Use Cases in Big Data Software and Analytics

Vol. 1, Fall 2017

Bloomington, Indiana

Saturday 4th November, 2017, 09:46

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

Contents

1 Preface **9**
1.0.1 Disclaimer 9
1.0.2 Citation 9
1.1 List of Papers 10

2 Biology **13**

3 Business **13**

4 Edge Computing **13**

5 Education **13**

6 Energy **13**

7 Environment **13**
2 hid231 Status: 90%
Using Big Data to Battle Air Pollution
Vegi, Karthik 13

8 Government **23**

9 Health **23**

10 Lifestyle **23**
3 hid332 Status: 97%
Big Data Analytics in Developing Countries
Judy Phillips 23
4 hid347 Status: 0%
Sociological Methods of Big Data
Jeramy Townsley 23

11 Machine Learning **23**

12 Media **23**
5 hid213 Status: Oct 28 2017 50%
Big Data and Face Identification
Yuchen Liu 23

6 hid336 Status: 0%
Big Data Analysis for Computer Network Defense
Jordan Simmons 23

13 Physics **29**

14 Security **29**
7 hid316 Status: 99%
Big Data on IoT Smart Refrigerators
Robert Gasiewicz 29
8 hid329 Status: 33%
Big Data Analytics and the Impact on Personal Privacy
Ashley Miller 40

15 Sports **40**

16 Technology **40**
9 hid233 Status: 50%
Big Data Applications in Virtual Assistants
Wang, Jiaan 40
10 hid306 Status: 85%
Why Deep Learning matters in IoT Data Analytics?
Murali Cheruvu 45

17 Text **56**

18 Theory **56**

19 Transportation **56**

20 TBD **56**
11 hid101 Status: not yet started
Benchmarking a BigData Docker deployment
Huiyi Chen 56
12 hid102 Status: unkown
Benchmarking a BigData Docker deployment
Gregor von Laszewski 56
13 hid104 Status: 5%
Big Data = Big Bias? Ethical Challenges of Big Data
Jones, Gabriel 56
14 hid105 Status: unkown
Benchmarking a BigData Docker deployment
Gregor von Laszewski 56
15 hid106 Status: 0%
Benchmarking a BigData Docker deployment
Qiaoyi Liu 66

16 hid107 Status: unkown
Benchmarking a BigData Docker deployment
Gregor von Laszewski 66
17 hid109 Status: 0%
Big Data and Application in Amazon
Shiqi Shen 66
18 hid111 Status: unkown
Benchmarking a BigData Docker deployment
Gregor von Laszewski 77
19 hid201 Status: not started
None
Arnav, Arnav 77
20 hid202 Status: 0%
This is my paper about the other abc
Himani Bhatt 87
21 hid204 Status: 20%
Big Data and Support Vector Machines
Chaturvedi, Dhawal 87
22 hid205 Status: 0%
This is my paper about the other abc
Chaudhary Mrunal L 87
23 hid208 Status: unkown
Algorithms for Big Data Analysis
Jyothi Pranavi Devineni 87
24 hid211 Status: unkown
Machine learning optimizations for big data
Khamkar, Ajinkya 87
25 hid212 Status: unkown
Not yet decided
Kumar, Saurabh 87
26 hid214 Status: 0%
Big Data and League of Legend
Gregor von Laszewski 87
27 hid215 Status: yet to start
to be decided
Mallala, Bharat 98
28 hid216 Status: not started
n/a
Millard, Mathew 98
29 hid218 Status: 0%
How Big Data Transform Education
Niu, Geng 98
30 hid219 Status: unkown
Benchmarking a BigData Docker deployment
Gregor von Laszewski 108

31 hid224 Status: 0%
Big Data Applications in the Energy and Utilities Sector
Rawat, Neha 108
32 hid225 Status: not started
...
Schwartzer, Matthew 115
33 hid228 Status: 0%
TBD
Swargam, Prashanth 115
34 hid229 Status: not yet started
TBD
ZhiCheng Zhu 120
35 hid230 Status: unkown
Big data with natural language processing
YuanMing Huang 120
36 hid232 Status: 0%
This is my paper about the other abc
Gregor von Laszewski 120
37 hid234 Status: 10%
Big Data and Edge Computing in Health Informatics for People with Disabilities.
Weixuan Wang 120
38 hid235 Status: 0%
Big data: An Opportunity for Historians
Yujie Wu 120
39 hid236 Status: not started
Benchmarking a BigData Docker deployment
Weipeng Yang 131
40 hid237 Status: 0%
Benchmarking a BigData Docker deployment
Gregor von Laszewski 131
41 hid301 Status: 10%
Prediction of psychological traits based on Big Data classification of associated social media footprints
Gagan Arora 142
42 hid302 Status: 30%
Hadoop and MongoDB in support of Big Data Applications and Analytics
Sushant Athaley 159
43 hid304 Status: 0%
Big Data and Analytics in Deep Space Telemetry and Navigation
Ricky Carmickle 174
44 hid305 Status: 0%
Big Data applied to zoning and city planning.
Andres Castro Benavides 174
45 hid308 Status: 0%
Parallel Computing and Big Data
Pravin Deshmukh 174

46 hid311 Status: 0%
Benchmarking a BigData Docker deployment
Gregor von Laszewski 174
47 hid312 Status: 15%
To be decided
Neil Eliason 174
48 hid313 Status: 25%
Big Data Applications in Laboratories
Tiffany Fabianac 182
49 hid314 Status: 0%
Benchmarking a BigData Docker deployment
Gregor von Laszewski 182
50 hid315 Status: 0%
Big Data Opportunity and Challenges with Smart Helmets
Garner, Jeffry 182
51 hid318 Status: 0%
Benchmarking a BigData Docker deployment
Gregor von Laszewski 182
52 hid319 Status: 0%
Mini Project: ESP8266 and Raspberry PI Robot Car
Mani Kumar Kagita 182
53 hid320 Status: 50%
Overview of Python Data Visualization Tools
Elena Kirzhner 182
54 hid321 Status: unkown
Benchmarking a BigData Docker deployment
Gregor von Laszewski 182
55 hid323 Status: unkown
None
Uma M Kugan 182
56 hid324 Status: unkown
TBD
Ashok Kuppuraj 182
57 hid326 Status: unkown
Big data on autonomous cars
Mohan Mahendrakar 182
58 hid328 Status: 70%
Big data analytics in data center network monitoring
Dhanya Mathew 182
59 hid330 Status: 35%
MQTT for Big Data and Edge Computing
Janaki Mudvari Khatiwada 182
60 hid331 Status: 10%
Big Data Applications in Using Neural Networks for Medical Image Analysis
Tyler Peterson 182

61 hid333 Status: 0%
Nartural language processing (NLP) for speech analysis and voice recognition
Ashok Reddy Singam, Anil Ravi 182
62 hid334 Status: 90%
Advancements in Drone Technology for the US Military
Peter Russell 182
63 hid335 Status: 90%
Big Health Data from Wearable Electronic Sensors (WES) and the Treatment of Opioid Addiction
Sean M. Shiverick 189
64 hid337 Status: 0%
Natural Language Processing (NLP) to analyze human speech data
Ashok Reddy Singam, Anil Ravi 209
65 hid338 Status: 0%
Benchmarking a BigData Docker deployment
Gregor von Laszewski 215
66 hid339 Status: 0%
Benchmarking a BigData Docker deployment
Hady Sylla 215
67 hid340 Status: 0%
Big data on the blockchain? Distributed networks and large-scale analytics
Timothy A. Thompson 215
68 hid341 Status: 0%
This is my paper about the other abc
Tibenkana, Jacob 215
69 hid342 Status: 0%
Still under consideration
Udoyen, Nsikan 215
70 hid345 Status: unkown
Big Data Analytics and influence on althetics.
Ross Wood 215
71 hid346 Status: unkown
This is my paper about the other abc
Gregor von Laszewski 215
72 hid348 Status: 30%
Security aspect of NOSQL database in Big Data Applications
Budhaditya Roy 215

Chapter 1

Preface

1.0.1 Disclaimer

The papers provided are contributed by students of the i523 class thought at Indiana University in Fall of 2017. The students were educated in plagiarizm and we hope that all papers meet the high standrads provided by the policies set at Indiana University in regrads to plagiarizm. In case you notice any issues, please contact Gregor von Laszewski (laszewski@gmail.com) so we cn address the issue with the student.

1.0.2 Citation

The proceedings is at this time available as a draft. To cite this proceedings you can use the following citation entry:

```
@Book{las17-i523,
  editor =    {Gregor von Laszewski},
  title =        {Use Cases in Big Data Software and Analytics},
  publisher =    {Indiana University},
  year =         {2017},
  volume =    {1},
  series =    {i523},
  address =   {Bloomington, IN},
  edition =   {1},
  month =     dec,
  url={https://github.com/laszewski/laszewski.github.io/raw/master/papers/vonLaszewski-i
}
```

Contributors to the volume can cite their contribution as follows. They just need to *FILLIN* the missing information

```
@InBook{las17-,
  author =    {FILLIN},
```

```
  editor =    {Gregor von Laszewski},
  title =        {Use Cases in Big Data Software and Analytics},
  chapter =      {FILLIN},
  publisher =    {Indiana University},
  year =         {2017},
  volume =    {1},
  series =    {i523},
  address =   {Bloomington, IN},
  edition =   {1},
  month =     dec,
  url={https://github.com/laszewski/laszewski.github.io/raw/master/papers/vonLaszewski-i
  pages =     {FILLIN},
}
```

1.1 List of Papers

HID	Author	Title
101	Huiyi Chen	Benchmarking a BigData Docker deployment
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
104	Jones, Gabriel	Big Data = Big Bias? Ethical Challenges of Big Data
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
106	Qiaoyi Liu	Benchmarking a BigData Docker deployment
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
109	Shiqi Shen	Big Data and Application in Amazon
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
201	Arnav, Arnav	None
202	Himani Bhatt	This is my paper about the other abc
hid203	error: yaml	This is my paper about the other abc
204	Chaturvedi, Dhawal	Big Data and Support Vector Machines
205	Chaudhary Mrunal L	This is my paper about the other abc
208	Jyothi Pranavi Devineni	Algorithms for Big Data Analysis
209	Han, Wenxuan	Clustering Algorithms in Big Data Analysis
hid210	error: yaml	Clustering Algorithms in Big Data Analysis
211	Khamkar, Ajinkya	Machine learning optimizations for big data
212,	Kumar, Saurabh	Not yet decided
213	Yuchen Liu	Big Data and Face Identification
0	Gregor von Laszewski	Big Data and League of Legend
215	Mallala, Bharat	to be decided
216	Millard, Mathew	n/a
0	Niu, Geng	How Big Data Transform Education
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
224	Rawat, Neha	Big Data Applications in the Energy and Utilities Sector
225	Schwartzer, Matthew	...
228	Swargam, Prashanth	TBD
229	ZhiCheng Zhu	TBD
230	YuanMing Huang	Big data with natural language processing
231	Vegi, Karthik	Using Big Data to Battle Air Pollution

0	Gregor von Laszewski	This is my paper about the other abc
233	Wang, Jiaan	Big Data Applications in Virtual Assistants
234	Weixuan Wang	Big Data and Edge Computing in Health Informatics for People with Disabilities.
235	Yujie Wu	Big data: An Opportunity for Historians
236	Weipeng Yang	Benchmarking a BigData Docker deployment
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
301	Gagan Arora	Prediction of psychological traits based on Big Data classification of associated social media footprints
302	Sushant Athaley	Hadoop and MongoDB in support of Big Data Applications and Analytics
304	Ricky Carmickle	Big Data and Analytics in Deep Space Telemetry and Navigation
305	Andres Castro Benavides	Big Data applied to zoning and city planning.
306	Murali Cheruvu	Why Deep Learning matters in IoT Data Analytics?
308	Pravin Deshmukh	Parallel Computing and Big Data
hid309	error: yaml	Parallel Computing and Big Data
hid310	error: yaml	Parallel Computing and Big Data
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
312	Neil Eliason	To be decided
313	Tiffany Fabianac	Big Data Applications in Laboratories
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
315	Garner, Jeffry	Big Data Opportunity and Challenges with Smart Helmets
316	Robert Gasiewicz	Big Data on IoT Smart Refrigerators
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
319	Mani Kumar Kagita	Mini Project: ESP8266 and Raspberry PI Robot Car
320	Elena Kirzhner	Overview of Python Data Visualization Tools
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
323	Uma M Kugan	None
324	Ashok Kuppuraj	TBD
325	J. Robert Langlois	The importance of data sharing and the replication of the sciences
326	Mohan Mahendrakar	Big data on autonomous cars
327	Paul Marks	The Impact of Self-Driving Cars on the Economy
328	Dhanya Mathew	Big data analytics in data center network monitoring
329	Ashley Miller	Big Data Analytics and the Impact on Personal Privacy
330	Janaki Mudvari Khatiwada	MQTT for Big Data and Edge Computing
331	Tyler Peterson	Big Data Applications in Using Neural Networks for Medical Image Analysis
332	Judy Phillips	Big Data Analytics in Developing Countries
337, 333	Ashok Reddy Singam, Anil Ravi	Nartural language processing (NLP) for speech analysis and voice recognition
334	Peter Russell	Advancements in Drone Technology for the US Military
335	Sean M. Shiverick	Big Health Data from Wearable Electronic Sensors (WES) and the Treatment of Opioid Addiction
336	Jordan Simmons	Big Data Analysis for Computer Network Defense
337, 333	Ashok Reddy Singam, Anil Ravi	Natural Language Processing (NLP) to analyze human speech data
0	Gregor von Laszewski	Benchmarking a BigData Docker deployment
0	Hady Sylla	Benchmarking a BigData Docker deployment
340	Timothy A. Thompson	Big data on the blockchain? Distributed networks and large-scale analytics
341	Tibenkana, Jacob	This is my paper about the other abc

342	Udoyen, Nsikan	Still under consideration
343	Borga Edionse Usifo	Big Data Applications and Manufacturing
345	Ross Wood	Big Data Analytics and influence on althetics.
0	Gregor von Laszewski	This is my paper about the other abc
347	Jeramy Townsley	Sociological Methods of Big Data
348	Budhaditya Roy	Security aspect of NOSQL database in Big Data Applications

Using Big Data to Battle Air Pollution

Karthik Vegi
Indiana University Bloomington
2619 East 2nd Street, Apt 11
Bloomington, IN 47401, USA
kvegi@iu.com

ABSTRACT

We have come a long way from the stone age to build large scale industries, big cities, bullet trains, and a booming automobile industry. Technological and industrial advances are making our cities smarter by the day and yet a nagging side-effect is air pollution. Air pollution is not only creating local health hazards like respiratory and heart problems, but also directly leading to an increase in temperatures and contributing to global warming. We show how the advances in *Big Data*, *Cloud Computing*, and *Internet Of Devices* can be used to combat air pollution.

KEYWORDS

i523, hid231, big data, environment, air pollution, global warming

1 INTRODUCTION

Air pollution is no longer a local problem. It is a global environmental issue which involves individual countries to come together and device measures to combat it [5]. It it causing about 3.7 million premature deaths worldwide from cardiovascular and respiratory diseases and also ruins the crops that feed the world [5]. Air pollution also has a direct effect on a number of environmental issues like global warming, depletion of ozone layer, acid rains and impacts wild-life [5].

Back in the year 1990, the job of a typical air quality scientist was to develop atmospheric dispersion models to evaluate the air pollution caused by industries and make sure that it is within the permissible level suggested by the *Environmental Protection Agency* [2]. These models gather historic data of many years from airports and weather balloons to predict the pollution with the help of meteorology theory [2]. Although the methods used to derive the values were good enough, the limitations with respect to the technology posed a real challenge which took weeks to run the simulations only to be cut-off in the middle due to power and storage issues [2]. The data processing engine was built on Sun-Solaris workstations with tapes handling the data storage [2]. The work-stations set up in major points in the country would communicate using a very slow network connection [2]. The data processing would be done locally and later written to all the servers which would then be split and distributed among many machines and consolidated in the end [2]. "If only we had that much more data and that much more ability to handle it, we could iterate through the model at a much finer scale. Real-time data processing remained a pipe dream" [2].

2 AIR POLLUTION AS A BIG DATA PROBLEM

The advent of *Big Data* and the technological advances changed the way the data is ingested and analyzed [2]. The network speeds have increased, wide range of sensors are available to collect data with a lot of precision which would feed the high speed data processing systems. Batch processing has become easier with *Hadoop* and *Map-Reduce*. The storage mechanisms have become cheaper and more disaster proof.

IBM is helping Beijing combat air pollution by analyzing huge amounts of data using a data analysis platform *Green Horizons* [3]. IBM has signed up partnerships with different cities in China and India to deploy *Big Data Analytics*, *Machine Learning* and *Internet Of Things* to improve traffic, keep a check on the pollution from industrial machines and other pollution causing agents [3]. *IBM* will deploy sensors in various place to collect data in real-time and analyze previous weather forecasts, and build improved iterative models over time [3]. The system continuously streams data form the sensors and improves the forecast by learning over time using *Machine Learning* algorithms [3]. Figure 1 shows the Green Horizons air quality management for Beijing.

[Figure 1 about here.]

IBM is collaborating with the United Nations to push the use of technological advances by every country for the common good of the world [3]. More and more cities and countries are opening air quality data to public where you can get reports in real time [1]. The *BreezoMeter* is the first mobile application that provides real-time information of the street's air quality information using geo-location maps [1]. *Copernicus* is another monitoring service that ingests data from satellites and on site sensors on land, air and sea to provide continuous information to the users [1]. *Open Data Week* is an intergovernmental organization where 34 states come together to bring reforms and discuss how to use technology and services like *Copernicus* that use *Big Data* to test prototypes of new products to ensure they operate within the permissible levels of pollution [1].

While these initiatives help bring awareness about the seriousness of the issue, each state and country should take strict measures to bring out reforms that will help eradicate pollution. *Big Data* might never replace the environmental responsibility but it will help to plan the vision for environmental awareness and its tools make it easier to achieve the vision [1]. These tools can also be used to gauge the alternative sources of energy and the feasibility of tapping into other natural resources ensuring responsible consumption of energy [1]. For example, *IBM Bluemix* analyzed data from a steel industry and the analysis uncovered an interesting insight that the furnace wastes a lot of energy to offset the temperature of the smoke which resulted in optimizing its operation [2]

3 BIG DATA TECHNIQUES TO COMBAT POLLUTION

3.1 Random Forest Approach for predicting air quality in Urban Sensing Systems

Air pollution in an urban setting is very important to monitor because of the population density. Air quality in these areas varies a lot in various parts of the city owing to traffic and presence of industries [6]. A random forest approach ingests data from meteorology, urban sensors, road information and real-time traffic and predicts the air quality with utmost precision [6]. Real-time air quality information consists of measuring the concentration of $PM_{2.5}$, PM_{10}, and NO_2 [6].

The *Air Quality Index* AQI is the measure that is used to understand how polluted the air is [6]. AQI is measured by reading the concentration of 6 pollutant gases namely, sulfur dioxide SO_2, nitrogen dioxide NO_2, air particles smaller than 10 μm PM_{10}, air particles smaller than 2.5 μm $PM_{2.5}$, carbon monoxide CO, and ozone O_3 [6]. Based on the level of AQI, the air quality is classified as shown in Figure 2.

[Figure 2 about here.]

Traffic Congestion Status TCS, explains the traffic status at the current hour [6]. Figure 3 shows how colors are used to represent the traffic congestion [6].

