoided.

produced by the **equation** environment. An unnumbered display equation is produced by the **displaymath** environment.

Again, in either environment, you can use any of the symbols and structures available in LaTeX; this section will just give a couple of examples of display equations in context. First, consider the equation, shown as an inline equation above:

$$\lim_{n\rightarrow\infty} x = 0 \tag{1}$$

Notice how it is formatted somewhat differently in the **displaymath** environment. Now, we'll enter an unnumbered equation:

$$\sum_{i=0}^{\infty} x + 1$$

and follow it with another numbered equation:

$$\sum_{i=0}^{\infty} x_i = \int_{0}^{\pi+2} f \tag{2}$$

just to demonstrate LaTeX's able handling of numbering.

2.3 Citations

Citations to articles [? ? ? ?], conference proceedings [?] or maybe books [? ?] listed in the Bibliography section of your article will occur throughout the text of your article. You should use BibTeX to automatically produce this bibliography; you simply need to insert one of several citation commands with a key of the item cited in the proper location in the `.tex` file [?]. The key is a short reference you invent to uniquely identify each work; in this sample document, the key is the first author's surname and a word from the title. This identifying key is included with each item in the `.bib` file for your article.

The details of the construction of the `.bib` file are beyond the scope of this sample document, but more information can be found in the *Author's Guide*, and exhaustive details in the *LaTeX User's Guide* by Lamport [?].

This article shows only the plainest form of the citation command, using `\cite`.

Some examples. A paginated journal article [?], an enumerated journal article [?], a reference to an entire issue [?], a monograph (whole book) [?], a monograph/whole book in a series (see 2a in spec. document) [?], a divisible-book such as an anthology or compilation [?] followed by the same example, however we only output the series if the volume number is given [?] (so Editor00a's series should NOT be present since it has no vol. no.), a chapter in a divisible book [?], a chapter in a divisible book in a series [?], a multi-volume work as book [?], an article in a proceedings (of a conference, symposium, workshop for example) (paginated proceedings article) [?], a proceedings article with all possible elements [?], an example of an enumerated proceedings article [?], an informally published work [?], a doctoral dissertation [?], a master's thesis: [?], an online document / world wide web resource [? ? ?], a video game (Case 1) [?] and (Case 2) [?] and [?] and (Case 3) a patent [?], work accepted for publication [?], 'YYYYb'-test for prolific author [?] and [?]. Other cites might contain 'duplicate' DOI and URLs (some SIAM articles) [?]. Boris / Barbara Beeton: multi-volume works as books [?] and [?].

A couple of citations with DOIs: [? ?].

Online citations: [? ? ?].

We use jabref to manage all citations. A paper without managing a bib file will be returned without review. in the bibtex file all urls are added to rfernces with the *url* filed. They are not to be included in the *howpublished* or *note* field.

2.4 Tables

Because tables cannot be split across pages, the best placement for them is typically the top of the page nearest their initial cite. To ensure this proper "floating" placement of tables, use the environment **table** to enclose the table's contents and the table caption. The contents of the table itself must go in the **tabular** environment, to be aligned properly in rows and columns, with the desired horizontal and vertical rules. Again, detailed instructions on **tabular** material are found in the *LaTeX User's Guide*.

Immediately following this sentence is the point at which Table 1 is included in the input file; compare the placement of the table here with the table in the printed output of this document.

[Table 1 about here.]

To set a wider table, which takes up the whole width of the page's live area, use the environment **table*** to enclose the table's contents and the table caption. As with a single-column table, this wide table will "float" to a location deemed more desirable. Immediately following this sentence is the point at which Table 2 is included in the input file; again, it is instructive to compare the placement of the table here with the table in the printed output of this document.

[Table 2 about here.]

It is strongly recommended to use the package booktabs [?] and follow its main principles of typography with respect to tables:

(1) Never, ever use vertical rules.
(2) Never use double rules.

It is also a good idea not to overuse horizontal rules.

2.5 Figures

Like tables, figures cannot be split across pages; the best placement for them is typically the top or the bottom of the page nearest their initial cite. To ensure this proper "floating" placement of figures, use the environment **figure** to enclose the figure and its caption.

This sample document contains examples of `.eps` files to be displayable with LaTeX. If you work with pdfLaTeX, use files in the `.pdf` format. Note that most modern TeX systems will convert `.eps` to `.pdf` for you on the fly. More details on each of these are found in the *Author's Guide*.

[Figure 1 about here.]

[Figure 2 about here.]

As was the case with tables, you may want a figure that spans two columns. To do this, and still to ensure proper "floating" placement of tables, use the environment **figure*** to enclose the figure and its caption. And don't forget to end the environment with **figure***, not **figure**!

[Figure 3 about here.]

[Figure 4 about here.]

2.6 Theorem-like Constructs

Other common constructs that may occur in your article are the forms for logical constructs like theorems, axioms, corollaries and proofs. ACM uses two types of these constructs: theorem-like and definition-like.

Here is a theorem:

Theorem 2.1. *Let f be continuous on $[a, b]$. If G is an antiderivative for f on $[a, b]$, then*

$$\int_a^b f(t)\,dt = G(b) - G(a).$$

Here is a definition:

Definition 2.2. If z is irrational, then by e^z we mean the unique number that has logarithm z:

$$\log e^z = z.$$

The pre-defined theorem-like constructs are **theorem**, **conjecture**, **proposition**, **lemma** and **corollary**. The pre-defined definition-like constructs are **example** and **definition**. You can add your own constructs using the *amsthm* interface [?]. The styles used in the `\theoremstyle` command are **acmplain** and **acmdefinition**.