[Figure 3 about here.]

3.2 RAQ Algorithm

The RAQ algorithm collects data from air monitoring station AQI, meteorology data MD, traffic congestion TCS, road information RI, and point of interest POI which is the specific location that someone is interested to visit [6]. The data refresh rate is one hour and the data is collected from different parts of the city which are divided in grids from G_1 to G_n [6]. The data is divided into training and testing data sets to train the model and evaluate the model [6]. Figure 4 shows the structure of the data.

[Figure 4 about here.]

A decision tree is used to split and classify the data and the results are aggregated by collecting the data from all the sub-trees [6]. Figure 5 illustrates the procedure of RAQ.

[Figure 5 about here.]

The *Random Forest* algorithm is employed using the tree type classifier to recursively partition the dataset and generate sub-trees and finally aggregate the results of each sub-tree [6]. Each sub-tress is constructed using *Bootstrap Aggregating* where each data set is divided into different buckets by using statistical samples [6]. Once the trees are constructed, each subset of data is fed into a decision tree and the estimated AQI index is calculated [6]. The final AQI index is determined as the maximum value out of all the individual values [6]. Figure 6 shows the step-by-step RAQ algorithm.

[Figure 6 about here.]

4 MACHINE LEARNING MODELS

Machine Learning deals with augmenting computers with the ability to learn from data and program themselves [4]. These algorithms can be used to evaluate the air quality [4].

4.1 Artificial Neural Network Model

Artificial Neural Network Model tries to solve the problem by simulating the functioning of brain and neurons [4]. The model architecture is a function of a sigmoid [4]. For this experiment, the air quality data was divided into training, test, and validation data with split of 60, 20, and 20 with a back propagation network of two hidden layers [4]. To ensure consistency, the air quality data for the training and test sets are derived from the same season [4]. The air quality is forecast by looking at the historic data where the input and output are represented by the air quality data measured at different times [4]. The model turns out to be reliable with a good prediction accuracy with the lowest mean square error of 3.7×10^{-4} [4]. The Artificial Neural Network Model is combined with *Markov Chains* to develop a new improved model with improved prediction accuracy where the ANN computes the primary values and the results are re-computed and improved by the markov transitional probability matrices [4]. Figure 7 shows the Artificial Neural Network Model with two hidden layers.

[Figure 7 about here.]

4.2 Least squares Support Vector Machine Model

Least squares support vector machine is a supervised learning model used for classification and regression analysis which arrives at the solution by solving the data represented in the form of linear equations [4]. For this model, the sample data was collected from 100 sensor points in different intervals of time and at different geographical locations that ranged from urban areas with population, areas near the airport, water surface areas like lakes, and sewage processing areas [4]. The sample data was a good split with 80 percent collected from urban sewage area and the other data collected from air surface areas [4]. The fluorescence content in the air was analyzed by a portable air quality measuring device developed in-house by Zhejiang University [4]. The fluorescence data captured using the device is highly dimensional and non-linear and therefore data pre-processing is essential to bring the dimensions down to a manageable level [4]. This eliminates the ambient noise and the temperature drift from the data [4]. The algorithm predicts the regression model by looking at the training data for each cluster [4]. Finally, the vector consine distance is used to classify the sample into clusters and the performance criterion such as *Root Mean Square Error* and *Mean Absolute Error* are computed which demonstrate the efficiency of the algorithm [4]. Figure 8 shows the pictorial representation of the algorithm.

[Figure 8 about here.]

5 CONCLUSION

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski and the teaching assistants for their support and suggestions in writing this paper.

REFERENCES

[1] Ferrovial Blog. 2017. Big data will control pollution in your city. Webpage. (April 2017). http://blog.ferrovial.com/en/2017/04/big-data-pollution-control-in-cities/

[2] Jay Hardikar. 2017. Environmental analysis in the era of cloud and big data platforms. Webpage. (Jan. 2017). https://www.ibm.com/blogs/bluemix/2017/01/environmental-analysis-era-cloud-big-data-platforms/
[3] Alexander Howard. 2015. How IBM Is Using Big Data To Battle Air Pollution In Cities. Webpage. (Sept. 2015). https://www.ibm.com/blogs/bluemix/2017/01/environmental-analysis-era-cloud-big-data-platforms/
[4] Gaganjot Kaur Kang, Jerry Gao, Sen Chiao, Shengqiang Lu, and Gang Xie. 2017. Air Quality Prediction: Big data and Machine Learning Approaches. *International Conference on Sustainable Environment and Agriculture* 1 (10 2017).
[5] Research Applications Laboratory. 2016. Air Pollution: A Global Problem. Webpage. (April 2016). https://ral.ucar.edu/pressroom/features/air-pollution-a-global-problem
[6] Ruiyun Yu, Yu Yang, Leyou Yang, Guangjie Han, and Oguti Ann Move. 2016. RAQ: A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems. *Sensors* 16 (2016), 1–86. http://www.mdpi.com/1424-8220/16/1/86

List of Figures

1 Green Horizons air quality management for Beijing [3] 5
2 AQI classification [6] 5
3 Traffic Congestion[6] 5
4 Structure of RAQ data[6] 6
5 Procedure of RAQ [6] 6
6 RAQ Algorithm [6] 6
7 Artificial Neural Network(ANN) Model [4] 6
8 Least squares Support Vector Machine Model [4] 7

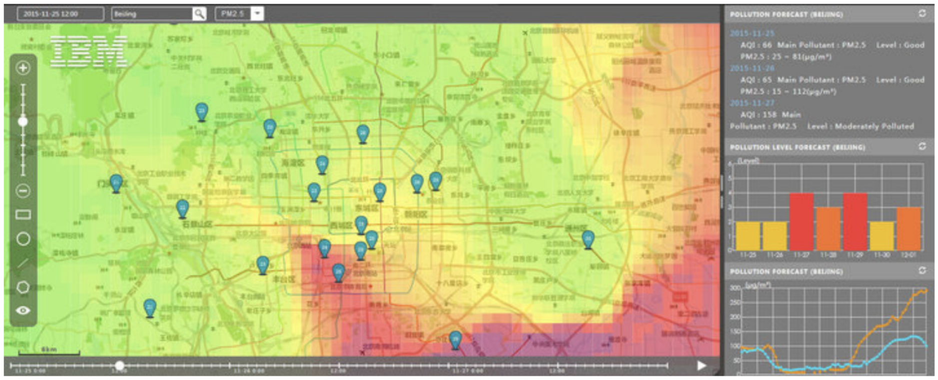


Figure 1: Green Horizons air quality management for Beijing [3]

AQI	Air Pollution Level
0–50	Excellent
51–100	Good
101–150	Lightly Polluted
151–200	Moderately Polluted
201–300	Heavily Polluted
300+	Severely Polluted

Figure 2: AQI classification [6]

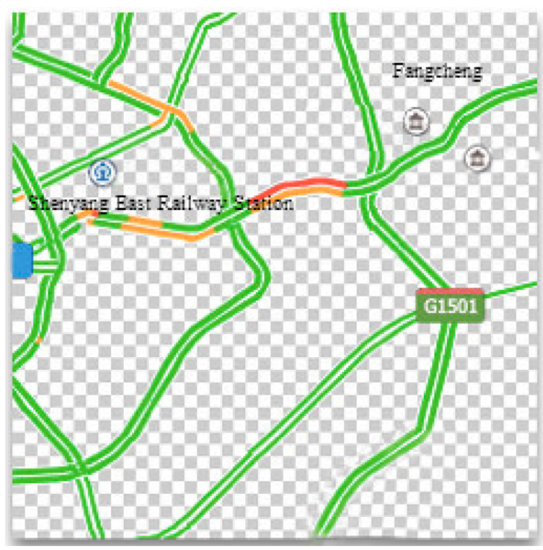


Figure 3: Traffic Congestion[6]

temperature	humility	pressure	wind	visibility	road_length	tfs	poi_number	aqi
Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Nominal
5.5	89.0	758.1	2.0	14.0	2185.0	2371.0	63.0	excellent

Figure 4: Structure of RAQ data[6]

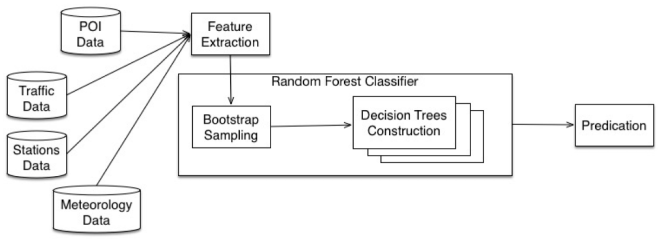


Figure 5: Procedure of RAQ [6]

Algorithm 1. RAQ	
Input:	A dataset S with features: F_{mt}, F_{mh}, F_{mp}, F_{mw}, F_{mv}, F_{ri}, F_{tcs}, F_{pn} and labeled AQI level; unlabeled dataset U; trees quantity T; features quantity m;
Output:	AQI level
	for T trees
	randomly select m features from S;
	for m features in each node
	calculate information gain by Equation (3);
	choose maximum gain to split the dataset in the node;
	remove used feature from feature candidates;
	input unlabeled data into trees;
	get predicted AQI level according to Equations (5) and (6);

Figure 6: RAQ Algorithm [6]

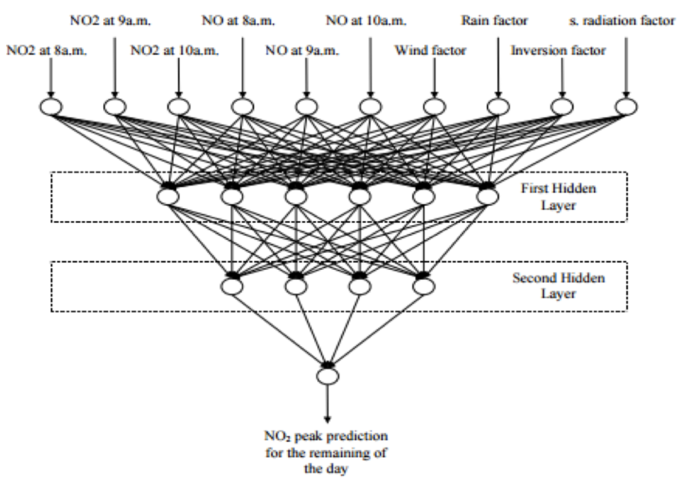


Figure 7: Artificial Neural Network(ANN) Model [4]

6

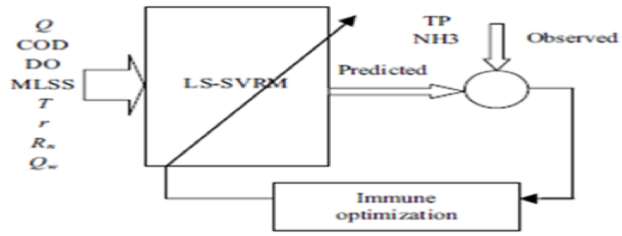


Figure 8: Least squares Support Vector Machine Model [4]

7

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib

bibtext _ label error
============================================


bibtext space label error
============================================


bibtext comma label error
============================================


latex report
============================================

        [2017-11-04 09.45.23] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
Typesetting of "report.tex" completed in 1.7s.
../README.yml
  15:27     error    trailing spaces  (trailing-spaces)


=================================================================================
Compliance Report
=================================================================================


name: Vegi, Karthik
hid:  231
paper1: Oct 29 17 100%
paper2: 90%

yamlcheck
---------------------------------------------------------------------------------
```

```
wordcount
--------------------------------------------------------------------------------

7
wc 231 paper2 7 565 content.tex
wc 231 paper2 7 1990 report.pdf
wc 231 paper2 7 257 report.bib

find "
--------------------------------------------------------------------------------

102: Do not use "these quotes" but use these ``these quotes''.

passed: False

find footnote
--------------------------------------------------------------------------------

112: \footnote{do not use footnotes}.

passed: False

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

62: In Figure \ref{f:fly} we show a fly. Please note that because we
    use
68: \begin{figure}[!ht]
69: \centering\includegraphics[width=\columnwidth]{images/fly.pdf}
70: \caption{Example caption}\label{f:fly}
85: or generate them by hand while using the provided template in
    Table\ref{t:mytable}. Not ethat
88: \begin{table}[htb]
91: \label{t:mytabble}

figures 1
tables 1
includegraphics 1
```

```
labels 2
refs 2
floats 2

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
105: Do not use Figure 1 user the ref for the figure while using its
     label
passed: False -> labels or refs used wrong

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
```

```
--------------------------------------------------------------------------------

passed: True

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Big Data Analysis for Computer Network Defense

Jordan Simmons
Indiana University
Smith Research Center
Bloomington, IN 47408, USA
jomsimm@iu.edu

ABSTRACT

Computer security threats and attacks are constantly evolving. Everyday, hackers are creating new techniques to bypass network security for the purpose of malicious attacks. To keep up with the changing intrusion technologies, the technologies that defend these attacks need to constantly evolve also. Modern day technologies use deep learning techniques to monitor network activity, and detect malicious code. We will provide an overview of network security and modern technologies being used to protect computer systems and networks.

KEYWORDS

i523,HID336, Computer Network Security, Big Data Analysis, Deep Learning, Intrusion Detection Systems,

1 INTRODUCTION

Everyday a different computer network is being breached with the intent to cause harm to the system or to steal valuable data. Computer hackers are constantly creating new ways to evade network security and create malicious code that can not be detected by security systems. As malicious technologies continue to advance, the technologies that defend against these technologies need to adapt with these advances. The problem with computer network defence is that the technologies used to breach systems constantly change. Once a solution is created to defend a technology, a new malicious technology could be created the next day. Today many security specialist are using deep learning technologies to monitor network intrusions, and detect malicious code. In order to better understand computer network defense, an overview of modern attacks, network data collection processes, and the technologies used to analyze network data is provided.

2 DATA COLLECTION

2.1 Network Intrusion Data Collection

2.2 Malware Data Collection

3 DEEP LEARNING FOR NETWORK INTRUSIONS

4 DEEP LEARNING ON MALWARE

5 CONCLUSION

ACKNOWLEDGMENTS

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

REFERENCES

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

- Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure
- Do not use the word *I* instead use *we* even if you are the sole author
- Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned
- If you want to say *and* do not use *&* but use the word *and*
- Use a space after . , :
- When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

- It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class
- Claims made without citations provided
- Need to paraphrase long quotations (whole sentences or longer)
- Need to quote directly cited material

A.6 Character Errors

- Erroneous use of quotation marks, i.e. use “quotes” , instead of " "
- To emphasize a word, use *emphasize* and not “quote”
- When using the characters & # % _ put a backslash before them so that they show up correctly
- Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.
- If you see a ﬁgure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

- Acknowledgement section missing
- Incorrect README file
- In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper
- The paper has less than 2 pages of text, i.e. excluding images, tables and figures
- The paper has more than 6 pages of text, i.e. excluding images, tables and figures
- Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

- Capitalization errors in referring to captions, e.g. Figure 1, Table 2
- Do use *label* and *ref* to automatically create figure numbers
- Wrong placement of figure caption. They should be on the bottom of the figure
- Wrong placement of table caption. They should be on the top of the table
- Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg
- Do not submit eps images. Instead, convert them to PDF
- The image files must be in a single directory named ”images”
- In case there is a powerpoint in the submission, the image must be exported as PDF
- Make the figures large enough so we can read the details. If needed make the figure over two columns
- Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.
- In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption
- Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)
- Do not use textwidth as a parameter for includegraphics
- Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

```
bibtext report
===========================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
I found no \citation commands---while reading file report.aux
Database file #1: report.bib
(There was 1 error message)
make[2]: *** [bibtex] Error 2

latex report
===========================================

        [2017-11-04 09.46.30] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
Empty ‘thebibliography’ environment.
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Typesetting of "report.tex" completed in 1.3s.
../README.yml
  8:81      error    line too long (84 > 80 characters)  (line-length)
  9:81      error    line too long (85 > 80 characters)  (line-length)
  10:81     error    line too long (82 > 80 characters)  (line-length)
  11:81     error    line too long (81 > 80 characters)  (line-length)
  12:52     error    trailing spaces  (trailing-spaces)
  25:81     error    line too long (82 > 80 characters)  (line-length)
  25:82     error    trailing spaces  (trailing-spaces)
  28:79     error    trailing spaces  (trailing-spaces)
  30:62     error    trailing spaces  (trailing-spaces)
  32:79     error    trailing spaces  (trailing-spaces)


================================================================================
Compliance Report
================================================================================


name: Jordan Simmons
hid:  336
paper1: Oct 25 17
```

```
paper2: In Progress

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

2
wc 336 paper2 2 457 report.tex
wc 336 paper2 2 1097 report.pdf
wc 336 paper2 2 50 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------


figures 0
tables 0
includegraphics 0
labels 0
refs 0
floats 0

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)
```

```
Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
I found no \citation commands---while reading file report.aux
Database file #1: report.bib
(There was 1 error message)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

15: note            = "",

passed: False

ascii
--------------------------------------------------------------------------------
```

```
================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Big Data Analytics and IoT Smart Refridgerators

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

ABSTRACT

The intent of this paper is to explore the rapid growth of IoT Smart Appliances, specifically with regard to refrigerators. As more devices are connected to the internet, to each other, and become readily available to consumers, there are many exciting new possibilities that offer both convenience and to make our lives more efficient. The scope of this paper will begin with a brief history of IoT, then move on to describe the current way in which this technology is being applied, and conclude with exploration and outlook on future development possibilities as well as potential risks.

KEYWORDS

i523, HID316, Big Data, IoT, Refrigerators, Smart Appliances, M2M, Samsung, Innit, Instacart, GrubHub

1 INTRODUCTION

The advent of the Internet of Things (IoT) began at the close of the last millennium when the world began connecting ordinary devices - electronics other than traditional computers - to the internet. With virtually unlimited possibilities, the unthinkable became reality when the concept of putting a wireless network card in a refrigerator went mainstream. Initial features were as simple as a large touchscreen with the news, the weather, and a doodle board.

From there, IoT Smart Refrigerators have evolved to become equipped with cameras, cooking recommendations, and even rudimentary food inventory and spoilage management systems. Now that food delivery services such as Instacart and GrubHub have become popular, there are already plans to integrate these services with smart refrigerators. As IoT has continued to expand throughout the marketplace and the concept of machine-to-machine (M2M) IoT has taken hold, there are now even more possibilities, which means a bright future in the kitchen no matter if you're an aspiring chef, a person trying to efficiently manage a family, or someone with specific health needs. However, along with the rapid advance of new features, there are also significant threats and blind spots with security.

2 EARLY HISTORY OF IOT AND NETWORKED APPLIANCES

Although the internet didn't yet exist in the minds of Hollywood producers in 1985, the opening scene from Back to the Future begins with a room full of ticking clocks, one of which is an alarm clock that rings and sets off a Rube Goldberg machine that has been configured by Doc Brown to automate the preparation of his breakfast. It's not unreasonable to believe that, in his many time travel escapades, Doc would've eventually *discovered* the internet and would've upgraded this rudimentary appliance.

That reality wouldn't come until five years later in 1990 when the first IoT device, a toaster, was turned off and on via the internet. At the October 1989 INTEROP Conference, John Ramkey used a Sunbeam Radiant Control toaster connected to a TCP/IP network to demonstrate that the device could be turned off and on[11]. Not only did Ramkey succeed at turning the toaster on and off, he used SNMP code delivered via his computer's parallel port to a larger relay to control power to the toaster. The SNMP code executed commands for a value, 1 through 10, for the toast's doneness as well as a calculation for the type of item being toasted. For example, while the command for wheat bread would tell the toaster to toast at a level of 2, the command for a frozen bagel would tell the toaster to toast at a level of 5. Additional innovations were later added, such as a Lego robotic arm to insert the bread into the toaster; a sight Doc Brown would've been proud to see.

By 1999, the Salt Lake City Tribune/Deseret News was predicting that household appliances like the refrigerator were going to be part of a future in which "everyone lived like the Jetsons" [12]. "The networked home is on the horizon", the Tribune/Deseret News' Michael Stroh wrote, "with a click, you call up your refrigerator on your office PC to see what's inside (a bar-code reader within the fridge keeps a running inventory). The refrigerator suggests lasagna but warns that you'll need to buy ricotta - and a few other items" [12]. Not surprisingly, it would be at least another decade before this concept became a viable reality.

3 IOT IS BORN

The first time the term *Internet of Things* was used wasn't until nine years later by Kevin Ashton, co-founder of the Auto-ID Center at the Massachusetts Institute of Technology (MIT). The Auto-ID Center was founded with the expressed purpose of creating a formal standard for Radio Frequency Identification (RFID) and other types of networked sensors. In 2009, Kevin wrote[5]: "I could be wrong, but I'm fairly sure the phrase *Internet of Things* started life as the title of a presentation I made at Procter & Gamble (P&G) in 1999. Linking the new idea of RFID in P&G's supply chain to the then-red-hot topic of the Internet was more than just a good way to get executive attention. It summed up an important insight which is still often misunderstood."

Even though Kevin briefly captured the momentary attention of the C-Suite at P&G, it wasn't another full decade until the true concept of IoT caught on in the marketplace. In 2011, the market research company Gartner, included IoT on their hype cycle chart for the very first time. By 2016, IoT was past-peak of inflated expectations was doing the usual nosedive into the trough of disillusionment[7].

[Figure 1 about here.]

4 IOT SMART REFRIGERATORS COME OF AGE

Internet refrigerators, on the other hand were a bit slower catching up. After many failed attempts in the mid-2000s at various gimmicky models, it seemed that the once rosy future painted by Mr. Stroh a decade earlier was simply not going to come to fruition. Hardware and network technology had not yet caught up. By 2014, murmurs of a new wave of internet fridges hit the marketplace and excitement began to build, and by 2016, the IoT Refrigerator was ready for primetime. On January 24, 2016, Samsung launched its Smart Hub Refrigerator complete with a massive 21.5 inch 1080P touchscreen and Android operating system. Another exciting new feature of the Smart Hub fridge was the interior cameras that allowed users to get a real-time look at the contents of their fridge from anywhere[9].

A year later, Samsung debuted version 2.0 of the Smart Hub fridge, this time with improvements such as third-party apps such as Spotify and individualized user profiles for family members. Users are also able to serve photos and other content to the screen as well. Interestingly, Samsung has opted to go with the its own proprietary voice control system called S-Voice, while its only current competitor in the IoT fridge marketplace, LG, will integrate with Amazon's Alexa. Only in Europe, with the LidL supermarket chain, will consumers be able to order groceries through the the fridge. It's a start, but there is much, much more on the horizon[8].

5 THE FUTURE OF IOT SMART REFRIGERATORS

The future of IoT Smart Refrigerators - and kitchen appliances working in concert in general - is brighter than perhaps Doc Brown or even John Ramkey could have ever imagined. Hardware, networking, and most importantly, software, have all caught up to be viable in fulfilling consumer demands and there are fresh new ideas already just beginning to hit the marketplace. The next phase of the IoT Smart Refrigerator will be one that is marked by progress in software. Structurally speaking, refrigerators are designed to last between 14-17 years[2], however, the average consumer might upgrade their personal computer 3 to 4 times during this time span. In other words, an IoT Smart Refrigerator made today, might only be 1/4 to 1/3 of the way through its average lifespan before its computer and networking components become obsolete.

One Silicon Valley company that seems to have a viable solution to this problem is Innit[4]. Innit has come up with the idea of having a cloud-based platform for the kitchen that partners with appliance manufacturers such as Jenn Air and Whirlpool to add their components and integrate their application with existing appliance platforms. The idea is that you can equip your entire kitchen, not just the refrigerator, with technology that can make anyone a culinary master with a bit of guidance[10]. Building upon Samsung's successful Smart Hub fridge platform, Innit takes the camera-in-your-fridge concept a step further by introducing image recognition software that can be used to interface with the cloud to generate recipes based on available ingredients, manage spoilage, and inventory - including placing orders for new food. The technology would also enable other kitchen appliances such as an oven or microwave to interact with one another to create a meal.

Aside from personal convenience, one of the most significant values derived from the advance of this sort of technology is that it could prevent an enormous amount of food waste. The United Nations' Food and Agriculture Organization estimates that up to 1.3 billion tons of food are wasted globally every year[3], which equates to roughly 30 percent of all food produced in the same time-frame. Ultimately, software like Innit's because it is connected to the cloud and utilizing big data to allow consumers to make informed decisions about what they eat, people will live and eat healthier and greener.

6 SMART AND DANGEROUS: AN IOT DOUBLE-EDGED SWORD

Yes - it is true - both today and in the future, your IoT Smart Refrigerator will help you live better, but as Swapnil Bhartiya points out in a recent article on InfoWorld[6], it could also kill you. It sounds ominous, but the rapid growth of IoT comes with a steep price: lack of security. Consumers can never really be sure if their software will be patched properly and for how long. It has been well-documented that hackers have been able to successfully commandeer smart devices and utilize them to aggressively launch DDoS that disabled a sizeable portion of the internet. An even bigger threat is that, once compromised, a vulnerable smart device will work as a Trojan Horse allowing nefarious users to access other devices on your local network. Once you throw Alexa into the mix, all bets are off.

One development that is offsetting this risk is the unification of IoT networks in the cloud. Samsung is now creating a SmartThings cloud in which all of its IoT devices will interact. This centraliztion makes security and big data much easier to manage. This unification is also occurring at the macro level with Cisco and Google's cloud[1] which will hopes to achieve the following goals:

(1) Freedom to access any resource while preserving security and compliance
(2) Ability to extend policy to cloud environments to optimize applications
(3) Extend visibility, threat detection and control across hybrid environments without slowing innovation

7 CONCLUSION

IoT has a very bright future ahead and the rapidly evolving IoT Smart Refrigerator will serve as the centerpiece not only to a smart, connected kitchen, but to a smart, connected, and secure home. While it was hardware and networking that delayed progress in the 1990s and software and implementation that led to stagnation in the 2000s, security serves as the next challenge to be overcome as IoT Smart Refrigerators join the burgeoning global network of IoT smart devices.

REFERENCES

[1] 2017. Cisco and Google Cloud. (2017). Retrieved October 30th, 2017 from https://www.cisco.com/c/en/us/solutions/strategic-partners/google-cloud.html

[2] 2017. The Expected Life of a Refrigerator. (2017). Retrieved October 30th, 2017 from http://homeguides.sfgate.com/expected-life-refrigerator-88577.html

[3] 2017. Food and Agriculture Organization of the United Nations: Food Loss and Food Waste. (2017). Retrieved October 30th, 2017 from http://www.fao.org/food-loss-and-food-waste/en/

[4] 2017. Innit. (2017). Retrieved October 30th, 2017 from http://www.innit.com

[5] Kevin Ashton. 2009. That 'Internet of Things' Thing. *RFID Journal* (jun 2009), 1. http://www.rfidjournal.com/articles/view?4986

[6] Swapnil Bhartiya. 2017. Your smart fridge may kill you: The dark side of IoT. (2017). Retrieved October 30th, 2017 from https://www.infoworld.com/article/3176673/internet-of-things/your-smart-fridge-may-kill-you-the-dark-side-of-iot.html

[7] Inc. Gartner. 2017. Technologies Underpin the Hype Cycle for the Internet of Things, 2016. (2017). Retrieved October 30th, 2017 from https://www.gartner.com/smarterwithgartner/7-technologies-underpin-the-hype-cycle-for-the-internet-of-things-2016/

[8] Rik Henderson. 2017. Samsung Family Hub 2.0 refrigerator preview: Spotify and sausages. (2017). Retrieved October 30th, 2017 from http://www.pocket-lint.com/review/139892-samsung-family-hub-2-0-refrigerator-preview-spotify-and-sausages

[9] Stuart Miles. 2016. Samsung Family Hub Refrigerator comes with giant 21.5-inch screen and camera to spy on your food. (2016). Retrieved October 30th, 2017 from http://www.pocket-lint.com/news/136305-samsung-family-hub-refrigerator-comes-with-giant-21-5-inch-screen-and-camera-to-spy-on-your-food

[10] Rohini Nambiar. 2016. Smart kitchens are a new phase in the Internet of Things, as Innit explains. (2016). Retrieved October 30th, 2017 from https://www.cnbc.com/2016/07/26/smart-kitchens-are-a-new-phase-in-the-internet-of-things-as-innit-explains.html

[11] John Ramkey. 2016. Toast of the IoT: The 1990 Interop Internet Toaster. *IEEE* 6, Article 1 (dec 2016), 3 pages. https://doi.org/10.1109/MCE.2016.2614740

[12] Michael Stroh. 1999. Network systems allow us to live more like the Jetsons. (1999). Retrieved October 30th, 2017 from https://news.google.com/newspapers?nid=336&dat=19990116&id=lu8jAAAAIBAJ&sjid=iewDAAAAIBAJ&pg=3607,488766&hl=en

List of Figures

1 2016 Gartner Hype-Cycle Chart. 5

image missing

Figure 1: 2016 Gartner Hype-Cycle Chart.

bibtext report
===

```
This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--no key, author in SFGate2017
Warning--no author, editor, organization, or key in SFGate2017
Warning--to sort, need author or key in SFGate2017
Warning--no key, author in Innit2017
Warning--no author, editor, organization, or key in Innit2017
Warning--to sort, need author or key in Innit2017
Warning--no key, author in FAO2017
Warning--no author, editor, organization, or key in FAO2017
Warning--to sort, need author or key in FAO2017
Warning--no key, author in Cisco2017
Warning--no author, editor, organization, or key in Cisco2017
Warning--to sort, need author or key in Cisco2017
Warning--no key, author in Cisco2017
Warning--no key, author in Cisco2017
Warning--no key, author in FAO2017
Warning--no key, author in FAO2017
Warning--no key, author in Innit2017
Warning--no key, author in Innit2017
Warning--no key, author in SFGate2017
Warning--no key, author in SFGate2017
Warning--no key, author in Cisco2017
Warning--no author, editor, organization, or key in Cisco2017
Warning--empty author in Cisco2017
Warning--no key, author in SFGate2017
Warning--no author, editor, organization, or key in SFGate2017
Warning--empty author in SFGate2017
Warning--no key, author in FAO2017
Warning--no author, editor, organization, or key in FAO2017
Warning--empty author in FAO2017
Warning--no key, author in Innit2017
Warning--no author, editor, organization, or key in Innit2017
Warning--empty author in Innit2017
Warning--no number and no volume in Ashton01
Warning--numpages field, but no articleno or eid field, in Ashton01
(There were 34 warnings)
```

bibtext _ label error
===

```
bibtext space label error
===========================================


bibtext comma label error
===========================================


latex report
===========================================

        [2017-11-04 09.46.10] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
Typesetting of "report.tex" completed in 0.9s.

================================================================================
Compliance Report
================================================================================


name: Robert Gasiewicz
hid:  316
paper1: 100% Oct 25 17
paper2: 99%
project: 10%

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

5
wc 316 paper2 5 1898 report.tex
wc 316 paper2 5 2046 report.pdf
wc 316 paper2 5 408 report.bib

find "
--------------------------------------------------------------------------------

passed: True
```

```
find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

52: \begin{figure}
54: %\includegraphics[width=\linewidth]{gartner2016.png}
56: \label{fig:Gartner2016}

figures 1
tables 0
includegraphics 1
labels 1
refs 0
floats 1

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
False : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
```

----

label errors

bibtex errors

```
This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--no key, author in SFGate2017
Warning--no author, editor, organization, or key in SFGate2017
Warning--to sort, need author or key in SFGate2017
Warning--no key, author in Innit2017
Warning--no author, editor, organization, or key in Innit2017
Warning--to sort, need author or key in Innit2017
Warning--no key, author in FAO2017
Warning--no author, editor, organization, or key in FAO2017
Warning--to sort, need author or key in FAO2017
Warning--no key, author in Cisco2017
Warning--no author, editor, organization, or key in Cisco2017
Warning--to sort, need author or key in Cisco2017
Warning--no key, author in Cisco2017
Warning--no key, author in Cisco2017
Warning--no key, author in FAO2017
Warning--no key, author in FAO2017
Warning--no key, author in Innit2017
Warning--no key, author in Innit2017
Warning--no key, author in SFGate2017
Warning--no key, author in SFGate2017
Warning--no key, author in Cisco2017
Warning--no author, editor, organization, or key in Cisco2017
Warning--empty author in Cisco2017
Warning--no key, author in SFGate2017
Warning--no author, editor, organization, or key in SFGate2017
Warning--empty author in SFGate2017
Warning--no key, author in FAO2017
Warning--no author, editor, organization, or key in FAO2017
Warning--empty author in FAO2017
Warning--no key, author in Innit2017
Warning--no author, editor, organization, or key in Innit2017
Warning--empty author in Innit2017
Warning--no number and no volume in Ashton01
Warning--numpages field, but no articleno or eid field, in Ashton01
```

```
(There were 34 warnings)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

9: number  =       "",

14: acmid          =       "",

15: note           =       "",

22: articleno      =       "",

24: volume  =       "",

25: number  =       "",

28: doi            =       "",

30: acmid          =       "",

31: note           =       "",

39: month =          "",

48: month =          "",

57: month = "",

66: month =          "",

71: author =      "",

75: month =          "",

80: author =      "",

84: month =          "",
```

```
93: month =          "",

98: author =       "",

102: month =          "",

111: month =          "",

116: author =       "",

120: month =          "",

passed: False

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Big Data Applications in Virtual Assistants

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

ABSTRACT

This paper provides

KEYWORDS

i523, HID233, Big data, Virtual Assistants, Artificial intelligence

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" [11] [13] [3] [5] [8] [12] [7] [6] [10] [2] [9] [4] [1].

2 FIGURES

3 LONG EXAMPLE

4 CONCLUSION

Put here an conclusion. Conclusions and abstracts must not have any citations in the section.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Ethan Baron. 2017. One bot to rule them all? Not likely, with Apple, Google, Amazon and Microsoft virtual assistants. Web Page. (Feb. 2017). http://www.mercurynews.com/2017/02/06/one-bot-to-rule-them-all-not-likely-with-apple-google-amazon-and-microsoft-virtual-assistants/ HID: 233, Accessed: 2017-10-24.

[2] Clint Boulton. 2016. Slack CEO describes 'Holy Grail' of virtual assistants. Web Page. (Oct. 2016). https://www.cio.com/article/3131536/collaboration/slack-ceo-describes-holy-grail-of-virtual-assistants.html HID: 233, Accessed: 2017-10-24.

[3] Mike Elgan. 2016. These three virtual assistants point the way to the future. Web Page. (June 2016). https://www.computerworld.com/article/3078829/artificial-intelligence/these-three-virtual-assistants-point-the-way-to-the-future.html HID: 233, Accessed: 2017-10-18.

[4] Darrell Etherington. 2014. The Virtual Assistant Could Be The Next Interpreter Of Enterprise Data, Starting With Google Now. Web Page. (Aug. 2014). https://techcrunch.com/2014/08/13/the-virtual-assistant-could-be-the-next-interpreter-of-enterprise-data-starting-with-google-now/ HID: 233, Accessed: 2017-10-24.

[5] Lars Hard. 2014. The Disruptive Potential of Artificial Intelligence Applications. Web Page. (Jan. 2014). http://data-informed.com/disruptive-potential-artificial-intelligence-applications/ HID: 233, Accessed: 2017-10-18.

[6] Eran Kinsbruner. 2017. Building the virtual assistant everyone wants. Web Page. (July 2017). https://www.infoworld.com/article/3210488/machine-learning/building-the-virtual-assistant-everyone-wants.html HID: 233, Accessed: 2017-10-24.

[7] Rob Marvin. 2017. What Are Virtual Assistants and What Can You Do With Them? Web Page. (June 2017). https://www.pcmag.com/article/354371/what-are-virtual-assistants-and-what-can-you-do-with-them HID: 233, Accessed: 2017-10-24.

[8] Susanne Mueller. 2016. Rhiza Launches Rhizabot, First Virtual Assistant for Analytics. Web Page. (Aug. 2016). http://rhiza.com/2016/08/10/rhiza-launches-rhizabot-first-virtual-assistant-analytics/ HID: 233, Accessed: 2017-10-18.

[9] Tom Simonite. 2016. How Alexa, Siri, and Google Assistant Will Make Money Off You. Web Page. (May 2016). https://www.technologyreview.com/s/601583/how-alexa-siri-and-google-assistant-will-make-money-off-you/ HID: 233, Accessed: 2017-10-24.

[10] Anubhav Srivastava. 2016. Why the virtual assistants market is on the upswing? Web Page. (July 2016). http://thinkbigdata.in/virtual-assistants-market-upswing/ HID: 233, Accessed: 2017-10-24.

[11] David Tal. 2015. Forecast — Rise of the big data-powered virtual assistants: Future of the Internet P3. Web page. (Nov. 2015). http://www.quantumrun.com/prediction/rise-big-data-powered-virtual-assistants-future-internet-p3 HID: 233, Accessed: 2017-10-18.

[12] Spotfire Blogging Team. 2012. Meet Your Companyfis New Virtual Assistant fi?! Big Data. Web Page. (May 2012). https://www.tibco.com/blog/2012/05/11/meet-your-companys-new-virtual-assistant-big-data/ HID: 233, Accessed: 2017-10-18.

[13] Richard Waters. 2015. Artificial intelligence: A virtual assistant for life. Web page. (Feb. 2015). https://www.ft.com/content/4f2f97ea-b8ec-11e4-b8e6-00144feab7de?mhq5j=e5 HID: 233, Accessed: 2017-10-18.