Another construct is **proof**, for example,

Proof. Suppose on the contrary there exists a real number L such that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L.$$

Then

$$l = \lim_{x\to c} f(x) = \lim_{x\to c} \left[gx \cdot \frac{f(x)}{g(x)} \right] = \lim_{x\to c} g(x) \cdot \lim_{x\to c} \frac{f(x)}{g(x)} = 0 \cdot L = 0,$$

which contradicts our assumption that $l \neq 0$. □

3 CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

A HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are different in the appendices. In the **appendix** environment, the command **section** is used to indicate the start of each Appendix, with alphabetic order designation (i.e., the first is A, the second B, etc.) and a title (if you include one). So, if you need hierarchical structure *within* an Appendix, start with **subsection** as the highest level. Here is an outline of the body of this document in Appendix-appropriate form:

A.1 Introduction

A.2 The Body of the Paper

A.2.1 Type Changes and Special Characters.

A.2.2 Math Equations.

Inline (In-text) Equations.

Display Equations.

A.2.3 Citations.

A.2.4 Tables.

A.2.5 Figures.

A.2.6 Theorem-like Constructs.

A Caveat for the TeX Expert.

A.3 Conclusions

A.4 References

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

B MORE HELP FOR THE HARDY

Of course, reading the source code is always useful. The file `acmart.pdf` contains both the user guide and the commented code.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

List of Figures

Figure 1: A sample black and white graphic.

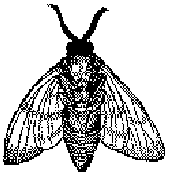


Figure 2: A sample black and white graphic that has been resized with the `includegraphics` command.

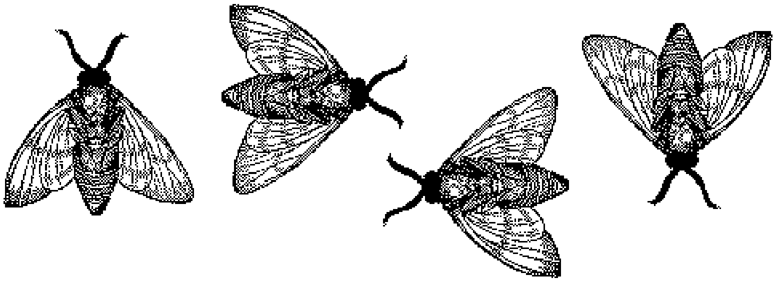


Figure 3: A sample black and white graphic that needs to span two columns of text.

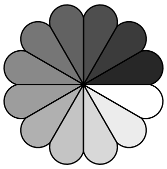


Figure 4: A sample black and white graphic that has been resized with the `includegraphics` command.

List of Tables

Table 1: Frequency of Special Characters

Non-English or Math	Frequency	Comments
Ø	1 in 1,000	For Swedish names
π	1 in 5	Common in math
$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage

Table 2: Some Typical Commands

Command	A Number	Comments
`\author`	100	Author
`\table`	300	For tables
`\table*`	400	For wider tables

Big Data Analytic Architecture for Real Time Traffic Control

Syam Sundar herle
Indiana University
Bloomington, Indiana 47408
syampara@iu.edu

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

This is a introduction.

2 THE BODY OF THE PAPER

Typically, the body of a paper is organized into a hierarchical structure, with numbered or unnumbered headings for sections, subsections, sub-subsections, and even smaller sections. The command `\section` that precedes this paragraph is part of such a hierarchy. LaTeX handles the numbering and placement of these headings for you, when you use the appropriate heading commands around the titles of the headings. If you want a sub-subsection or smaller part to be unnumbered in your output, simply append an asterisk to the command name. Examples of both numbered and unnumbered headings will appear throughout the balance of this sample document.

Because the entire article is contained in the **document** environment, you can indicate the start of a new paragraph with a blank line in your input file; that is why this sentence forms a separate paragraph.

2.1 Type Changes and *Special* Characters

We have already seen several typeface changes in this sample. You can indicate italicized words or phrases in your text with the command `\textit`; emboldening with the command `\textbf` and typewriter-style (for instance, for computer code) with `\texttt`. But remember, you do not have to indicate typestyle changes when such changes are part of the *structural* elements of your article; for instance, the heading of this subsection will be in a sans serif[1] typeface, but that is handled by the document class file. Take care with the use of the curly braces in typeface changes; they mark the beginning and end of the text that is to be in the different typeface.

You can use whatever symbols, accented characters, or non-English characters you need anywhere in your document; you can find a complete list of what is available in the *LaTeX User's Guide* [26].

2.2 Math Equations

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections.

2.2.1 Inline (In-text) Equations. A formula that appears in the running text is called an inline or in-text formula. It is produced by the **math** environment, which can be invoked with the usual `\begin . . . \end` construction or with the short form `$ . . . $`. You can use any of the symbols and structures, from α to ω, available in LaTeX [26]; this section will simply show a few examples of in-text equations in context. Notice how this equation:
$\lim_{n\to\infty} x = 0$,
set here in in-line math style, looks slightly different when set in display style. (See next section).

2.2.2 Display Equations. A numbered display equation—one set off by vertical space from the text and centered horizontally—is produced by the **equation** environment. An unnumbered display equation is produced by the **displaymath** environment.