We include an appendix with common issues that we see when students submit papers.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To do this change TODO with DONE.

```
bibtext report
===========================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib

bibtext _ label error
===========================================


bibtext space label error
===========================================


bibtext comma label error
===========================================


latex report
===========================================

        [2017-11-04 09.45.29] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
Missing character: ""
Missing character: ""
Missing character: ""
Typesetting of "report.tex" completed in 0.8s.

================================================================================
Compliance Report
================================================================================


name: Wang, Jiaan
hid:  233
paper1: Nov 03 17 100%
paper2: 50%
project: 10%

yamlcheck
--------------------------------------------------------------------------------
```

```
wordcount
--------------------------------------------------------------------------------

1
wc 233 paper2 1 175 content.tex
wc 233 paper2 1 456 report.pdf
wc 233 paper2 1 483 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------



figures 0
tables 0
includegraphics 0
labels 0
refs 0
floats 0

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
```

```
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

passed: True

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------
```

```
passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Why Deep Learning matters in IoT Data Analytics?

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

ABSTRACT

The Deep Learning is unique in all machine learning algorithms to analyze supervised and unsupervised datasets. Big Data challenges, such as high volumes, multi-dimensionality and feature engineering, are well addressed using Deep Learning algorithms. Deep Leaning, with Edge and distributed Mesh computing, is best suited to handle IoT Analytics from millions of sensors producing petabytes of time-series data.

KEYWORDS

i523, hid306, IoT, Deep Learning, Big Data Analytics

1 INTRODUCTION

Supervised machine learning algorithms: decision trees, linear regression, Support Vector Machines (SVMs), Naive Bayes, neural networks, etc. are popular for classification and regression problems by analyzing labeled training data. K-means clustering algorithms are good for unsupervised datasets to categorize based on the identified patterns in unlabeled data. While there are so many factors - nature of the domain, sample size of the dataset and number of attributes defining characteristics of the data - decide which machine learning algorithm works better, Deep Learning algorithms are, getting greater traction, addressing complex analytics tasks, including high-dimensionality and automatic creation of new features from existing complex hierarchical features, very well.

2 NEURAL NETWORKS

Neural Networks are modeled after human brain, the way they solve complex problems. *Perceptron*, the first generation neural network, created a simple mathematical model, mimicking neuron - the basic unit of the brain, by taking several binary inputs and produced single binary output. *Sigmoid Neuron* improved learning by giving some *weightage* to the input based on importance of the corresponding input to the output so that tiny changes in the output due to the minor adjustments in the input weights (or biases) can be measured effectively. Neural Network is, a *directed graph*, organized by layers and layers are created by number of interconnected neurons (or nodes). Every neuron in a layer is connected with all the neurons from the previous layer; there will no interaction of neurons within a layer. As shown in Figure (1), a typical Neural Network contains three layers: input (left), hidden (middle) and output (right) [3]. The middle layer is called *hidden* only because the neurons of this layer are neither an input nor an output but the actual processing happen in the hidden layer. As data passes through layer by layer, each neuron acts as an *activation function* to process the input. The performance of a Neural Network is measured using *cost or error function* and the dependent input *weight* variables. *Forward-propagation* and *back-propagation* are two techniques, neural network uses repeatedly until all the input variables are adjusted or calibrated to predict accurate output. During, forward-propagation, information moves in forward direction and passes through all the layers by applying certain weights to the input parameters. *Back-propagation* method minimizes the error in the *weights* by applying an algorithm called *gradient descent* at each iteration step.

[Figure 1 about here.]

3 DEEP LEARNING

Deep Learning is an advanced neural network, with multiple hidden layers (thousands or even more deep), that can work well with supervised (labeled) and unsupervised (unlabeled) datasets. Applications, such as speech, image and behavior patterns, having complex relationships in large-set of attributes, are best suited for Deep Learning Neural Networks. Deep Learning vectorizes the input and converts it into output vector space by decomposing complex geometric and polynomial equations into a series of simple transformations. These transformations go through neuron activation functions at each layer parameterized by input weights. For it to be effective, the cost function of the neural network must guarantee two mathematical properties: *continuity* and *differentiability*.

[Figure 2 about here.]

3.1 Feature Engineering

The dataset with too many dimensions, also known as attributes or features, create large sparsity and make it difficult to process. *Curse of dimensionality* is a scenario where the value added by the dimensions is much smaller in comparison to the processing cost. However, in certain applications, such as face recognition and patient electronic medical records, the complexity created by multiple dimensions might add value to the context. *Feature Engineering* is an exploratory analysis to identify the features that collectively contribute to better predictive modeling by removing irrelevant features and creating new features, using the training information to identify the patterns, from existing interrelated features [6]. *Principal Component Analysis* (PCA) is a technique to analyze the interdependency among the features and keep only the principal, most relevant, features with minimum loss in the model. With enough training, Deep Learning makes neurons learn new features themselves, in an unsupervised manner, from existing features distributed in several hidden layers. *Stacked Autoencoder* (AE) is, a Deep Belief Network algorithm, to create advanced predictive models for large datasets having thousands or even millions of dimensions, automatically, with complex hierarchical attributes in non-linear fashion for simpler computing. Though AE is sophisticated, it is very difficult to understand the algorithm logic and so unable to reuse the learnings from the modeling to other systems.

3.2 Deep Neural Networks

Convolutional Neural Network (CNN), also called multilayer perceptron (MLP), is a deep feedforward network, consists of (1) convolutional layers - to identify the features using weights and biases, followed by (2) fully connected layers - where each neuron is connected from all the neurons of previous layers - to provide non-linearity, sub-sampling or max-pooling, performance and control data overfitting [2]. CNN is used in image and voice recognition applications by effectively using multiples copies of same neuron and reusing group of neurons in several places to make them *modular*. CNNs are constrained by *fixed-size* vectorized inputs and outputs. *Recursive Neural Network* (RNN) is, another type of Deep Learning, that uses same shared feature weights recursively for processing sequential data, emitted by sensors or the way spoken words are processed in natural language processing (NLP), to produce arbitrary size input and output vectors. RNN uses a technique called *loop*, where several copies of the same chunk of network (module), each instance passing a message to the next, to persist the information. Long Short Term Memory (LSTM) is an advanced RNN to learn and remember *longer* sequences by composing series of repeated modules of neural network and a concept called *cell state*, a memory unit, to memorize the learning by adding and removing information using *input*, *output* and *forget* gates, in a regularized fashion while data flows through the layers [9]. The Convolutional and Recursive Neural Networks can complement each other to produce better and effective models where problem space has both - hierarchical features and temporal data. Deep Learning can also work well with related *Reinforcement Learning* algorithms where the focus is on how to maximize the learning based on rewards and punishments.

[Figure 3 about here.]

[Figure 4 about here.]

4 IOT DATA ANALYTICS

Internet of Things (IoT) is getting lots of traction, due to the massive volumes of data, making it *Big Data*; however, business needs to convert this data into *information* whether to monitor and control the devices or to analyze the sensor data for betterment. Time series data has non-stationary time aspects collected at certain intervals over a short period of time and correlate this sequence of data with past or future sequences. Stock prices and IoT sensor data are examples of time series data. *InfluxDB*, an open source time series database, is offering high write performance, data compaction through down-sampling and automatic deletion of expired old time series data, to address IoT data storage challenges [5].

4.1 Complexity

Unique traits of IoT data, such as noise, high dimensionality and high streaming of time-series data in real-time, make it challenging to process using traditional machine learning algorithms [10]. Autoregressive Moving Average Model (ARIMA), converts time-series from non-stationary into stationary, but only for short-time predictions. Deep Learning, using LSTM, can detect anomalies in the IoT Data and train time series data very well. Deep Learning algorithms involve complex mathematics - geometry, matrix algebra, differential calculus, statistics and probability, and intensive distributed computing to train the massive amounts of sensor data.

4.2 Scalability

Deep Learning, by design, allows parallel programming, as each module - with all the dependencies among neurons - can run independently and parallelly from other modules within the network. Using Graphics Process Unit (GPU), module networks can achieve parallel programming without needing much of Central Processing Unit (CPU) allocation. Though GPU is intended for graphical processing, it works efficiently to run thousands of small mathematical functions, such as matrix multiplications, in parallel. Cloud computing and Edge Analytics offer flexible scale out options, using virtualization and containerization, for distributed processing. Sophisticated algorithms and distributed computing make Deep Learning scale and perform well to process huge datasets.

4.3 Case Study

Hewlett Packard (HP) Labs has given a presentation of their experiments to check how effectively Deep Learning algorithms can be applied for IoT Sensor Data Analytics. Sample data - vision, speech, text and sensor data such as signals, have been collected from scripted video and accelerometer from 52 subjects on average 20 minutes of activity recognition per subject - 12,000 measurements per minute per person with 16 classifications, such as walk to bed, enter bed, lie down, roll left, roll right and speak. They have analyzed and trained the sample data using various supervised learning algorithms including SVMs, Decision Trees and traditional Neural Networks; compared the results with Recurrent, Deep Learning, Neural Network. Deep Learning showed 95% or more accuracy in various scenarios, performed much better than all the other algorithms, without sophisticated feature engineering. However, Deep Learning algorithms were slow and expensive for results to converge as the sample dataset is huge with lots of instances (10^6-10^9) and very large number of features ($>10^6$). They have concluded the presentation with scale-out hardware options using CPU/GPU clusters and futuristic Edge Analytics and distributed Mesh Computing alternatives for better scalability and performance [11].

5 CONCLUSION

In contrast to traditional machine learning solutions, Deep Learning not only scales well with high volumes of input data but also facilitates in automatic decomposition of complex data representations of unsupervised and uncategorized data. Automatic discovery of new features, from convolutional or recurrent neural networks, makes Deep Learning predominant among all machine learning algorithms. It is very difficult to understand fuzzy and complex logic of Deep Learning, perhaps, more adoption helps getting better handle at them. Deep Learning algorithms need deep research in validating the process of advanced Big Data Analytics tasks, such as IoT sensor time-series data, semantic learning, scalability, data tagging and reliability of the predictive models without extreme generalization.

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski and the Teaching Assistants for their support and valuable suggestions.

REFERENCES

[1] Mark Chang. 2016. Applied Deep Learning 11/03 Convolutional Neural Networks. (Oct. 2016). https://www.slideshare.net/ckmarkohchang/applied-deep-learning-1103-convolutional-neural-networks
[2] Christopher Olah. 2014. Conv Nets: A Modular Perspective. (July 2014). http://colah.github.io/posts/2014-07-Conv-Nets-Modular/
[3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org
[4] Vikas Gupta. 2017. Understanding Feedforward Neural Networks. (Oct. 2017). https://www.learnopencv.com/understanding-feedforward-neural-networks/
[5] Influx. [n. d.]. *InfluxDB is the Time Series Database in the TICK stack*. Technical Report. Influx. https://www.influxdata.com/time-series-platform/influxdb/
[6] Jason Brownlee. 2014. Discover Feature Engineering, How to Engineer Features and How to Get Good at It. (Sept. 2014). https://machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/
[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. (May 2015). http://www.nature.com/nature/journal/v521/n7553/fig_tab/nature14539_F5.html
[8] Nicholas Leonard. 2016. Language modeling a billion words. (July 2016). http://torch.ch/blog/2016/07/25/nce.html
[9] Christopher Olah. 2015. Understanding LSTM Networks. (Aug. 2015). http://colah.github.io/posts/2015-08-Understanding-LSTMs/
[10] Rajesh Sampathkumar. 2016. Time Series Analysis of Sensor Data. (Aug. 2016). http://www.thedatateam.in/time-series-analysis-of-sensor-data/
[11] Natalia Vassilieva. 2016. *Sense Making in an IOT World: Sensor Data Analysis with Deep Learning*. Technical Report. Hewlett Packard Labs. http://on-demand.gputechconf.com/gtc/2016/presentation/s6773-natalia-vassilieva-sensor-data-analysis.pdf

List of Figures

1 Simple Neural Network [3, 4]. 5
2 Deep Neural Network with three hidden layers [3]. 5
3 Sample Convolutional Neural Network [1]. 5
4 Recursive Neural Network Loop and LSTM Cell State [7, 8]. 6

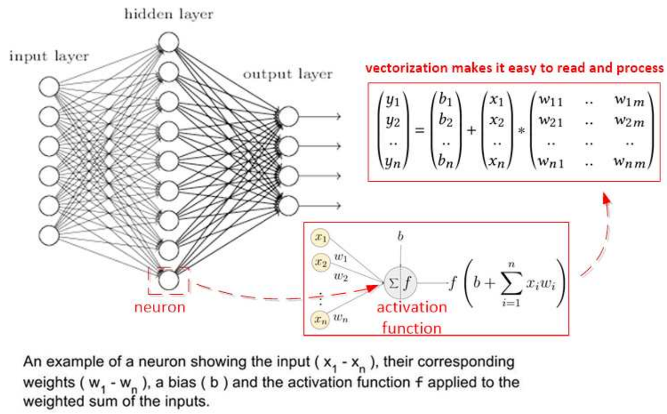


Figure 1: Simple Neural Network [3, 4].

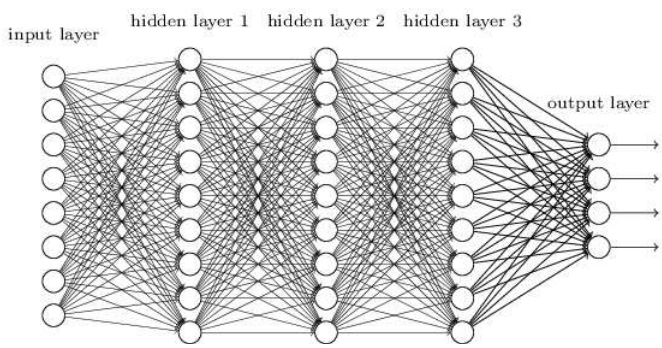


Figure 2: Deep Neural Network with three hidden layers [3].

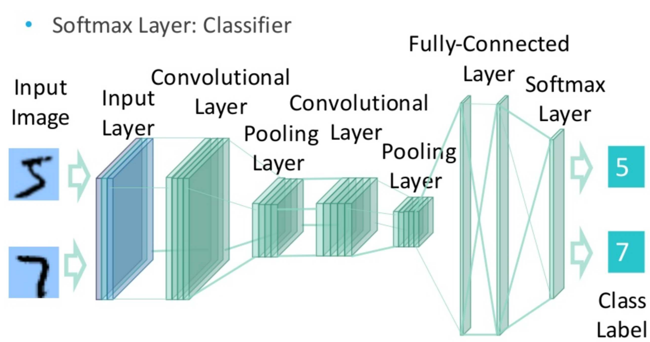


Figure 3: Sample Convolutional Neural Network [1].

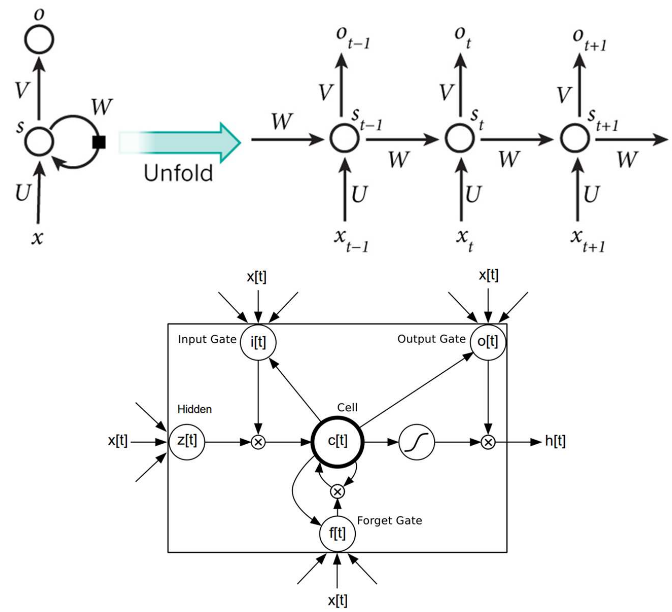


Figure 4: Recursive Neural Network Loop and LSTM Cell State [7, 8].

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--empty address in Goodfellow2016
Warning--empty year in Influx
(There were 2 warnings)

bibtext _ label error
============================================


bibtext space label error
============================================


bibtext comma label error
============================================


latex report
============================================

        [2017-11-04 09.45.59] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
Typesetting of "report.tex" completed in 0.9s.
../README.yml
  8:63      error    trailing spaces  (trailing-spaces)
  9:63      error    trailing spaces  (trailing-spaces)
  11:61     error    trailing spaces  (trailing-spaces)
  12:60     error    trailing spaces  (trailing-spaces)
  25:67     error    trailing spaces  (trailing-spaces)
  26:66     error    trailing spaces  (trailing-spaces)
  27:63     error    trailing spaces  (trailing-spaces)
  28:53     error    trailing spaces  (trailing-spaces)
  29:62     error    trailing spaces  (trailing-spaces)
  30:61     error    trailing spaces  (trailing-spaces)
  31:55     error    trailing spaces  (trailing-spaces)
  41:10     error    too many spaces after colon  (colons)
```

```
================================================================================
Compliance Report
================================================================================


name: Cheruvu, Murali
hid:  306
paper1: 100%; 10/26/2017
paper2: 85%

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

6
wc 306 paper2 6 1841 report.tex
wc 306 paper2 6 1926 report.pdf
wc 306 paper2 6 273 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

45: \begin{figure}
47: \includegraphics[width=0.5\textwidth]{images/neuralnetwork}
48: \caption{Simple Neural Network \cite{Goodfellow2016, Gupta2017}.}
```

```
    \label{fig:figure1}
56: \begin{figure}
58: \includegraphics[width=0.5\textwidth]{images/deepnetwork}
59: \caption{Deep Neural Network with three hidden layers
    \cite{Goodfellow2016}.} \label{fig:figure2}
71: \begin{figure}
73: \includegraphics[width=0.5\textwidth]{images/cnn}
74: \caption{Sample Convolutional Neural Network \cite{Chang2016}.}
    \label{fig:figure3}
77: \begin{figure}
79: \includegraphics[width=0.5\textwidth]{images/rnn}
80: \caption{Recursive Neural Network Loop and LSTM Cell State
    \cite{LeCun2015, Leonard2016}.} \label{fig:figure4}

figures 4
tables 0
includegraphics 4
labels 4
refs 0
floats 4

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
False : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

47: \includegraphics[width=0.5\textwidth]{images/neuralnetwork}

58: \includegraphics[width=0.5\textwidth]{images/deepnetwork}

73: \includegraphics[width=0.5\textwidth]{images/cnn}

79: \includegraphics[width=0.5\textwidth]{images/rnn}

passed: False
```

```
bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--empty address in Goodfellow2016
Warning--empty year in Influx
(There were 2 warnings)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

passed: True

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {
```

```
find cite {
--------------------------------------------------------------------------------

passed: True
```

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 [**?**].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 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

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction}` and NOT `\section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use "quotes" , instead of " "

To emphasize a word, use *emphasize* and not "quote"

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬃgure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inﬂate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named "images"

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

List of Figures

1 Example caption 4

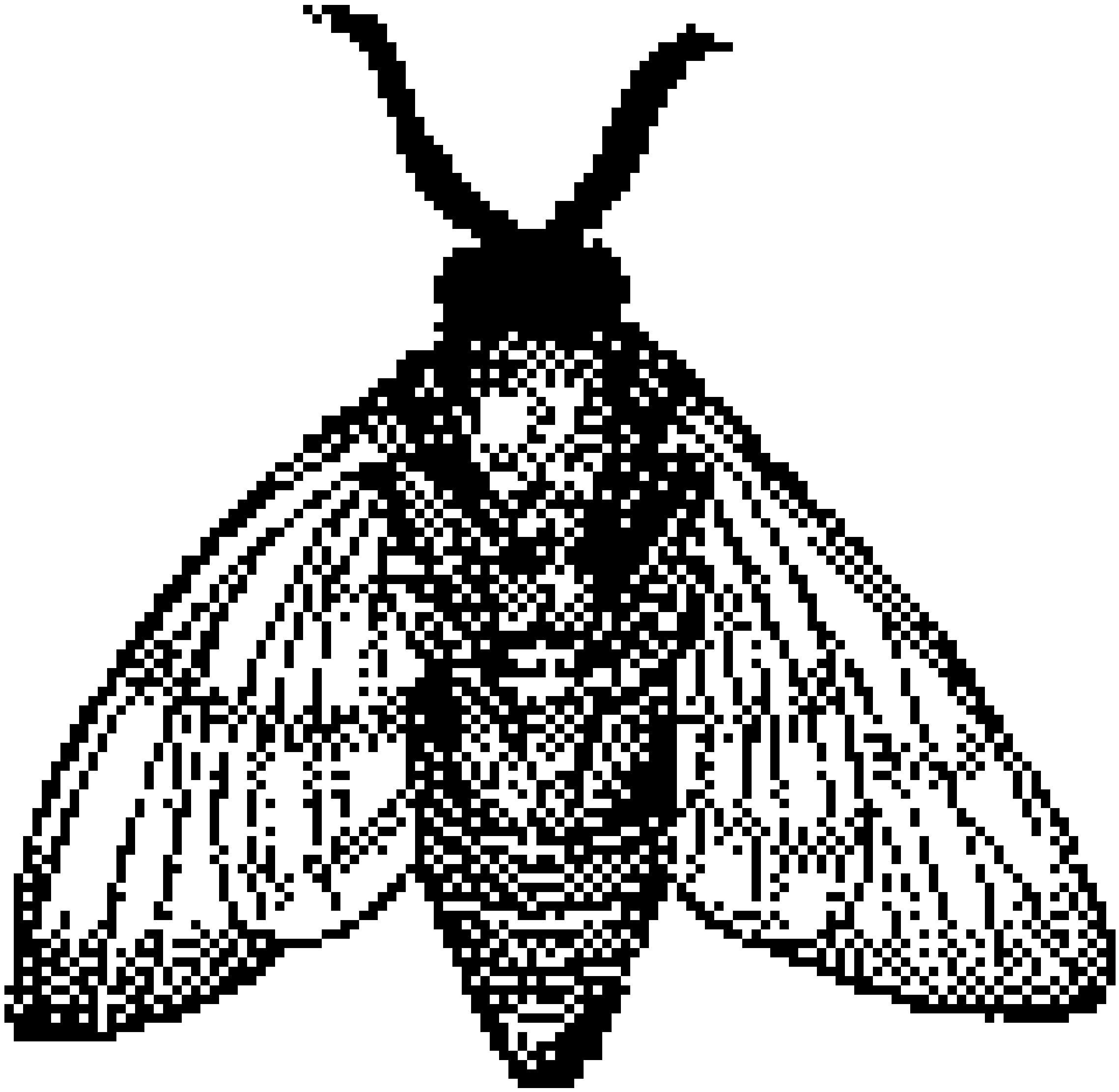


Figure 1: Example caption

4

List of Tables

1 My caption 6

Table 1: My caption

1	2	3
4	5	6
7	8	9

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtext _ label error
============================================


bibtext space label error
============================================


bibtext comma label error
============================================


latex report
============================================

        [2017-11-04 09.44.42] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
p.1   L62   : [editor00] undefined
p.1   L89   : 't:mytable' undefined
Empty `thebibliography' environment.
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
There were undefined citations.
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
There were undefined references.
Typesetting of "report.tex" completed in 1.2s.

================================================================================
Compliance Report
```

```
================================================================================


name: Lipe-Melton, Josh
hid:  105
paper1: 100% October 27, 2017

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

6
wc 105 paper2 6 518 report.tex
wc 105 paper2 6 1163 report.pdf
wc 105 paper2 6 50 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

67: In Figure \ref{f:fly} we show a fly. Please note that because we
    use
73: \begin{figure}[!ht]
74: \centering\includegraphics[width=\columnwidth]{images/fly.pdf}
75: \caption{Example caption}\label{f:fly}
90: or generate them by hand while using the provided template in
    Table\ref{t:mytable}. Not ethat
```

```
93: \begin{table}[htb]
96: \label{t:mytabble}

figures 1
tables 1
includegraphics 1
labels 2
refs 2
floats 2

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------



label errors



bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtex_empty_fields
--------------------------------------------------------------------------------
```

```
entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

15: note            = "",

passed: False

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

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 [?].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 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

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

```
\section{Introduction} and NOT \section{INTRODUCTION}
```

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use "quotes" , instead of " "

To emphasize a word, use *emphasize* and not "quote"

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬀigure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named "images"

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

```
/includegraphics[width=\columnwidth]{images/myimage.pdf}
```

re

2

List of Figures

1 Example caption 4

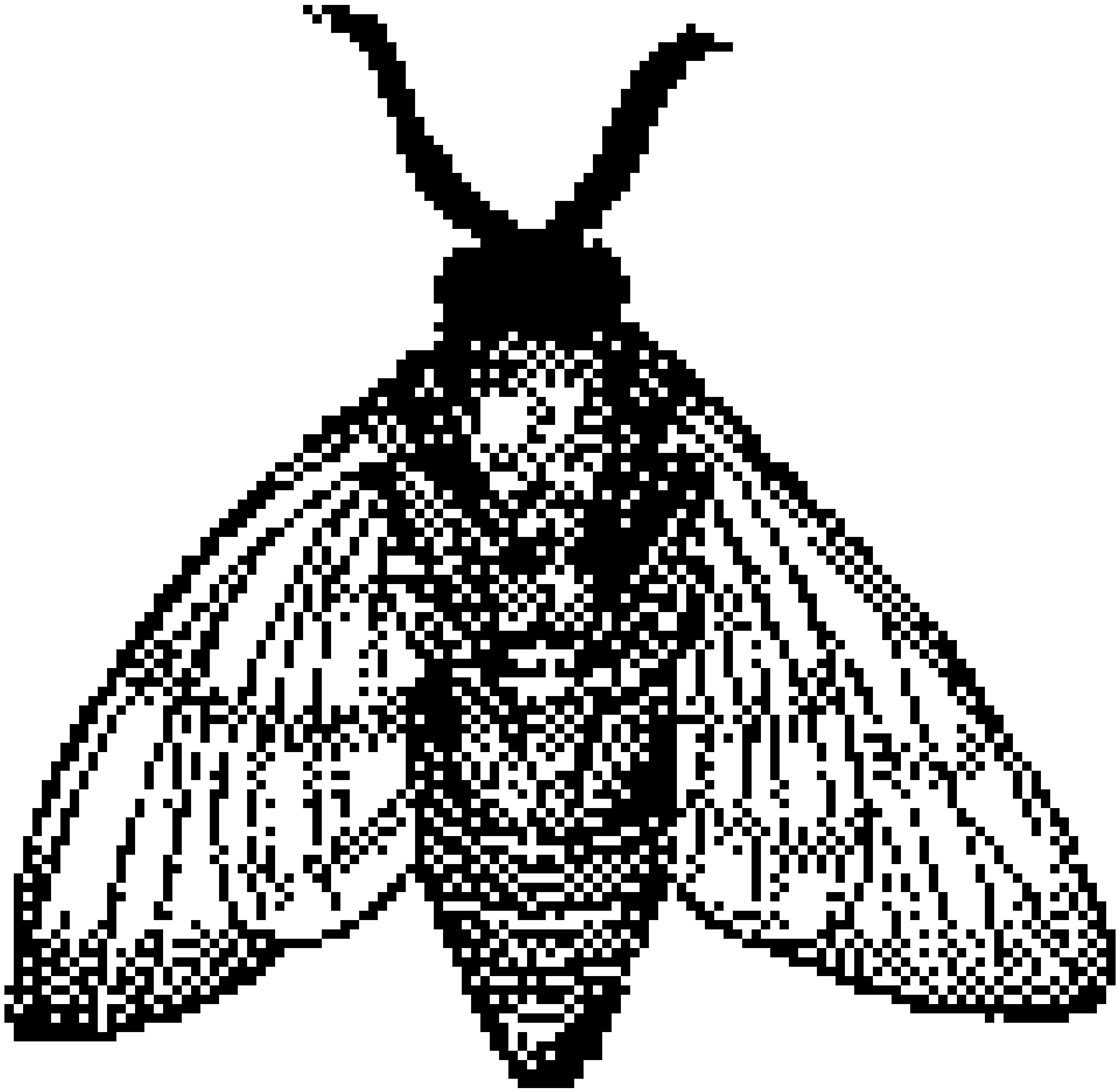


Figure 1: Example caption

List of Tables

1 My caption 6

Table 1: My caption

1	2	3
4	5	6
7	8	9

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtext _ label error
============================================


bibtext space label error
============================================


bibtext comma label error
============================================


latex report
============================================

        [2017-11-04 09.44.48] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
p.1   L62   : [editor00] undefined
p.1   L89   : 't:mytable' undefined
Empty `thebibliography' environment.
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
There were undefined citations.
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
There were undefined references.
Typesetting of "report.tex" completed in 1.2s.
../README.yml
  8:81      error    line too long (81 > 80 characters)  (line-length)
  8:81      error    trailing spaces  (trailing-spaces)
```

```
9:79      error    trailing spaces  (trailing-spaces)
10:81     error    line too long (83 > 80 characters)  (line-length)
11:81     error    line too long (83 > 80 characters)  (line-length)
11:83     error    trailing spaces  (trailing-spaces)
12:81     error    line too long (81 > 80 characters)  (line-length)
12:81     error    trailing spaces  (trailing-spaces)
13:81     error    line too long (82 > 80 characters)  (line-length)
13:82     error    trailing spaces  (trailing-spaces)
14:80     error    trailing spaces  (trailing-spaces)
15:81     error    line too long (83 > 80 characters)  (line-length)
15:83     error    trailing spaces  (trailing-spaces)
16:81     error    line too long (86 > 80 characters)  (line-length)
16:86     error    trailing spaces  (trailing-spaces)
17:81     error    line too long (84 > 80 characters)  (line-length)
17:84     error    trailing spaces  (trailing-spaces)
30:81     error    line too long (87 > 80 characters)  (line-length)
31:81     error    line too long (88 > 80 characters)  (line-length)
31:88     error    trailing spaces  (trailing-spaces)
32:81     error    line too long (88 > 80 characters)  (line-length)
32:88     error    trailing spaces  (trailing-spaces)
33:81     error    line too long (88 > 80 characters)  (line-length)
33:88     error    trailing spaces  (trailing-spaces)
34:81     error    line too long (87 > 80 characters)  (line-length)
34:87     error    trailing spaces  (trailing-spaces)
35:81     error    line too long (90 > 80 characters)  (line-length)
36:81     error    line too long (86 > 80 characters)  (line-length)
36:86     error    trailing spaces  (trailing-spaces)
37:81     error    line too long (88 > 80 characters)  (line-length)
37:88     error    trailing spaces  (trailing-spaces)
38:81     error    line too long (88 > 80 characters)  (line-length)
38:80     error    trailing spaces  (trailing-spaces)
49:81     error    line too long (89 > 80 characters)  (line-length)
49:89     error    trailing spaces  (trailing-spaces)
50:81     error    line too long (88 > 80 characters)  (line-length)
50:88     error    trailing spaces  (trailing-spaces)
51:81     error    line too long (87 > 80 characters)  (line-length)
51:87     error    trailing spaces  (trailing-spaces)
52:81     error    line too long (87 > 80 characters)  (line-length)
53:81     error    line too long (91 > 80 characters)  (line-length)
53:91     error    trailing spaces  (trailing-spaces)
54:81     error    line too long (88 > 80 characters)  (line-length)
55:81     error    line too long (88 > 80 characters)  (line-length)
55:88     error    trailing spaces  (trailing-spaces)
56:66     error    trailing spaces  (trailing-spaces)
63:24     error    trailing spaces  (trailing-spaces)
```

```
================================================================================
Compliance Report
================================================================================


name: Shiqi Shen
hid:  109
paper1: 100% Oct 27th
paper2: in progress 50%
project: in progress

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

6
wc 109 paper2 6 518 report.tex
wc 109 paper2 6 1163 report.pdf
wc 109 paper2 6 50 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

67: In Figure \ref{f:fly} we show a fly. Please note that because we
    use
```

```
73: \begin{figure}[!ht]
74: \centering\includegraphics[width=\columnwidth]{images/fly.pdf}
75: \caption{Example caption}\label{f:fly}
90: or generate them by hand while using the provided template in
    Table\ref{t:mytable}. Not ethat
93: \begin{table}[htb]
96: \label{t:mytabble}

figures 1
tables 1
includegraphics 1
labels 2
refs 2
floats 2

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
```

```
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

15: note            = "",

passed: False

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

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 [?].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 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

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬁgure and not a figure in text you copied from a text that has the fi combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named ”images”