Again, in either environment, you can use any of the symbols and structures available in LaTeX; this section will just give a couple of examples of display equations in context. First, consider the equation, shown as an inline equation above:

$$\lim_{n\to\infty} x = 0 \tag{1}$$

Notice how it is formatted somewhat differently in the **displaymath** environment. Now, we'll enter an unnumbered equation:

$$\sum_{i=0}^{\infty} x + 1$$

and follow it with another numbered equation:

$$\sum_{i=0}^{\infty} x_i = \int_0^{\pi+2} f \tag{2}$$

just to demonstrate LaTeX's able handling of numbering.

2.3 Citations

Citations to articles [6–8, 19], conference proceedings [8] or maybe books [26, 34] listed in the Bibliography section of your article will occur throughout the text of your article. You should use BibTeX to automatically produce this bibliography; you simply need to insert one of several citation commands with a key of the item cited in the proper location in the `.tex` file [26]. The key is a short reference you invent to uniquely identify each work; in this sample document, the key is the first author's surname and a word from the title. This identifying key is included with each item in the `.bib` file for your article.

[1] Another footnote here. Let's make this a rather long one to see how it looks. Footnotes must be avoided.

The details of the construction of the `.bib` file are beyond the scope of this sample document, but more information can be found in the *Author's Guide*, and exhaustive details in the *LaTeX User's Guide* by Lamport [26].

This article shows only the plainest form of the citation command, using `\cite`.

Some examples. A paginated journal article [2], an enumerated journal article [11], a reference to an entire issue [10], a monograph (whole book) [25], a monograph/whole book in a series (see 2a in spec. document) [18], a divisible-book such as an anthology or compilation [13] followed by the same example, however we only output the series if the volume number is given [14] (so Editor00a's series should NOT be present since it has no vol. no.), a chapter in a divisible book [37], a chapter in a divisible book in a series [12], a multi-volume work as book [24], an article in a proceedings (of a conference, symposium, workshop for example) (paginated proceedings article) [4], a proceedings article with all possible elements [36], an example of an enumerated proceedings article [16], an informally published work [17], a doctoral dissertation [9], a master's thesis: [5], an online document / world wide web resource [1, 30, 38], a video game (Case 1) [29] and (Case 2) [28] and [27] and (Case 3) a patent [35], work accepted for publication [31], 'YYYYb'-test for prolific author [32] and [33]. Other cites might contain 'duplicate' DOI and URLs (some SIAM articles) [23]. Boris / Barbara Beeton: multi-volume works as books [21] and [20].

A couple of citations with DOIs: [22, 23].

Online citations: [38–40].

We use jabref to manage all citations. A paper without managing a bib file will be returned without review. in the bibtex file all urls are added to rfernces with the *url* filed. They are not to be included in the *howpublished* or *note* field.

2.4 Tables

Because tables cannot be split across pages, the best placement for them is typically the top of the page nearest their initial cite. To ensure this proper "floating" placement of tables, use the environment **table** to enclose the table's contents and the table caption. The contents of the table itself must go in the **tabular** environment, to be aligned properly in rows and columns, with the desired horizontal and vertical rules. Again, detailed instructions on **tabular** material are found in the *LaTeX User's Guide.*

Immediately following this sentence is the point at which Table 1 is included in the input file; compare the placement of the table here with the table in the printed output of this document.

[Table 1 about here.]

To set a wider table, which takes up the whole width of the page's live area, use the environment **table*** to enclose the table's contents and the table caption. As with a single-column table, this wide table will "float" to a location deemed more desirable. Immediately following this sentence is the point at which Table 2 is included in the input file; again, it is instructive to compare the placement of the table here with the table in the printed output of this document.

[Table 2 about here.]

It is strongly recommended to use the package booktabs [15] and follow its main principles of typography with respect to tables:

(1) Never, ever use vertical rules.
(2) Never use double rules.

It is also a good idea not to overuse horizontal rules.

2.5 Figures

Like tables, figures cannot be split across pages; the best placement for them is typically the top or the bottom of the page nearest their initial cite. To ensure this proper "floating" placement of figures, use the environment **figure** to enclose the figure and its caption.

This sample document contains examples of `.eps` files to be displayable with LaTeX. If you work with pdfLaTeX, use files in the `.pdf` format. Note that most modern TeX systems will convert `.eps` to `.pdf` for you on the fly. More details on each of these are found in the *Author's Guide.*

As was the case with tables, you may want a figure that spans two columns. To do this, and still to ensure proper "floating" placement of tables, use the environment **figure*** to enclose the figure and its caption. And don't forget to end the environment with **figure***, not **figure**!

2.6 Theorem-like Constructs

Other common constructs that may occur in your article are the forms for logical constructs like theorems, axioms, corollaries and proofs. ACM uses two types of these constructs: theorem-like and definition-like.

Here is a theorem:

THEOREM 2.1. *Let f be continuous on $[a, b]$. If G is an antiderivative for f on $[a, b]$, then*

$$\int_a^b f(t)\,dt = G(b) - G(a).$$

Here is a definition:

Definition 2.2. If z is irrational, then by e^z we mean the unique number that has logarithm z:

$$\log e^z = z.$$

The pre-defined theorem-like constructs are **theorem**, **conjecture**, **proposition**, **lemma** and **corollary**. The pre-defined definition-like constructs are **example** and **definition**. You can add your own constructs using the *amsthm* interface [3]. The styles used in the `\theoremstyle` command are **acmplain** and **acmdefinition**.

Another construct is **proof**, for example,

PROOF. Suppose on the contrary there exists a real number L such that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L.$$

Then

$$l = \lim_{x\to c} f(x) = \lim_{x\to c} \left[gx \cdot \frac{f(x)}{g(x)} \right] = \lim_{x\to c} g(x) \cdot \lim_{x\to c} \frac{f(x)}{g(x)} = 0 \cdot L = 0,$$

which contradicts our assumption that $l \neq 0$. □

3 CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

A HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are different in the appendices. In the **appendix** environment, the command **section** is used to indicate the start of each Appendix, with alphabetic order designation (i.e., the first is A, the second B, etc.) and a title (if you include one). So, if you need hierarchical structure *within* an Appendix, start with **subsection** as the highest level. Here is an outline of the body of this document in Appendix-appropriate form:

A.1 Introduction

A.2 The Body of the Paper

A.2.1 Type Changes and Special Characters.

A.2.2 Math Equations.

Inline (In-text) Equations.

Display Equations.

A.2.3 Citations.

A.2.4 Tables.

A.2.5 Figures.

A.2.6 Theorem-like Constructs.

A Caveat for the TEX Expert.

A.3 Conclusions

A.4 References

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

B MORE HELP FOR THE HARDY

Of course, reading the source code is always useful. The file `acmart.pdf` contains both the user guide and the commented code.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

[1] Rafal Ablamowicz and Bertfried Fauser. 2007. CLIFFORD: a Maple 11 Package for Clifford Algebra Computations, version 11. (2007). Retrieved February 28, 2008 from http://math.tntech.edu/rafal/cliff11/index.html

[2] Patricia S. Abril and Robert Plant. 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan. 2007), 36–44. https://doi.org/10.1145/1188913.1188915

[3] American Mathematical Society 2015. *Using the amsthm Package.* American Mathematical Society. http://www.ctan.org/pkg/amsthm

[4] Sten Andler. 1979. Predicate Path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN symposium on Principles of Programming Languages (POPL '79).* ACM Press, New York, NY, 226–236. https://doi.org/10.1145/567752.567774

[5] David A. Anisi. 2003. *Optimal Motion Control of a Ground Vehicle.* Master's thesis. Royal Institute of Technology (KTH), Stockholm, Sweden.

[6] Mic Bowman, Saumya K. Debray, and Larry L. Peterson. 1993. Reasoning About Naming Systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (November 1993), 795–825. https://doi.org/10.1145/161468.161471

[7] Johannes Braams. 1991. Babel, a Multilingual Style-Option System for Use with LaTeX's Standard Document Styles. *TUGboat* 12, 2 (June 1991), 291–301.

[8] Malcolm Clark. 1991. Post Congress Tristesse. In *TeX90 Conference Proceedings.* TeX Users Group, 84–89.

[9] Kenneth L. Clarkson. 1985. *Algorithms for Closest-Point Problems (Computational Geometry).* Ph.D. Dissertation. Stanford University, Palo Alto, CA. UMI Order Number: AAT 8506171.

[10] Jacques Cohen (Ed.). 1996. Special issue: Digital Libraries. *Commun. ACM* 39, 11 (Nov. 1996).

[11] Sarah Cohen, Werner Nutt, and Yehoshua Sagic. 2007. Deciding equivalances among conjunctive aggregate queries. *J. ACM* 54, 2, Article 5 (April 2007), 50 pages. https://doi.org/10.1145/1219092.1219093

[12] Bruce P. Douglass, David Harel, and Mark B. Trakhtenbrot. 1998. Statecarts in use: structured analysis and object-orientation. In *Lectures on Embedded Systems*, Grzegorz Rozenberg and Frits W. Vaandrager (Eds.). Lecture Notes in Computer Science, Vol. 1494. Springer-Verlag, London, 368–394. https://doi.org/10.1007/3-540-65193-4_29

[13] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

[14] Ian Editor (Ed.). 2008. *The title of book two* (2nd. ed.). University of Chicago Press, Chicago, Chapter 100. https://doi.org/10.1007/3-540-09237-4

[15] Simon Fear. 2005. *Publication quality tables in LaTeX.* http://www.ctan.org/pkg/booktabs

[16] Matthew Van Gundy, Davide Balzarotti, and Giovanni Vigna. 2007. Catch me, if you can: Evading network signatures with web-based polymorphic worms. In *Proceedings of the first USENIX workshop on Offensive Technologies (WOOT '07).* USENIX Association, Berkley, CA, Article 7, 9 pages.

[17] David Harel. 1978. *LOGICS of Programs: AXIOMATICS and DESCRIPTIVE POWER.* MIT Research Lab Technical Report TR-200. Massachusetts Institute of Technology, Cambridge, MA.

[18] David Harel. 1979. *First-Order Dynamic Logic.* Lecture Notes in Computer Science, Vol. 68. Springer-Verlag, New York, NY. https://doi.org/10.1007/3-540-09237-4

[19] Maurice Herlihy. 1993. A Methodology for Implementing Highly Concurrent Data Objects. *ACM Trans. Program. Lang. Syst.* 15, 5 (November 1993), 745–770. https://doi.org/10.1145/161468.161469

[20] Lars Hörmander. 1985. *The analysis of linear partial differential operators. III.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 275. Springer-Verlag, Berlin, Germany. viii+525 pages. Pseudodifferential operators.

[21] Lars Hörmander. 1985. *The analysis of linear partial differential operators. IV.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 275. Springer-Verlag, Berlin, Germany. vii+352 pages. Fourier integral operators.