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

List of Figures

1 Example caption 4

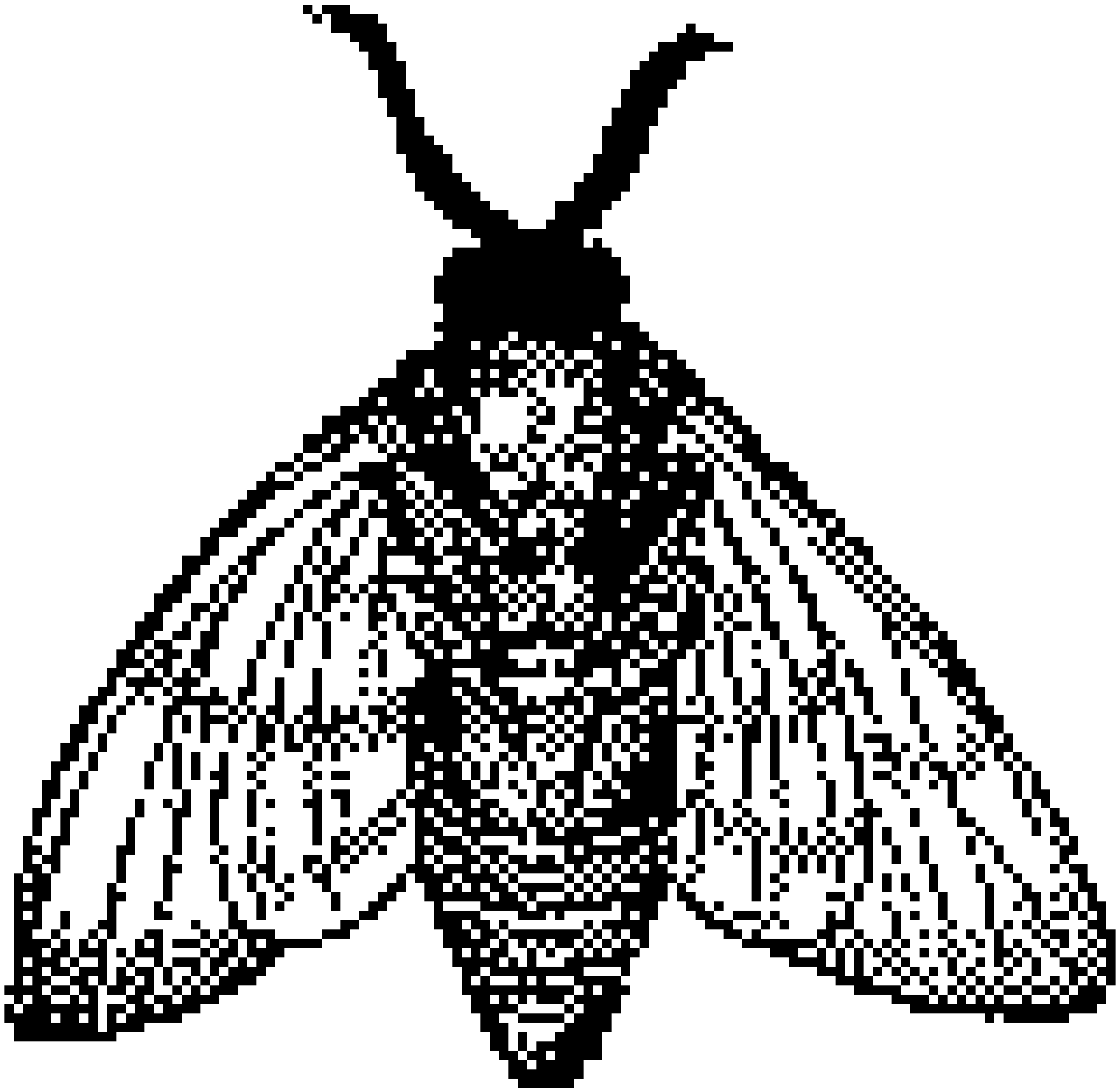


Figure 1: Example caption

List of Tables

1 My caption 6

Table 1: My caption

1	2	3
4	5	6
7	8	9

6

```
bibtex report
===========================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn’t find a database entry for "editor00"
(There was 1 warning)

bibtex _ label error
===========================================


bibtex space label error
===========================================


bibtex comma label error
===========================================


latex report
===========================================

        [2017-11-04 09.44.55] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
p.1   L62   : [editor00] undefined
p.1   L89   : ’t:mytable’ undefined
Empty ‘thebibliography’ environment.
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
There were undefined citations.
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent’s must not be the same. Added a new anchor.
The anchor of a bookmark and its parent’s must not be the same. Added a new anchor.
There were undefined references.
Typesetting of "report.tex" completed in 1.4s.
../README.yml
  9:81      error    line too long (83 > 80 characters)  (line-length)
  10:81     error    line too long (81 > 80 characters)  (line-length)
```

```
  11:81     error    line too long (82 > 80 characters)  (line-length)
  12:81     error    line too long (81 > 80 characters)  (line-length)
  13:81     error    line too long (81 > 80 characters)  (line-length)


================================================================================
Compliance Report
================================================================================


name: Arnav, Arnav
hid:  201
paper1: 20th Oct 2017 100%
paper2: not started
project: not started

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

6
wc 201 paper2 6 518 report.tex
wc 201 paper2 6 1163 report.pdf
wc 201 paper2 6 50 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
```

```
--------------------------------------------------------------------------------

67: In Figure \ref{f:fly} we show a fly. Please note that because we
    use
73: \begin{figure}[!ht]
74: \centering\includegraphics[width=\columnwidth]{images/fly.pdf}
75: \caption{Example caption}\label{f:fly}
90: or generate them by hand while using the provided template in
    Table\ref{t:mytable}. Not ethat
93: \begin{table}[htb]
96: \label{t:mytabble}

figures 1
tables 1
includegraphics 1
labels 2
refs 2
floats 2

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors
```

```
This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

15: note            = "",

passed: False

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

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 [?].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 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

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬀigure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inﬂate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named ”images”

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

2

List of Figures

1 Example caption 4

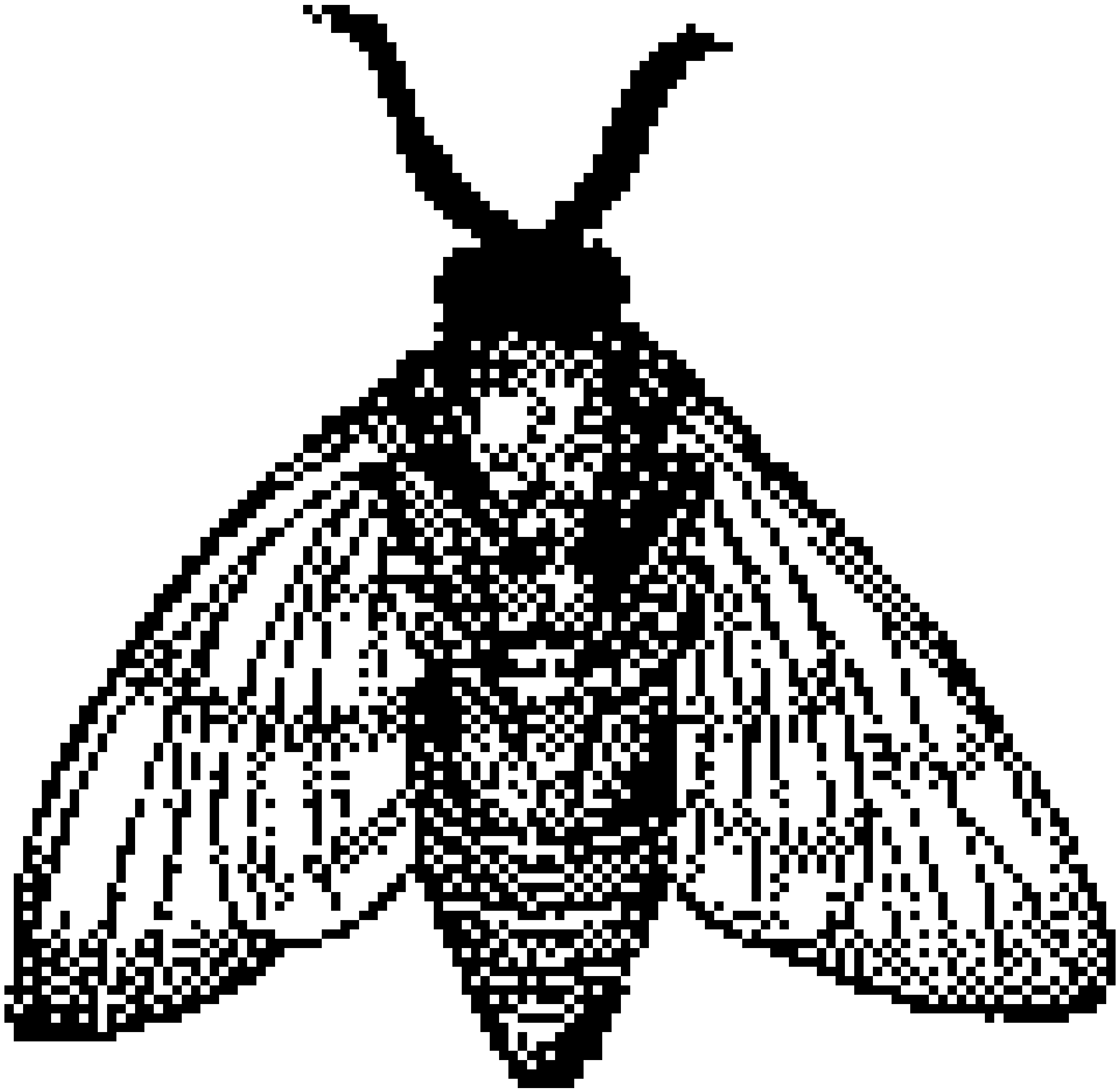


Figure 1: Example caption

4

List of Tables

1 My caption 6

Table 1: My caption

1	2	3
4	5	6
7	8	9

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtext _ label error
============================================


bibtext space label error
============================================


bibtext comma label error
============================================


latex report
============================================

        [2017-11-04 09.45.06] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
p.1   L62   : [editor00] undefined
p.1   L89   : 't:mytable' undefined
Empty `thebibliography' environment.
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
There were undefined citations.
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
There were undefined references.
Typesetting of "report.tex" completed in 1.2s.
../README.yml
  1:1       warning  missing document start "---"  (document-start)
  6:4       error    wrong indentation: expected 4 but found 3  (indentation)
```

```
  6:11      error    trailing spaces  (trailing-spaces)
  11:15     error    trailing spaces  (trailing-spaces)
  14:81     error    line too long (88 > 80 characters)  (line-length)
  14:88     error    trailing spaces  (trailing-spaces)
  20:4      error    wrong indentation: expected 4 but found 3  (indentation)
  20:11     error    trailing spaces  (trailing-spaces)
  26:72     error    trailing spaces  (trailing-spaces)
  30:4      error    wrong indentation: expected 4 but found 3  (indentation)
  31:11     error    trailing spaces  (trailing-spaces)
  36:61     error    trailing spaces  (trailing-spaces)
  38:4      error    duplication of key "type" in mapping  (key-duplicates)


================================================================================
Compliance Report
================================================================================


name: Lu, Junjie
hid:  214
paper1: 100% Oct 29th
paper2: 0%
project: 0%

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

6
wc 214 paper2 6 518 report.tex
wc 214 paper2 6 1163 report.pdf
wc 214 paper2 6 50 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True
```

```
find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

67: In Figure \ref{f:fly} we show a fly. Please note that because we
    use
73: \begin{figure}[!ht]
74: \centering\includegraphics[width=\columnwidth]{images/fly.pdf}
75: \caption{Example caption}\label{f:fly}
90: or generate them by hand while using the provided template in
    Table\ref{t:mytable}. Not ethat
93: \begin{table}[htb]
96: \label{t:mytabble}

figures 1
tables 1
includegraphics 1
labels 2
refs 2
floats 2

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
```

```
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

15: note            = "",

passed: False

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
```

```
--------------------------------------------------------------------------------

passed: True
```

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 [?].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 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

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below.* Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show.* It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

```
\section{Introduction} and NOT \section{INTRODUCTION}
```

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬃgure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named ”images”

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

```
/includegraphics[width=\columnwidth]{images/myimage.pdf}
```

re

List of Figures

1 Example caption 4

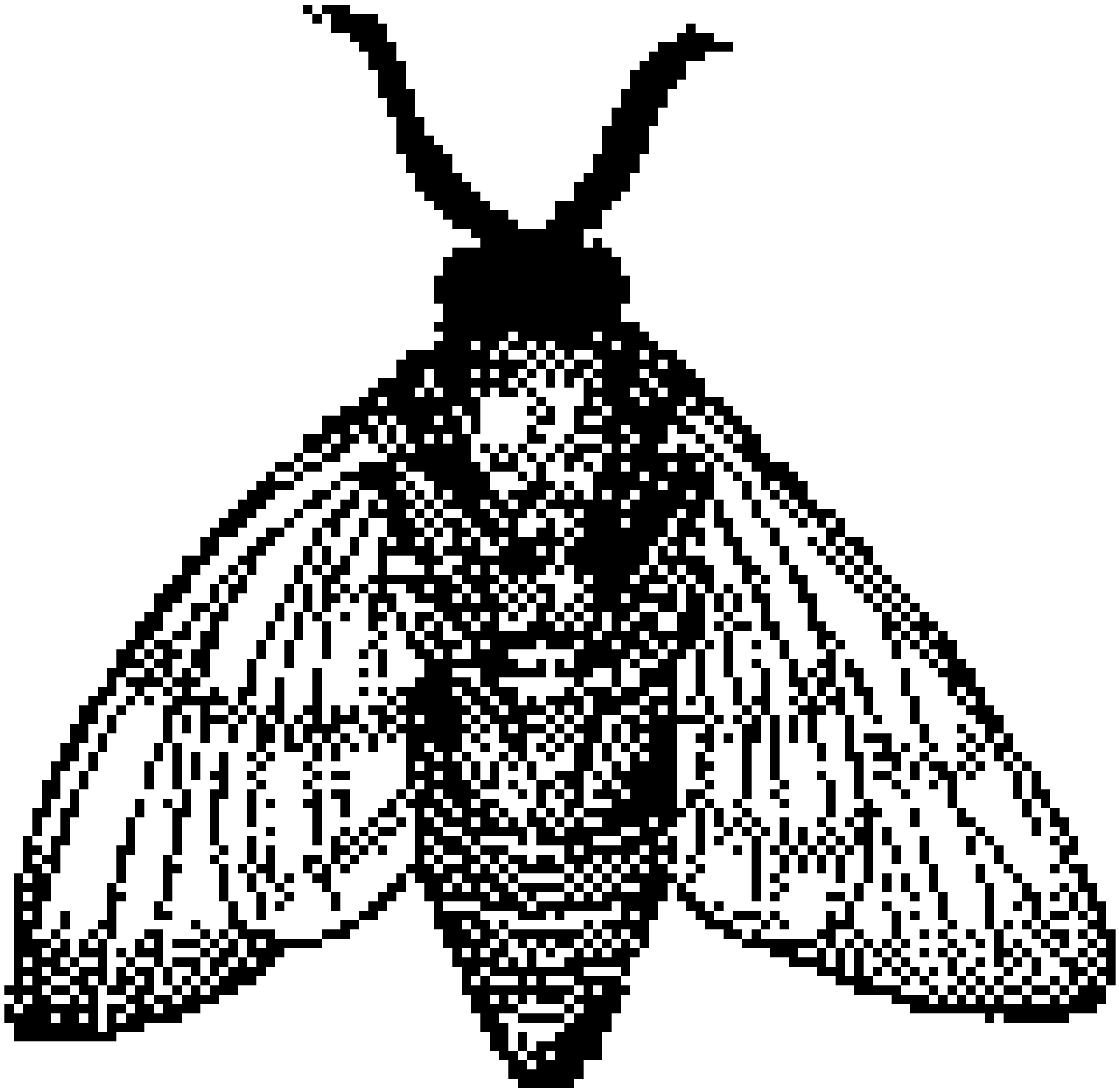


Figure 1: Example caption

4

List of Tables

1 My caption 6

Table 1: My caption

1	2	3
4	5	6
7	8	9

```
bibtext report
===========================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtext _ label error
===========================================


bibtext space label error
===========================================


bibtext comma label error
===========================================


latex report
===========================================

        [2017-11-04 09.45.12] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
p.1   L62   : [editor00] undefined
p.1   L89   : 't:mytable' undefined
Empty `thebibliography' environment.
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
There were undefined citations.
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
There were undefined references.
Typesetting of "report.tex" completed in 1.1s.
../README.yml
  9:81      error    line too long (82 > 80 characters)  (line-length)
  10:81     error    line too long (81 > 80 characters)  (line-length)
```

```
  11:81     error    line too long (81 > 80 characters)  (line-length)


================================================================================
Compliance Report
================================================================================


name: Niu, Geng
hid:  218
paper1: 100%
paper2: in progress

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

6
wc 218 paper2 6 518 report.tex
wc 218 paper2 6 1163 report.pdf
wc 218 paper2 6 50 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

67: In Figure \ref{f:fly} we show a fly. Please note that because we
```

```
    use
73: \begin{figure}[!ht]
74: \centering\includegraphics[width=\columnwidth]{images/fly.pdf}
75: \caption{Example caption}\label{f:fly}
90: or generate them by hand while using the provided template in
    Table\ref{t:mytable}. Not ethat
93: \begin{table}[htb]
96: \label{t:mytabble}

figures 1
tables 1
includegraphics 1
labels 2
refs 2
floats 2

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
```

```
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

15: note            = "",

passed: False

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

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 [?].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table??. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 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

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬁgure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named ”images”

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

List of Figures

1 Example caption 4

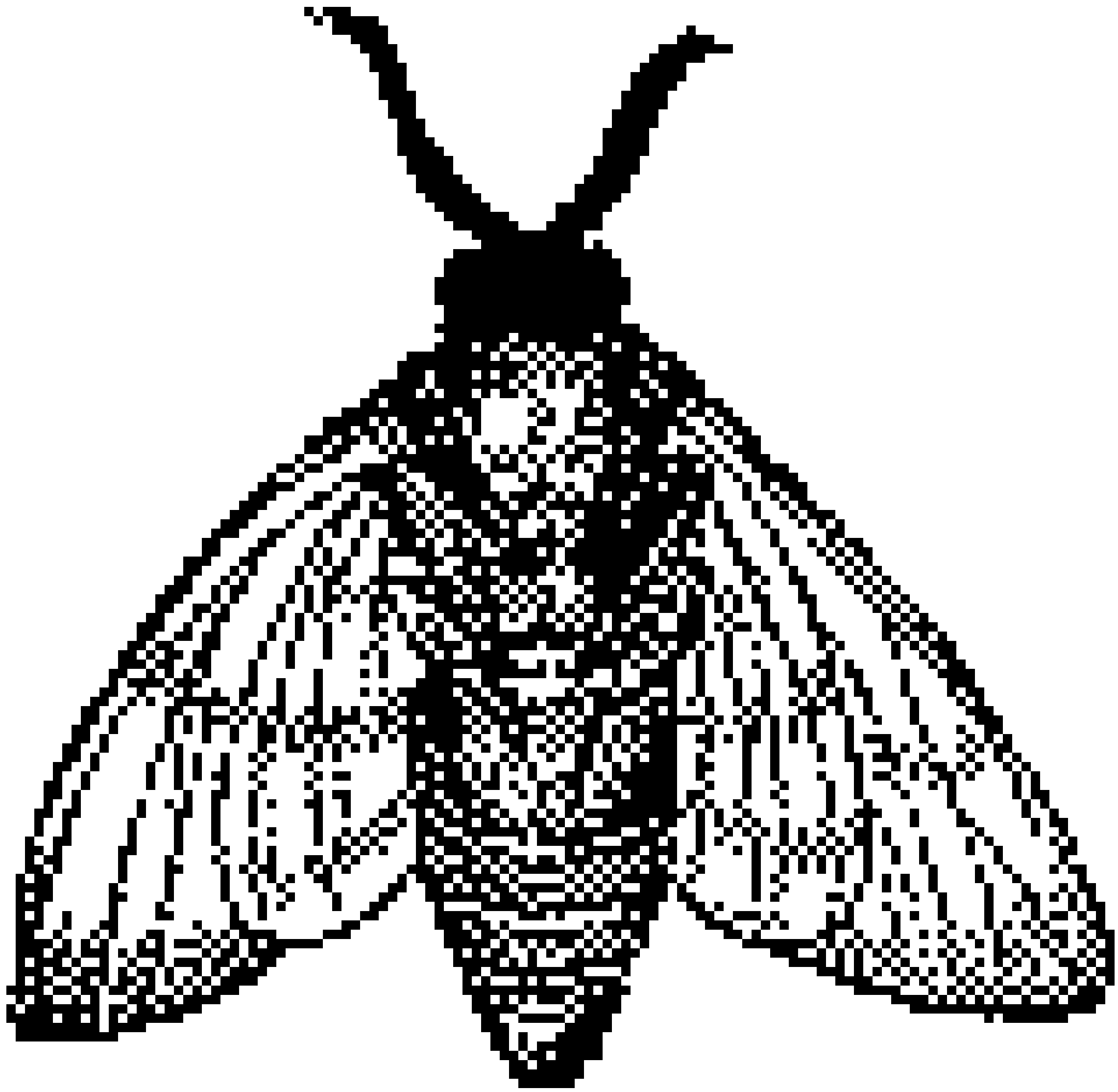


Figure 1: Example caption

4

List of Tables

1 My caption 6

Table 1: My caption

1	2	3
4	5	6
7	8	9

Big Data Applications in Aviation Industry

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

ABSTRACT

Data generated by aviation industry is being increased enormously. The data generated by all the components of aviation industry can be analysed for reducing the operational costs, predict customer behaviour, analyse customer satisfaction. These applications of big data in aviation industry makes it a prominent player. Hence, collecting this data, storing and processing them for desired results can help the aviation industry in boosting their profits and improve customer satisfaction.Various applications of Big data, their challenges and models are discussed here.

KEYWORDS

HID228, I523, Big Data, Aviation Industry,Analytics ,

1 INTRODUCTION

Big Data has transformed the businesses were being conducted. Every sector is integrated with data and generating huge amounts of data every day. All the companies are following data driven approach to crunch their competition. With the advent of concepts like Internet of things, the generation of data is increasing by many folds. This brings scope for a new business which handles the storage and analysis of data.

The solutions offered by Big Data in many industrial sections have revolutionised the respective businesses. In aviation industry, where data as big as 20tb is generated from an aircraft flying for an hour, Big Data can offer influential solutions in terms of dealing with the massive data. This data ,if is processed in an efficient way would increase the customer satisfaction at a reduced running and operational costs which in turn increases the profits.

2 APPLICATIONS

2.1 Baggage Handling

All the customer check-in their bag and have a doubt if their bags are being transported with them. There are several cases where customers raise some complaints about their bag being missed are bag transported to another destination. Traditional barcode system was used to handle this task. As the number of airline users increased, this solution was not profitable for customers and airline operators. However, this is being replaced by the new technology which uses radio frequencies to track real-time location of the bag. Bags which are checked in at the kiosk are assigned with a microchip. These chips will send the data related to the location of the bag frequently. The data generated by these chips is processed and stored. The processed data is available to the customers through mobile application or a web interface.

2.2 Flight Safety

All the flights have many sensors which generates a lot of data related to flight status and incidents. According to, a Being 737 generates nearly 20tb of data for one hour and an average cross international plane travelling for 6 hours generates 240 tb of data. Most of these data is related to safety and status of various equipment on the flights. A lot of this data should be filtered and mined to generate a meaningful and usable data. Southwest Airlines partnered with NASA for crunching this data and generating a meaningful data. NASA uses machine learning algorithms to mine this data.

This collected data from the flight can be analysed to decide a desired value for variables like altitude, wind speed, thrust, weight of the aircraft are proposed to the pilot for increased fuel economy. This data can also be helpful in deciding the nature of services according to the nature of the location and fuel costs.

2.3 Personalized promotion

In the advent of smart devices, all the industries including airline industry have come closer to the customer. Variables which are considered as characteristics are studied from the customer data available through their interaction with customers. These details range from preferences to behaviour of the customer. This data is analysed to study the behaviour of the customer and improve his experience with the airline industry.

2.4 Pricing strategies

Pricing is an important strategy to generate profits. It is quite often to see a price bump of the airfare during the payment or checkout process. This is because of increase in demand for the journey. This demand data is analysed in the servers and shown on the customers screen in less than minute. This calculations and analysis requires high computing power and efficient algorithms.

3 DATA SOURCES

3.1 In-Flight Data

QAR Data: Quick Access Recorder records the statistics of the flight like speed, height, speed, altitude, at any instance during flight. This data is stored in servers and processed

ACARS Data: Aircraft Communications Addressing and Reporting System is a online data transmission system which transmits data to ground through the aircraftfis satellite communication system. ACARS records values of different parameters during a event. An event is an action performed by the aircraft. The sensors mounted on the flightsfi brakes, wings, doors, send data to the ground staff using ACARS. Aircraft connection sensors and equipment monitoring system also uses this ACARS to transmit the data to ground.

3.2 Data from mobile and web applications

Now-a-days all the customer interactions with airline industry is through web. All the web applications and mobile applications which are developed for interacting with customer are smart enough to store the variables which are used to study the customer behaviour. This portion of data sources generates the data at increasing rates due to the evolution of customerfis interaction with internet.

3.3 Historical Data

Data available from the previous analytics and recordings constitutes a major portion. These are generally excel sheets or other forms of data stored in servers or files. These can be used for predictive analysis of the flight.

3.4 Other Sources

Other sources like weather sensors, internet, analysis from third party vendors which help airline industry in scheduling and predicting flight delays .

3.5 Service Oriented Model

4 DATA STORAGE AND PROCESSING

4.1 Apache Spark for Realtime data

5 CHALLENGES IN IMPLEMENTING BIG DATA

5.1 Information Security

6 CONCLUSION

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
I found no \citation commands---while reading file report.aux
I found no \bibstyle command---while reading file report.aux
(There were 2 error messages)
make[2]: *** [bibtex] Error 2

latex report
============================================

        [2017-11-04 09.45.17] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Typesetting of "report.tex" completed in 0.9s.
../README.yml
  8:81      error    line too long (133 > 80 characters)  (line-length)
  9:81      error    line too long (110 > 80 characters)  (line-length)
  9:110     error    trailing spaces  (trailing-spaces)
  10:81     error    line too long (87 > 80 characters)  (line-length)
  10:87     error    trailing spaces  (trailing-spaces)
  11:81     error    line too long (114 > 80 characters)  (line-length)
  11:114    error    trailing spaces  (trailing-spaces)
  12:81     error    line too long (114 > 80 characters)  (line-length)
  13:81     error    line too long (169 > 80 characters)  (line-length)
  13:169    error    trailing spaces  (trailing-spaces)
  14:81     error    line too long (155 > 80 characters)  (line-length)


================================================================================
Compliance Report
================================================================================


name: Swargam, Prashanth
hid:  228
paper1: Oct 20 17  100%
paper2: in progress
```

```
yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

2
wc 228 paper2 2 1003 report.tex
wc 228 paper2 2 922 report.pdf
wc 228 paper2 2 261 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

13: \renewcommand\footnotetextcopyrightpermission[1]{} % removes
    footnote with conference information in first column

passed: False

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------


figures 0
tables 0
includegraphics 0
labels 0
refs 0
floats 0

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
```

```
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
I found no \citation commands---while reading file report.aux
I found no \bibstyle command---while reading file report.aux
(There were 2 error messages)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

passed: True

ascii
--------------------------------------------------------------------------------
```

```
non ascii found 8217
non ascii found 8217
non ascii found 8217

================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

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 [**?**].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 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

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below.* Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show.* It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

\section{Introduction} and NOT \section{INTRODUCTION}

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬀigure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named ”images”

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

/includegraphics[width=\columnwidth]{images/myimage.pdf}

re

List of Figures

1 Example caption 4

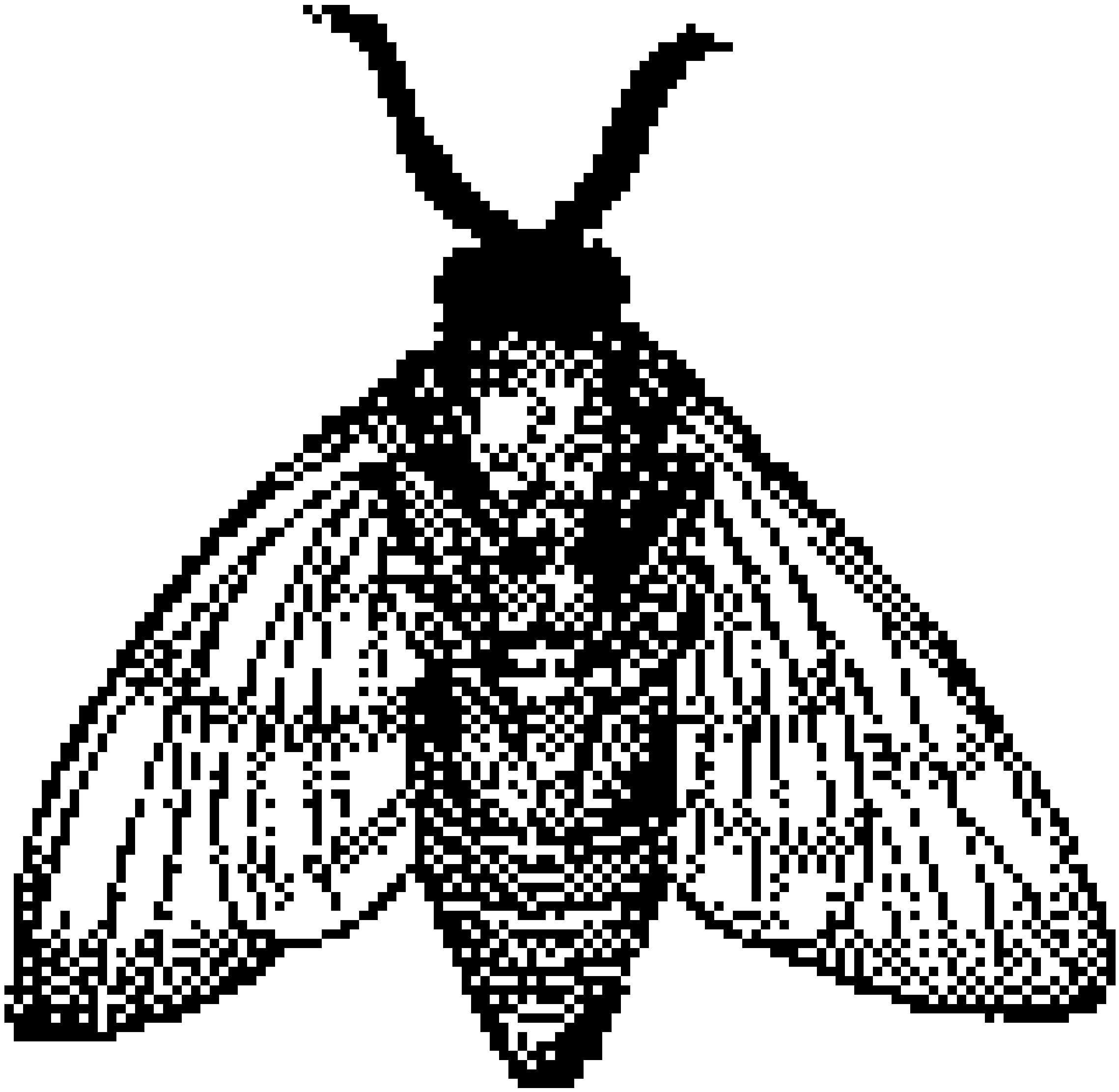


Figure 1: Example caption

4

List of Tables

1 My caption 6

Table 1: My caption

1	2	3
4	5	6
7	8	9