[22] IEEE 2004. IEEE TCSC Executive Committee. In *Proceedings of the IEEE International Conference on Web Services (ICWS '04).* IEEE Computer Society, Washington, DC, USA, 21–22. https://doi.org/10.1109/ICWS.2004.64

[23] Markus Kirschmer and John Voight. 2010. Algorithmic Enumeration of Ideal Classes for Quaternion Orders. *SIAM J. Comput.* 39, 5 (Jan. 2010), 1714–1747. https://doi.org/10.1137/080734467

[24] Donald E. Knuth. 1997. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms (3rd. ed.).* Addison Wesley Longman Publishing Co., Inc.

[25] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY.

[26] Leslie Lamport. 1986. *LaTeX: A Document Preparation System.* Addison-Wesley, Reading, MA.

[27] Newton Lee. 2005. Interview with Bill Kinder: January 13, 2005. Video. *Comput. Entertain.* 3, 1, Article 4 (Jan.-March 2005). https://doi.org/10.1145/1057270.1057278

[28] Dave Novak. 2003. Solder man. Video. In *ACM SIGGRAPH 2003 Video Review on Animation theater Program: Part I - Vol. 145 (July 27–27, 2003)*. ACM Press, New York, NY, 4. https://doi.org/99.9999/woot07-S422

[29] Barack Obama. 2008. A more perfect union. Video. (5 March 2008). Retrieved March 21, 2008 from http://video.google.com/videoplay?docid=6528042696351994555

[30] Poker-Edge.Com. 2006. Stats and Analysis. (March 2006). Retrieved June 7, 2006 from http://www.poker-edge.com/stats.php

[31] Bernard Rous. 2008. The Enabling of Digital Libraries. *Digital Libraries* 12, 3, Article 5 (July 2008). To appear.

[32] Mehdi Saeedi, Morteza Saheb Zamani, and Mehdi Sedighi. 2010. A library-based synthesis methodology for reversible logic. *Microelectron. J.* 41, 4 (April 2010), 185–194.

[33] Mehdi Saeedi, Morteza Saheb Zamani, Mehdi Sedighi, and Zahra Sasanian. 2010. Synthesis of Reversible Circuit Using Cycle-Based Approach. *J. Emerg. Technol. Comput. Syst.* 6, 4 (Dec. 2010).

[34] S.L. Salas and Einar Hille. 1978. *Calculus: One and Several Variable.* John Wiley and Sons, New York.

[35] Joseph Scientist. 2009. The fountain of youth. (Aug. 2009). Patent No. 12345, Filed July 1st., 2008, Issued Aug. 9th., 2009.

[36] Stan W. Smith. 2010. An experiment in bibliographic mark-up: Parsing metadata for XML export. In *Proceedings of the 3rd. annual workshop on Librarians and Computers (LAC '10)*, Reginald N. Smythe and Alexander Noble (Eds.), Vol. 3. Paparazzi Press, Milan Italy, 422–431. https://doi.org/99.9999/woot07-S422

[37] Asad Z. Spector. 1990. Achieving application requirements. In *Distributed Systems* (2nd. ed.), Sape Mullender (Ed.). ACM Press, New York, NY, 19–33. https://doi.org/10.1145/90417.90738

[38] Harry Thornburg. 2001. Introduction to Bayesian Statistics. (March 2001). Retrieved March 2, 2005 from http://ccrma.stanford.edu/~jos/bayes/bayes.html

[39] TUG 2017. Institutional members of the TEX Users Group. (2017). Retrieved May 27, 2017 from http://wwtug.org/instmem.html

[40] Boris Veytsman. [n. d.]. acmart—Class for typesetting publications of ACM. ([n. d.]). Retrieved May 27, 2017 from http://www.ctan.org/pkg/acmart

List of Tables

Table 1: Frequency of Special Characters

Non-English or Math	Frequency	Comments
Ø	1 in 1,000	For Swedish names
π	1 in 5	Common in math
$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage

Table 2: Some Typical Commands

Command	A Number	Comments
`\author`	100	Author
`\table`	300	For tables
`\table*`	400	For wider tables

Optimizing Mass Transit Bus Routes with Big Data

Matthew Schwartzer
Indiana University
919 E 10th St
Bloomington, Indiana 43017-6221
mabschwa@indiana.edu

ABSTRACT

This paper provides a sample of a LATEX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523, hid225, LATEX, public tranist, route optimization

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size[1].

ACKNOWLEDGMENTS

The authors would like to thank Prof..

REFERENCES

[1] Keven Richly, Ralf Teusner, Alexander Immer, Fabian Windheuser, and Lennard Wolf. 2015. Optimizing Routes of Public Transportation Systems by Analyzing the Data of Taxi Rides. In *Proceedings of the 1st International ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics (UrbanGIS'15).* ACM, New York, NY, USA, 70–76. https://doi.org/10.1145/2835022.2835035

Big Data Applications in Self-Driving Cars

Borga Edionse Usifo
Indiana University Bloomington
107 S Indiana Ave
Bloomington, Indiana 47405
busifo@iu.edu

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Huiyi Chen
Institute for Clarity in Documentation
2451 E. 10TH ST., 612
Bloomington, Indiana 47408
huiychen@indiana.edu

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

This is my Intro

2 THE BODY OF THE PAPER

3 CONCLUSIONS

This is my conclusion.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

REFERENCES

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Lars Thørväld
The Thørväld Group
1 Thørväld Circle
Hekla, Iceland
larst@affiliation.org

Valerie Béranger
Inria Paris-Rocquencourt
Rocquencourt, France

Aparna Patel
Rajiv Gandhi University
Rono-Hills
Doimukh, Arunachal Pradesh, India

Huifen Chan
Tsinghua University
30 Shuangqing Rd
Haidian Qu, Beijing Shi, China

Charles Palmer
Palmer Research Laboratories
8600 Datapoint Drive
San Antonio, Texas 78229
cpalmer@prl.com

John Smith
The Thørväld Group
jsmith@affiliation.org

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size.