```
bibtext report
===========================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtext _ label error
===========================================


bibtext space label error
===========================================


bibtext comma label error
===========================================


latex report
===========================================

        [2017-11-04 09.45.35] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
p.1   L62   : [editor00] undefined
p.1   L89   : 't:mytable' undefined
Empty `thebibliography' environment.
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
There were undefined citations.
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
There were undefined references.
Typesetting of "report.tex" completed in 1.2s.
../README.yml
  9:81      error    line too long (86 > 80 characters)  (line-length)
  19:1      error    trailing spaces  (trailing-spaces)
```

```
  22:81     error    line too long (89 > 80 characters)  (line-length)
  22:89     error    trailing spaces  (trailing-spaces)
  23:77     error    trailing spaces  (trailing-spaces)
  24:81     error    line too long (106 > 80 characters)  (line-length)
  24:106    error    trailing spaces  (trailing-spaces)
  25:81     error    line too long (109 > 80 characters)  (line-length)
  25:109    error    trailing spaces  (trailing-spaces)
  33:9      error    trailing spaces  (trailing-spaces)
  34:1      error    trailing spaces  (trailing-spaces)
  37:81     error    line too long (88 > 80 characters)  (line-length)
  37:88     error    trailing spaces  (trailing-spaces)
  38:81     error    line too long (87 > 80 characters)  (line-length)
  38:87     error    trailing spaces  (trailing-spaces)
  39:49     error    trailing spaces  (trailing-spaces)
  46:9      error    trailing spaces  (trailing-spaces)


================================================================================
Compliance Report
================================================================================


name: Wu, Yujie
hid:  235
paper1: 100%, 10/27/2017
paper2: in progress

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

6
wc 235 paper2 6 518 report.tex
wc 235 paper2 6 1163 report.pdf
wc 235 paper2 6 50 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------
```

```
passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

67: In Figure \ref{f:fly} we show a fly. Please note that because we
    use
73: \begin{figure}[!ht]
74: \centering\includegraphics[width=\columnwidth]{images/fly.pdf}
75: \caption{Example caption}\label{f:fly}
90: or generate them by hand while using the provided template in
    Table\ref{t:mytable}. Not ethat
93: \begin{table}[htb]
96: \label{t:mytabble}

figures 1
tables 1
includegraphics 1
labels 2
refs 2
floats 2

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------
```

```
passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

15: note            = "",

passed: False

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
```

```
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

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 [?].

2 FIGURES

In Figure 1 we show a fly. Please note that because we use just columwidth that the size of the figure will change to the columnwidth of the paper once we change the layout to final. CHnaging the layout to final should not be done by you. All figures will be listed at the end.

[Figure 1 about here.]

When copying the example, please do not check in the images from the examples into your images directory as you will not need them for your paper. Instead use images that you like to include. If you do not have any images, do not dreate the images folder.

3 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

4 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

5 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

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use "quotes" , instead of " "

To emphasize a word, use *emphasize* and not "quote"

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬀigure and not a figure in text you copied from a text that has the ﬁ combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named "images"

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

2

List of Figures

1 Example caption 4

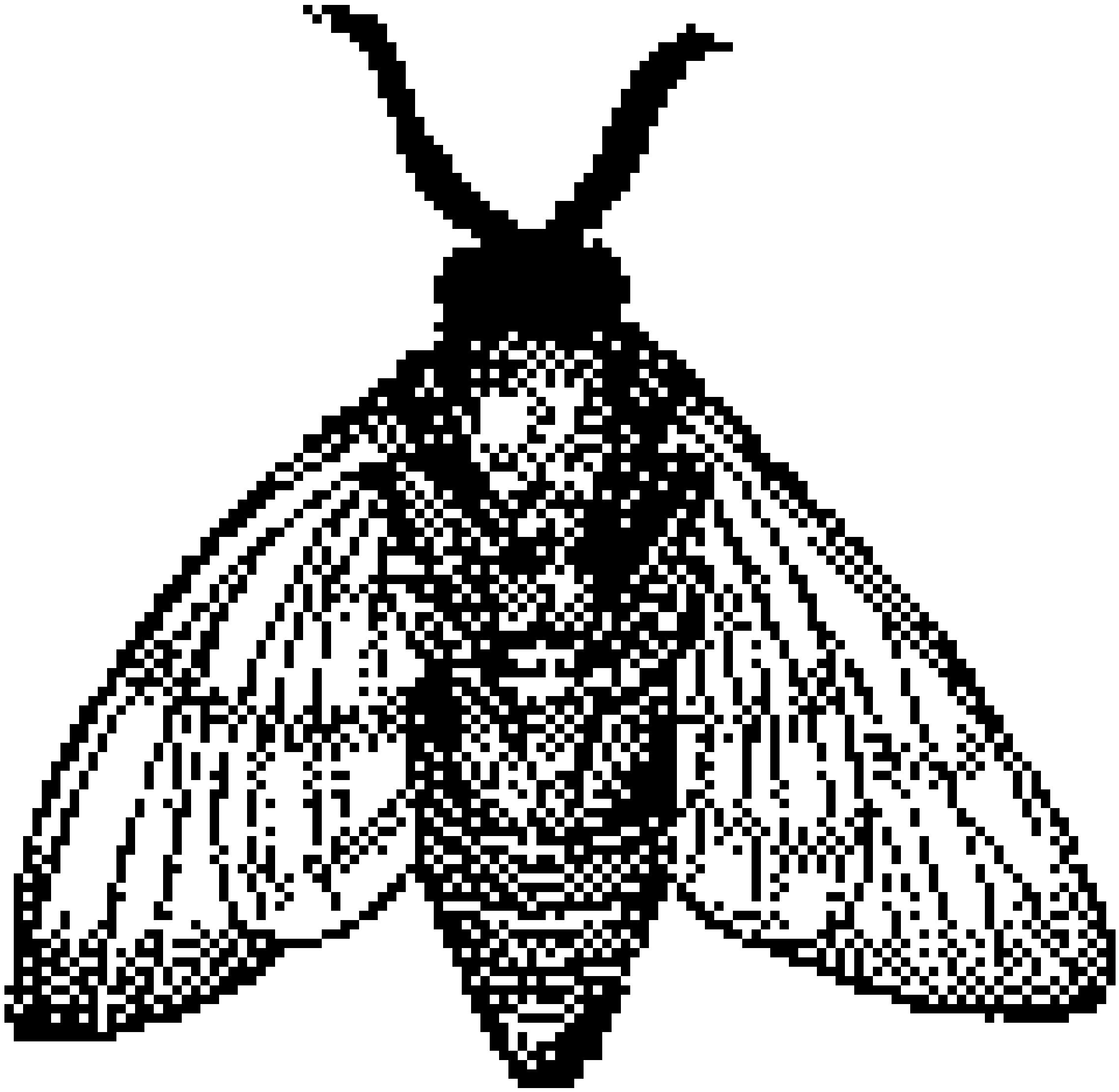


Figure 1: Example caption

List of Tables

1 My caption 6

Table 1: My caption

1	2	3
4	5	6
7	8	9

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtext _ label error
============================================


bibtext space label error
============================================


bibtext comma label error
============================================


latex report
============================================

        [2017-11-04 09.45.41] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
p.1   L62   : [editor00] undefined
p.1   L89   : 't:mytable' undefined
Empty `thebibliography' environment.
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
There were undefined citations.
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
There were undefined references.
Typesetting of "report.tex" completed in 1.1s.
../README.yml
  8:81      error    line too long (85 > 80 characters)  (line-length)
  9:81      error    line too long (84 > 80 characters)  (line-length)
```

```
  10:81     error    line too long (84 > 80 characters)  (line-length)
  11:81     error    line too long (83 > 80 characters)  (line-length)
  12:81     error    line too long (86 > 80 characters)  (line-length)
  13:81     error    line too long (83 > 80 characters)  (line-length)
  14:81     error    line too long (81 > 80 characters)  (line-length)
  15:81     error    line too long (84 > 80 characters)  (line-length)
  16:81     error    line too long (84 > 80 characters)  (line-length)
  37:1      error    trailing spaces  (trailing-spaces)
  39:4      error    wrong indentation: expected 4 but found 3  (indentation)
  39:17     error    trailing spaces  (trailing-spaces)
  41:4      error    wrong indentation: expected 7 but found 3  (indentation)
  43:4      error    wrong indentation: expected 7 but found 3  (indentation)
  47:81     error    line too long (117 > 80 characters)  (line-length)
  47:117    error    trailing spaces  (trailing-spaces)
  48:81     error    line too long (121 > 80 characters)  (line-length)
  49:81     error    line too long (123 > 80 characters)  (line-length)
  50:75     error    trailing spaces  (trailing-spaces)


================================================================================
Compliance Report
================================================================================


name: Ahmed, Tousif
hid:  237
paper1: 100%, October 27, 2017
paper2: in progress
project: in progress

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

6
wc 237 paper2 6 518 report.tex
wc 237 paper2 6 1163 report.pdf
wc 237 paper2 6 50 report.bib

find "
--------------------------------------------------------------------------------

passed: True
```

```
find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

67: In Figure \ref{f:fly} we show a fly. Please note that because we
    use
73: \begin{figure}[!ht]
74: \centering\includegraphics[width=\columnwidth]{images/fly.pdf}
75: \caption{Example caption}\label{f:fly}
90: or generate them by hand while using the provided template in
    Table\ref{t:mytable}. Not ethat
93: \begin{table}[htb]
96: \label{t:mytabble}

figures 1
tables 1
includegraphics 1
labels 2
refs 2
floats 2

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction
```

```
find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--I didn't find a database entry for "editor00"
(There was 1 warning)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

15: note            = "",

passed: False

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
```

```
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Prediction of psychological traits based on Big Data classification of associated social media footprints

Gagan Arora
Indiana University
2709 E 10th St
Bloomington, Indiana 47401
gkarora@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

Big Data, Edge Computing i523, psychological traits, Big Data, Facebook Data, Social media, digital foot prints, five factor model, personality traits

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.

2 INTRODUCTION

With the advancement of digital media and social media networks, there has been enormous amount of human activities, which is recorded as the digital footprints. According to IBM, in 2012 on an average 500 MB of personal data is uploaded to the online digital database. This data is either in the form of social media activities such as Facebook likes, Facebook comments, profile picture upload, tweets or in the form offline transactions where person goes to grocery shopping and pays using credit card. According to [1] China is investing heavy technological resources to mine this data along with personfis financial transactions to build social credit system. This project is expected to be implemented by 2020. There has been studies [4] fi?! [13], which analyzed the behavior outcomes of the digital profile with the actual characteristics of an individual. Interesting thing about these studies is that human behavior can be mapped statistically to define similarities and differences between individuals. This can further be used to build recommendation-based system to enrich social medial networks such as Facebook, LinkedIn, and Twitter etc. These studies [4] fi?! [13] further contributes in radically improving our behavior understanding of humans. [9] discusses about the predictability of individualfis psychological traits using statistical approach to arrive at the personality traits with certain confidence level. Psychological traits automation can further be used to enrich the quality of recommendation based systems and online search engines. [6] suggest how these studies [4] fi?! [13] can be used to improve online marketing systems. With so many advantages on one side, on other side it possesses biggest challenge to the Data privacy [5] - [11]. Reason why these studies [4] fi?! [13] provide better estimate of human psychological traits as compared to results of psychometric test because these study results [4] fi?! [13] takes the data of prolonged history. However, psychometric tests on the other hands is for few minutes or hours where human can manipulate response in order to achieve desire results. Thus, these studies [4] fi?! [13] can also be leveraged in employee hiring process where many companies still relies on psychometric tests.

3 DATA SOURCE OF BIG DATA IN DIGITAL WORLD

This section discusses how we can import, store and preprocess digital big data. This data can be fetched online via REST api or its direct available to download from website such as mypersonality.org. This site stores the social media data of close to six million participants. There are other sites like Stanford network analysis project, which contains enormous amount of data in the form of product reviews, Tweets, and social media data. Social medial sites like Instagram and Twitter provides public rest APIs through which we can access data, which is public. Other example is Amazon.com, which provides elegant web services to access product reviews. For preprocessing of this data, Python provides excellent libraries to access [via web service call] and preprocess data.

4 HUMAN BEHAVIOR AND PERSONALITY

[12] talks about various models, which can be used to describe human personality. Among all, five-factor model [FFM] is proved to be the best model to describe human behavior, psychological traits and preferences: Openness, Conscientiousness, Extroversion, Agreeableness and Emotional stability. We have data, we have psychological traits, and biggest challenge lies in extracting value out of big data and mapping the result to psychological traits. To accomplish this challenge we can perform singular value decomposition to map the qualitative data to quantitative data. To elaborate this further let us take an example: we have a Facebook likes of 10 million people and we filter down top 100 Facebook pages, which are of relevance. Top 100 relevant pages are those, which can predict factors mentioned in FFM. Now we will prepare Boolean matrix with Facebook user profile on vertical axis and Facebook page as horizontal axis. In simple words row represents Facebook user and column represents Facebook page. We will mark the coordinate as one if corresponding Facebook user [on vertical axis] likes a page [on horizontal axis] otherwise zero. Therefore, matrix will look like this:

$$
\begin{array}{c}
\\ user_1 \\ user_2 \\ user_3 \\ user_n
\end{array}
\begin{array}{c}
\begin{array}{cccc} fbPage_1 & fbPage_2 & \dots & fbPage_n \end{array} \\
\begin{pmatrix}
1 & 0 & \dots & 1 \\
0 & 1 & \dots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \dots & 1
\end{pmatrix}
\end{array}
$$

These 100 Facebook pages is clustered, based on the five factors mentioned in FFM. First twenty pages will represent first factor, second twenty pages will represent second factor and so on. Next step would be to build correlation matrix that represents how each person is correlated with each other based on the five factors. This matrix will be N by N where is N is number of Facebook users in this experiment. This matrix will help to determine how similar Facebook users are. Which will help us to build the recommendation based systems because similar peoples tends to like same pages and share same psychological traits. This correlation matrix will look like this:

$$
\begin{array}{c}
\\ user_1 \\ user_2 \\ user_3 \\ user_n
\end{array}
\begin{array}{c}
\begin{array}{cccc} user_1 & user_2 & \dots & user_n \end{array} \\
\begin{pmatrix}
1 & .75 & \dots & .85 \\
.75 & 1 & \dots & .91 \\
\vdots & \vdots & \ddots & \vdots \\
.85 & .91 & \dots & 1
\end{pmatrix}
\end{array}
$$

Step 1: Build binary matrix with Facebook user profile on vertical axis and Facebook page as horizontal axis.
Step 2: Populate the binary matrix with one and zero depending on if person has liked the page or not.
Step 3: Sort Facebook page columns depending on the factors mentioned in FFM.
Step 4: Use this matrix to build correlational matrix represents how each person is correlated with each other based on the five factors.
Step 5: Apply k-mean algorithm to group Facebook users of similar factors mentioned in FFM.