2 THE BODY OF THE PAPER

Typically, the body of a paper is organized into a hierarchical structure, with numbered or unnumbered headings for sections, subsections, sub-subsections, and even smaller sections. The command `\section` that precedes this paragraph is part of such a hierarchy. LaTeX handles the numbering and placement of these headings for you, when you use the appropriate heading commands around the titles of the headings. If you want a sub-subsection or smaller part to be unnumbered in your output, simply append an asterisk to the command name. Examples of both numbered and unnumbered headings will appear throughout the balance of this sample document.

Because the entire article is contained in the **document** environment, you can indicate the start of a new paragraph with a blank line in your input file; that is why this sentence forms a separate paragraph.

2.1 Type Changes and *Special* Characters

We have already seen several typeface changes in this sample. You can indicate italicized words or phrases in your text with the command `\textit`; emboldening with the command `\textbf` and typewriter-style (for instance, for computer code) with `\texttt`. But remember, you do not have to indicate typestyle changes when such changes are part of the *structural* elements of your article; for instance, the heading of this subsection will be in a sans serif[1] typeface, but that is handled by the document class file. Take care with the use of the curly braces in typeface changes; they mark the beginning and end of the text that is to be in the different typeface.

You can use whatever symbols, accented characters, or non-English characters you need anywhere in your document; you can find a complete list of what is available in the *LaTeX User's Guide* [25].

2.2 Math Equations

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections.

2.2.1 Inline (In-text) Equations. A formula that appears in the running text is called an inline or in-text formula. It is produced by the **math** environment, which can be invoked with the usual `\begin . . . \end` construction or with the short form `$ . . . $`. You can use any of the symbols and structures, from α to ω, available in LaTeX [25]; this section will simply show a few examples of in-text equations in context. Notice how this equation:
$\lim_{n\rightarrow\infty}x=0$,
set here in in-line math style, looks slightly different when set in display style. (See next section).

2.2.2 Display Equations. A numbered display equation—one set off by vertical space from the text and centered horizontally—is produced by the **equation** environment. An unnumbered display equation is produced by the **displaymath** environment.

Again, in either environment, you can use any of the symbols and structures available in LaTeX; this section will just give a couple of examples of display equations in context. First, consider the equation, shown as an inline equation above:

$$\lim_{n\rightarrow\infty}x=0 \tag{1}$$

Notice how it is formatted somewhat differently in the **displaymath** environment. Now, we'll enter an unnumbered equation:

$$\sum_{i=0}^{\infty} x + 1$$

and follow it with another numbered equation:

$$\sum_{i=0}^{\infty}x_i=\int_{0}^{\pi+2} f \tag{2}$$

just to demonstrate LaTeX's able handling of numbering.

2.3 Citations

Citations to articles [6–8, 18], conference proceedings [8] or maybe books [25, 33] listed in the Bibliography section of your article will

[1] Another footnote here. Let's make this a rather long one to see how it looks. Footnotes must be avoided.

occur throughout the text of your article. You should use BibTeX to automatically produce this bibliography; you simply need to insert one of several citation commands with a key of the item cited in the proper location in the `.tex` file [25]. The key is a short reference you invent to uniquely identify each work; in this sample document, the key is the first author's surname and a word from the title. This identifying key is included with each item in the `.bib` file for your article.

The details of the construction of the `.bib` file are beyond the scope of this sample document, but more information can be found in the *Author's Guide*, and exhaustive details in the *LaTeX User's Guide* by Lamport [25].

This article shows only the plainest form of the citation command, using `\cite`.

Some examples. A paginated journal article [2], an enumerated journal article [11], a reference to an entire issue [10], a monograph (whole book) [24], a monograph/whole book in a series (see 2a in spec. document) [17], a divisible-book such as an anthology or compilation [13] followed by the same example, however we only output the series if the volume number is given [14] (so Editor00a's series should NOT be present since it has no vol. no.), a chapter in a divisible book [36], a chapter in a divisible book in a series [12], a multi-volume work as book [23], an article in a proceedings (of a conference, symposium, workshop for example) (paginated proceedings article) [4], a proceedings article with all possible elements [35], an example of an enumerated proceedings article [15], an informally published work [16], a doctoral dissertation [9], a master's thesis: [5], an online document / world wide web resource [1, 29, 37], a video game (Case 1) [28] and (Case 2) [27] and [26] and (Case 3) a patent [34], work accepted for publication [30], 'YYYYb'-test for prolific author [31] and [32]. Other cites might contain 'duplicate' DOI and URLs (some SIAM articles) [22]. Boris / Barbara Beeton: multi-volume works as books [20] and [19].

A couple of citations with DOIs: [21, 22].

Online citations: [37–39].

We use jabref to manage all citations. A paper without managing a bib file will be returned without review. in the bibtex file all urls are added to rfernces with the *url* filed. They are not to be included in the *howpublished* or *note* field.

2.4 Theorem-like Constructs

Other common constructs that may occur in your article are the forms for logical constructs like theorems, axioms, corollaries and proofs. ACM uses two types of these constructs: theorem-like and definition-like.

Here is a theorem:

THEOREM 2.1. *Let f be continuous on $[a, b]$. If G is an antiderivative for f on $[a, b]$, then*

$$\int_a^b f(t)\,dt = G(b) - G(a).$$

Here is a definition:

Definition 2.2. If z is irrational, then by e^z we mean the unique number that has logarithm z:

$$\log e^z = z.$$

The pre-defined theorem-like constructs are **theorem**, **conjecture**, **proposition**, **lemma** and **corollary**. The pre-defined definition-like constructs are **example** and **definition**. You can add your own constructs using the *amsthm* interface [3]. The styles used in the `\theoremstyle` command are **acmplain** and **acmdefinition**.

Another construct is **proof**, for example,

PROOF. Suppose on the contrary there exists a real number L such that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L.$$

Then

$$l = \lim_{x\to c} f(x) = \lim_{x\to c} \left[gx \cdot \frac{f(x)}{g(x)} \right] = \lim_{x\to c} g(x) \cdot \lim_{x\to c} \frac{f(x)}{g(x)} = 0 \cdot L = 0,$$

which contradicts our assumption that $l \neq 0$. □

3 CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

Generated by bibtex from your `.bib` file. Run latex, then bibtex, then latex twice (to resolve references) to create the `.bbl` file. Insert that `.bbl` file into the `.tex` source file and comment out the command `\thebibliography`.

4 MORE HELP FOR THE HARDY

Of course, reading the source code is always useful. The file `acmart.pdf` contains both the user guide and the commented code.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the matlab code of the *BEPS* method.

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304 and Young Scientsts' Support Program (http://www.nnsf.cn/youngscientsts).

REFERENCES

[1] Rafal Ablamowicz and Bertfried Fauser. 2007. CLIFFORD: a Maple 11 Package for Clifford Algebra Computations, version 11. (2007). Retrieved February 28, 2008 from http://math.tntech.edu/rafal/cliff11/index.html

[2] Patricia S. Abril and Robert Plant. 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan. 2007), 36–44. https://doi.org/10.1145/1188913.1188915

[3] American Mathematical Society 2015. *Using the amsthm Package*. American Mathematical Society. http://www.ctan.org/pkg/amsthm

[4] Sten Andler. 1979. Predicate Path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN symposium on Principles of Programming Languages (POPL '79)*. ACM Press, New York, NY, 226–236. https://doi.org/10.1145/567752.567774

[5] David A. Anisi. 2003. *Optimal Motion Control of a Ground Vehicle*. Master's thesis. Royal Institute of Technology (KTH), Stockholm, Sweden.

[6] Mic Bowman, Saumya K. Debray, and Larry L. Peterson. 1993. Reasoning About Naming Systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (November 1993), 795–825. https://doi.org/10.1145/161468.161471

[7] Johannes Braams. 1991. Babel, a Multilingual Style-Option System for Use with LaTeX's Standard Document Styles. *TUGboat* 12, 2 (June 1991), 291–301.

[8] Malcolm Clark. 1991. Post Congress Tristesse. In *TeX90 Conference Proceedings*. TeX Users Group, 84–89.

[9] Kenneth L. Clarkson. 1985. *Algorithms for Closest-Point Problems (Computational Geometry)*. Ph.D. Dissertation. Stanford University, Palo Alto, CA. UMI Order Number: AAT 8506171.

[10] Jacques Cohen (Ed.). 1996. Special issue: Digital Libraries. *Commun. ACM* 39, 11 (Nov. 1996).

[11] Sarah Cohen, Werner Nutt, and Yehoshua Sagic. 2007. Deciding equivalances among conjunctive aggregate queries. *J. ACM* 54, 2, Article 5 (April 2007), 50 pages. https://doi.org/10.1145/1219092.1219093

[12] Bruce P. Douglass, David Harel, and Mark B. Trakhtenbrot. 1998. Statecarts in use: structured analysis and object-orientation. In *Lectures on Embedded Systems*, Grzegorz Rozenberg and Frits W. Vaandrager (Eds.). Lecture Notes in Computer Science, Vol. 1494. Springer-Verlag, London, 368–394. https://doi.org/10.1007/3-540-65193-4_29

[13] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

[14] Ian Editor (Ed.). 2008. *The title of book two* (2nd. ed.). University of Chicago Press, Chicago, Chapter 100. https://doi.org/10.1007/3-540-09237-4

[15] Matthew Van Gundy, Davide Balzarotti, and Giovanni Vigna. 2007. Catch me, if you can: Evading network signatures with web-based polymorphic worms. In *Proceedings of the first USENIX workshop on Offensive Technologies (WOOT '07)*. USENIX Association, Berkley, CA, Article 7, 9 pages.

[16] David Harel. 1978. *LOGICS of Programs: AXIOMATICS and DESCRIPTIVE POWER*. MIT Research Lab Technical Report TR-200. Massachusetts Institute of Technology, Cambridge, MA.

[17] David Harel. 1979. *First-Order Dynamic Logic*. Lecture Notes in Computer Science, Vol. 68. Springer-Verlag, New York, NY. https://doi.org/10.1007/3-540-09237-4

[18] Maurice Herlihy. 1993. A Methodology for Implementing Highly Concurrent Data Objects. *ACM Trans. Program. Lang. Syst.* 15, 5 (November 1993), 745–770. https://doi.org/10.1145/161468.161469

[19] Lars Hörmander. 1985. *The analysis of linear partial differential operators. III*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 275. Springer-Verlag, Berlin, Germany. viii+525 pages. Pseudodifferential operators.

[20] Lars Hörmander. 1985. *The analysis of linear partial differential operators. IV*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 275. Springer-Verlag, Berlin, Germany. vii+352 pages. Fourier integral operators.

[21] IEEE 2004. IEEE TCSC Executive Committee. In *Proceedings of the IEEE International Conference on Web Services (ICWS '04)*. IEEE Computer Society, Washington, DC, USA, 21–22. https://doi.org/10.1109/ICWS.2004.64

[22] Markus Kirschmer and John Voight. 2010. Algorithmic Enumeration of Ideal Classes for Quaternion Orders. *SIAM J. Comput.* 39, 5 (Jan. 2010), 1714–1747. https://doi.org/10.1137/080734467

[23] Donald E. Knuth. 1997. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms (3rd. ed.)*. Addison Wesley Longman Publishing Co., Inc.

[24] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY.

[25] Leslie Lamport. 1986. *LaTeX: A Document Preparation System*. Addison-Wesley, Reading, MA.

[26] Newton Lee. 2005. Interview with Bill Kinder: January 13, 2005. Video. *Comput. Entertain.* 3, 1, Article 4 (Jan.-March 2005). https://doi.org/10.1145/1057270.1057278

[27] Dave Novak. 2003. Solder man. Video. In *ACM SIGGRAPH 2003 Video Review on Animation theater Program: Part I - Vol. 145 (July 27–27, 2003)*. ACM Press, New York, NY, 4. https://doi.org/99.9999/woot07-S422

[28] Barack Obama. 2008. A more perfect union. Video. (5 March 2008). Retrieved March 21, 2008 from http://video.google.com/videoplay?docid=6528042696351994555

[29] Poker-Edge.Com. 2006. Stats and Analysis. (March 2006). Retrieved June 7, 2006 from http://www.poker-edge.com/stats.php

[30] Bernard Rous. 2008. The Enabling of Digital Libraries. *Digital Libraries* 12, 3, Article 5 (July 2008). To appear.

[31] Mehdi Saeedi, Morteza Saheb Zamani, and Mehdi Sedighi. 2010. A library-based synthesis methodology for reversible logic. *Microelectron. J.* 41, 4 (April 2010), 185–194.

[32] Mehdi Saeedi, Morteza Saheb Zamani, Mehdi Sedighi, and Zahra Sasanian. 2010. Synthesis of Reversible Circuit Using Cycle-Based Approach. *J. Emerg. Technol. Comput. Syst.* 6, 4 (Dec. 2010).

[33] S.L. Salas and Einar Hille. 1978. *Calculus: One and Several Variable*. John Wiley and Sons, New York.

[34] Joseph Scientist. 2009. The fountain of youth. (Aug. 2009). Patent No. 12345, Filed July 1st., 2008, Issued Aug. 9th., 2009.

[35] Stan W. Smith. 2010. An experiment in bibliographic mark-up: Parsing metadata for XML export. In *Proceedings of the 3rd. annual workshop on Librarians and Computers (LAC '10)*, Reginald N. Smythe and Alexander Noble (Eds.), Vol. 3. Paparazzi Press, Milan Italy, 422–431. https://doi.org/99.9999/woot07-S422

[36] Asad Z. Spector. 1990. Achieving application requirements. In *Distributed Systems* (2nd. ed.), Sape Mullender (Ed.). ACM Press, New York, NY, 19–33. https://doi.org/10.1145/90417.90738

[37] Harry Thornburg. 2001. Introduction to Bayesian Statistics. (March 2001). Retrieved March 2, 2005 from http://ccrma.stanford.edu/~jos/bayes/bayes.html

[38] TUG 2017. Institutional members of the TeX Users Group. (2017). Retrieved May 27, 2017 from http://wwtug.org/instmem.html

[39] Boris Veytsman. [n. d.]. acmart—Class for typesetting publications of ACM. ([n. d.]). Retrieved May 27, 2017 from http://www.ctan.org/pkg/acmart

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Lars Thørväld
The Thørväld Group
1 Thørväld Circle
Hekla, Iceland
larst@affiliation.org

Valerie Béranger
Inria Paris-Rocquencourt
Rocquencourt, France

Aparna Patel
Rajiv Gandhi University
Rono-Hills
Doimukh, Arunachal Pradesh, India

Huifen Chan
Tsinghua University
30 Shuangqing Rd
Haidian Qu, Beijing Shi, China

Charles Palmer
Palmer Research Laboratories
8600 Datapoint Drive
San Antonio, Texas 78229
cpalmer@prl.com

John Smith
The Thørväld Group
jsmith@affiliation.org

Julius P. Kumquat
The Kumquat Consortium
jpkumquat@consortium.net

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

ACM proceedings, LaTeX, text tagging

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

Big Data for Edge Computing

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Gregor von Laszewski
Indiana University
Smith Research Center
Bloomington, IN 47408, USA
laszewski@gmail.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

Big Data, Edge Computing i523

1 INTRODUCTION

Put here an introduction about your topic. We just need one sample refernce so the paper compiles in LaTeX so we put it here [1].

2 CONCLUSION

Put here an conclusion. Conlcusions and abstracts must not have any citations in the section.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4

My great Big Dat Paper

Ben Trovato
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
trovato@corporation.com

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

KEYWORDS

i523

1 INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes, a specified live area, centered on the page, specified size of margins, specified column width and gutter size [1].

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

[1] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. https://doi.org/10.1007/3-540-09237-4