5 COMPUTER BASED PERSONALITY JUDGEMENT AND HUMAN BASED PERSONALITY JUDGEMENT

Research [14] has shown computer based personality judgments are more accurate than those made by humans. According to [14] perceiving and judging peoplefis personality is an important component of living society. Many cognitive decision made by humans are based on the judgement they have in their mind. This research [14] has shown how advance machine learning algorithms and statistical tools can be used to predict the personality traits and compared the results with the human judgments. This research also addresses the issue of substantiating the qualitative aspects of behavior with the quantitative parameters. Computer based personality judgment is not only based on machine learning or statistics but computer vision algorithms can also be used to distinguish facial emotions and concluding psychological traits.

6 SOCIAL NETWORK AS A PERSONALITY TRAIT PREDICTOR

[10] studies suggest how valuable social network is in predicting the psychological traits. According to [10], It is considered as one of the valuable digital footprints to predict intimate personal traits. For instance, number of friends and their location can be used to grade first factor of FFM, which is openness. Personfis romantic partner can be detected depending on the social network overlap of each friend, which can further be analyzed to predict onefis sexual preference. These predictions can further be statistically analyzed to [14] to know how accurate predictions are. We can use social network data on the algorithm discussed in the fiHuman Behavior and Personalityfi and conclude a very strong predictions on the psychological traits of a person. It has been in the news that 2016 elections were strategized with the help of the social media big data which will be discussed in the next section.

7 SOCIAL MEDIA BIG DATA AND ITS IMPACT ON POLITICAL ELECTIONS

[2] suggests how last year elections were revolutionized by the impact of big data of social media. Using statistical and machine learning algorithms on social media big data, political parties filtered down the data to identify their likely supporters and then channelized their strategy to win their votes. These strategies were less expensive than conducting campaigns at various places. Traditional analysis is generally based on the survey which is in the sense is limited [3] but now with the ease of big social media data, analysis is more accurate and conclusive. There has been sophisticated tools available that can predict the personfis race depending on his or her name and location. In recent election, political parties also combined social media data and public data [from census Bureau] to run sophisticated machine learning algorithm to pinpoint their supports. All these mentioned ways helped the political parties to micro target their supporters and gained their votes.

8 SOCIAL ACTIVITY fi THE PREDICTOR OF PERSONALITY

[10] suggests Facebook profile of a user is not static rather it also contains enriched records of digital footprints such as likes, comments, reactions to other posts. Such activities materializes the connections between user and content. This information along with the other activities such as playlist, browsing logs, online shopping activities and google queries can be used to develop sophisticated highly predictive FFM set for a user and with a very high confidence level can predict userfis age, gender, intelligence religious view and sexual orientation [10]. Very interesting example from the [10] suggests fiUsers who liked Hello Kitty brand tended to have high openness, low conscientiousness, and low agreeablenessfi fi?! strange but very interesting! [10] research further elaborate the importance of comments. Semantic analysis on comment can be analyzed to infer onefis personality as shown by the research: [8] and [7].

9 CONCLUSION

We discussed various ways in which social medial data can be utilized to build five factor personality model for a user. Main purpose here is to review the literature work done in this field and also presented the algorithm which can be used to translate qualitative data to quantitative data and hoe value can be extracted to build FFM for a user. We discussed computer based personality judgments are better than the human based personality judgments. We also touched based where social network can be used to predict userfis personality. As discussed earlier, these researches [4] fi?! [13] have proved to impact the general election last year in United states. Finally we concluded by showing evidences how social activity can be used to build the FFM for a user.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski and all the TA's for their support and suggestions to write this paper.

REFERENCES

[1] [n. d.]. ([n. d.]). https://www.ft.com/content/f772a9ce-60c4-11e7-91a7-502f7ee26895
[2] [n. d.]. ([n. d.]). https://www.nbcnews.com/politics/elections/how-big-data-broke-american-politics-n732901
[3] [n. d.]. ([n. d.]). https://www.vox.com/conversations/2017/3/16/14935336/big-data-politics-donald-trump-2016-elections-polarization
[4] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. 2011. The Social fMRI: Measuring, Understanding, and Designing Social Mechanisms in the Real World. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*. ACM, New York, NY, USA, 445–454. https://doi.org/10.1145/2030112.2030171
[5] Declan Butler. 2007. Data sharing threatens privacy. 449 (11 2007), 644–5.
[6] Ye Chen, Dmitry Pavlov, and John F. Canny. 2009. Large-scale Behavioral Targeting. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*. ACM, New York, NY, USA, 209–218. https://doi.org/10.1145/1557019.1557048
[7] Adam D. I. Kramer and Kerry Rodden. 2008. Word usage and posting behaviors: Modeling blogs with unobtrusive data collection methods. (01 2008), 1125-1128 pages.
[8] Samuel Gosling, Sam Gaddis, and Simine Vazire. 2007. Personality Impressions Based on Facebook Profiles. 7 (01 2007).
[9] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5802–5805. https://doi.org/10.1073/pnas.1218772110 arXiv:http://www.pnas.org/content/110/15/5802.full.pdf
[10] Renaud Lambiotte and Michal Kosinski. 2014. Tracking the Digital Footprints of Personality. 102 (12 2014), 1934–1939.
[11] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust De-anonymization of Large Sparse Datasets. (06 2008), 111-125 pages.
[12] Lewis R. Goldberg. 1993. The structure of phenotypic personality traits. 48 (02 1993), 26–34.
[13] Kern ML Dziurzynski L Ramones SM Agrawal M Shah A Kosinski M Stillwell D Seligman ME Ungar LH. Schwartz H, Eichstaedt JC. 2013. Personality, gender, and age in the language of social media: the open-vocabulary approach. (2013). https://www.ncbi.nlm.nih.gov/pubmed/24086296
[14] Youyou Wu, Michal Kosinski, and David Stillwell. 2015. Computer-based personality judgments are more accurate than those made by humans. 112 (01 2015).

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction}` and NOT `\section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

- Claims made without citations provided
- Need to paraphrase long quotations (whole sentences or longer)
- Need to quote directly cited material

A.6 Character Errors

- Erroneous use of quotation marks, i.e. use “quotes” , instead of ” ”
- To emphasize a word, use *emphasize* and not “quote”
- When using the characters & # % _ put a backslash before them so that they show up correctly
- Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.
- If you see a ffigure and not a figure in text you copied from a text that has the fi combined as a single character

A.7 Structural Issues

- Acknowledgement section missing
- Incorrect README file
- In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper
- The paper has less than 2 pages of text, i.e. excluding images, tables and figures
- The paper has more than 6 pages of text, i.e. excluding images, tables and figures
- Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

- Capitalization errors in referring to captions, e.g. Figure 1, Table 2
- Do use *label* and *ref* to automatically create figure numbers
- Wrong placement of figure caption. They should be on the bottom of the figure
- Wrong placement of table caption. They should be on the top of the table
- Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg
- Do not submit eps images. Instead, convert them to PDF
- The image files must be in a single directory named ”images”
- In case there is a powerpoint in the submission, the image must be exported as PDF
- Make the figures large enough so we can read the details. If needed make the figure over two columns
- Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to ffoat. For this class, you should place all figures at the end of the report.
- In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption
- Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)
- Do not use textwidth as a parameter for includegraphics
- Figures should be reasonably sized and often you just need to add columnwidth

e.g.

```
/includegraphics[width=\columnwidth]{images/myimage.pdf}
```

re

4

bibtext report
===

```
This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--no key, author in ref1
Warning--no author, editor, organization, or key in ref1
Warning--to sort, need author or key in ref1
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
```

```
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Warning--no key, author in ref14
Warning--no author, editor, organization, or key in ref14
Warning--to sort, need author or key in ref14
Warning--no key, author in ref15
Warning--no author, editor, organization, or key in ref15
Warning--to sort, need author or key in ref15
Warning--no key, author in ref1
Warning--no key, author in ref1
Warning--no key, author in ref14
Warning--no key, author in ref15
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
```

```
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Warning--no key, author in ref1
Warning--no author, editor, organization, or key in ref1
Warning--empty author in ref1
Warning--empty year in ref1
Warning--no key, author in ref14
Warning--no author, editor, organization, or key in ref14
Warning--empty author in ref14
Warning--empty year in ref14
Warning--no key, author in ref15
Warning--no author, editor, organization, or key in ref15
Warning--empty author in ref15
Warning--empty year in ref15
Warning--numpages field, but no articleno or eid field, in ref2
Warning--no journal in ref9
Warning--numpages field, but no articleno or eid field, in ref8
Warning--no journal in ref16
Warning--page numbers missing in both pages and numpages fields in ref16
Warning--no journal in ref13
Warning--no journal in ref11
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
```

```
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Warning--no journal in ref12
Warning--page numbers missing in both pages and numpages fields in ref12
(There were 64 error messages)
make[2]: *** [bibtex] Error 2

latex report
===========================================

        [2017-11-04 09.45.47] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
```

```
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Typesetting of "report.tex" completed in 1.3s.
../README.yml
  16:81     error    line too long (119 > 80 characters)  (line-length)
  22:81     error    line too long (117 > 80 characters)  (line-length)
  22:117    error    trailing spaces  (trailing-spaces)
  27:4      error    wrong indentation: expected 4 but found 3  (indentation)
  27:17     error    trailing spaces  (trailing-spaces)
  29:4      error    wrong indentation: expected 7 but found 3  (indentation)
  31:4      error    wrong indentation: expected 7 but found 3  (indentation)
  33:56     error    trailing spaces  (trailing-spaces)
```

```
  35:81     error    line too long (115 > 80 characters)  (line-length)
  35:115    error    trailing spaces  (trailing-spaces)
  36:81     error    line too long (115 > 80 characters)  (line-length)
  36:115    error    trailing spaces  (trailing-spaces)
  37:81     error    line too long (89 > 80 characters)  (line-length)
  37:89     error    trailing spaces  (trailing-spaces)
  38:81     error    line too long (131 > 80 characters)  (line-length)
  38:131    error    trailing spaces  (trailing-spaces)
  39:81     error    line too long (130 > 80 characters)  (line-length)
  39:130    error    trailing spaces  (trailing-spaces)
  40:62     error    trailing spaces  (trailing-spaces)
  41:25     error    trailing spaces  (trailing-spaces)


================================================================================
Compliance Report
================================================================================


name: Arora, Gagan
hid:  301
paper1: 100% Oct 29 17
paper2: 10%
project: in progress

yamlcheck
--------------------------------------------------------------------------------



wordcount
--------------------------------------------------------------------------------


4
wc 301 paper2 4 1960 report.tex
wc 301 paper2 4 2832 report.pdf
wc 301 paper2 4 1379 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True
```

```
find input{format/i523}
--------------------------------------------------------------------------------

5: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------



figures 0
tables 0
includegraphics 0
labels 0
refs 0
floats 0

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors
```

```
This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--no key, author in ref1
Warning--no author, editor, organization, or key in ref1
Warning--to sort, need author or key in ref1
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
```

```
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3085 of file ACM-Reference-Format.bst
Warning--no key, author in ref14
Warning--no author, editor, organization, or key in ref14
Warning--to sort, need author or key in ref14
Warning--no key, author in ref15
Warning--no author, editor, organization, or key in ref15
Warning--to sort, need author or key in ref15
Warning--no key, author in ref1
Warning--no key, author in ref1
Warning--no key, author in ref14
Warning--no key, author in ref15
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
```

```
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3131 of file ACM-Reference-Format.bst
Warning--no key, author in ref1
Warning--no author, editor, organization, or key in ref1
Warning--empty author in ref1
Warning--empty year in ref1
Warning--no key, author in ref14
Warning--no author, editor, organization, or key in ref14
Warning--empty author in ref14
Warning--empty year in ref14
Warning--no key, author in ref15
Warning--no author, editor, organization, or key in ref15
Warning--empty author in ref15
Warning--empty year in ref15
Warning--numpages field, but no articleno or eid field, in ref2
Warning--no journal in ref9
Warning--numpages field, but no articleno or eid field, in ref8
Warning--no journal in ref16
Warning--page numbers missing in both pages and numpages fields in ref16
Warning--no journal in ref13
Warning--no journal in ref11
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
```

```
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "Schwartz H, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones
while executing---line 3229 of file ACM-Reference-Format.bst
Warning--no journal in ref12
Warning--page numbers missing in both pages and numpages fields in ref12
(There were 64 error messages)

bibtex_empty_fields
--------------------------------------------------------------------------------

entries in general should not be empty in bibtex

find ""
--------------------------------------------------------------------------------

passed: True

ascii
--------------------------------------------------------------------------------

non ascii found 8217
```

```
non ascii found 8211
non ascii found 8211
non ascii found 8217
non ascii found 8211
non ascii found 8211
non ascii found 8211
non ascii found 8211
non ascii found 8217
non ascii found 8217
non ascii found 8217
non ascii found 8220
non ascii found 8221
non ascii found 8217
non ascii found 8211
non ascii found 8217
non ascii found 8220
non ascii found 8221
non ascii found 8211
non ascii found 8217
non ascii found 8217
non ascii found 8211

================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

89: \textbf{\textit{Step 1}}: Build binary matrix with Facebook user
    profile on vertical axis and Facebook page as horizontal
    axis.\newline

90: \textbf{\textit{Step 2}}: Populate the binary matrix with one and
    zero depending on if person has liked the page or not.\newline

91: \textbf{\textit{Step 3}}: Sort Facebook page columns depending on
    the factors mentioned in FFM.\newline

92: \textbf{\textit{Step 4}}: Use this matrix to build correlational
    matrix represents how each person is correlated with each other
    based on the five factors.\newline

passed: False
```

```
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Hadoop and MongoDB in support of Big Data Applications and Analytics

Sushant Athaley
Indiana University
sathaley@iu.edu

ABSTRACT

Big data processing is beyond capability of traditional tool. It requires specialized tools like Hadoop and MongoDB. We will explore Haddop and MongoDB technically as a tool and how they provide support/help in big data analysis. TBD

KEYWORDS

i523, hid302, big data, Hadoop, MongoDB

1 INTRODUCTION

Describe about big data, hadoop and mongodb. Describe what this paper will do. Papers organization.

2 BIG DATA

Big Data is defined in lot many different ways but one of the interesting ways it has been defined is in terms of three V's which are Volume, Velocity, and Variety. Big data is generated in great *volume* typically in the gigabyte or more which makes data processing difficult. Data *velocity* has been increased due to the real-time data streaming from various applications like social media or different type of sensors recording data continuously. Big data comes in *variety* of format like structured or unstructured data. Data varies in various format like text, pictures, audio, videos, 3D, social media and so on. These big data characteristics pose challenges in terms of overall data lifecycle management. Some of the examples of big data usage are the recommendation service, predictive analytics, data analytics, pattern identification, and machine learning. Traditional systems are good for small or medium data processing but unable to provide support for the big data. Big data need specialized technologies and tools to handle its characteristics. The technologies which can solve big data problem should have capabilities like distributed computing system, massively parallel processing, NoSQL, and analytical database [1, Ch. 1, p. 4]. Can Hadoop or MongoDB be those technologies who can provide that support?

3 HADOOP

introduce hadoop, architecture, support to big data, real life examples Apache foundation describes Hadoop as "The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures" [2]. In other words, Hadoop provides a framework to store data in the distributed manner and provides the capability to run data analysis in the distributed way.

"Currently Hadoop project includes following modules:

- **Hadoop Common**: The common utilities that support the other Hadoop modules.
- **Hadoop Distributed File System (HDFS)**: A distributed file system that provides high-throughput access to application data.
- **Hadoop YARN**: A framework for job scheduling and cluster resource management.
- **Hadoop MapReduce**: A YARN-based system for parallel processing of large data sets" [2].

3.1 Hadoop Common

Hadoop Common are the collection of the utilities to support the Hadoop modules. This is the core package which provide essential and basic service of the framework.

3.2 Hadoop Distributed File System (HDFS)

Hadoop Distributed File System (HDFS) is the default distributed file system provided by the Hadoop. HDFS serves as storage mechanism in the Hadoop framework. HDFC specifically designed to process large data set and run on low cost hardware. It is high fault tolerant which contains mechanism for quick fault detection and auto recovery. HDFS is designed to port across heterogeneous hardware and software platform. It does data computation on same node instead of moving data to the server which is faster as well as avoid network congestion. It provides scalability by adding or removing nodes in the HDFS cluster and can support hundreds of nodes in single cluster [4]. Figure 1 shows HDFS architecture.

[Figure 1 about here.]

HDFS is based on master/slave architecture where NameNode is the master server and DataNodes are the slave nodes. There can be only one NameNode server which manages file system name space and all read write requests. NameNode doesn't store any data but contains all the meta-data about files and DataNodes. DataNode contains actual data and they can be multiple in numbers usually one per node. DataNodes are responsible for the create, delete, replicate of the datablocks on the node as per the instruction by the NameNode. DataNode also sends block-report to NameNode which has list of all blocks on the DataNode. DataNode sends heartbeat message to NameNode which helps in identifying the failure nodes. If heartbeat is not received by NameNode in specified interval then that DataNode is marked as dead and NameNode usage different DataNode. Figure 2 and 3 depicts read and write in HDFS respectively.

[Figure 2 about here.]

[Figure 3 about here.]

3.3 Hadoop YARN

Hadoop YARN (Yet Another Resource Negotiator) provides cluster resource management which helps in running multiple distributed application in Hadoop. YARN consists of 3 components *ResourceManager (RM), NodeManager (NM)* and *ApplicationMaster (AM)*. ResourceManager is the master process which manages resources across the nodes. NodeManager is responsible for the container and provide resource usage to the ResourceManager. ApplicationMaster is responsible for getting resources from ResourceManager and work with NodeManager to execute the task [3]. YARN makes it possible to run different application on Hadoop platform which makes it scalable and integrable [1, Ch. 3, p. 65]. Figure 4 shows YARN architecture.

[Figure 4 about here.]

3.4 Hadoop MapReduce

Hadoop MapReduce is a framework which provides capability to process vast amount of data in distributed manner. Processing is done in parallel on various nodes utilizing local machine processor and memory which results in high computation power. Framework provides fault tolerance along with supporting large clusters usually thousands of nodes. Typical framework processing is to split input data into independent chunks and then processed by *map* tasks in parallel. Sort the output of the map task and then provide that as input to *reduce* task for aggregate processing. Two important class in this framework are org.apache.hadoop.mapreduce.Mapper and org.apache.hadoop.mapreduce.Reducer. They respectively provides map and reduce method to process the data.

3.5 Big Data Support

4 MONGODB

5 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as https://www.tablesgenerator.com/ or other such tools (please google for them). They even allow you to manage tables as CSV.

or generate them by hand while using the provided template in Table**??**. Not ethat the caption is before the tabular environment.

[Table 1 about here.]

6 LONG EXAMPLE

If you like to see a more elaborate example, please look at report-long.tex.

7 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] Shiva Achari. 2015. *Hadoop Essentials.* Packt Publishing, Birmingham.
[2] Apache. [n. d.]. Apache Hadoop. ([n. d.]). http://hadoop.apache.org/
[3] Apache. 2017. Apache Hadoop YARN. web. (2017). http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html
[4] Apache. 2017. HDFS Architecture. web. (2017). http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html

We include an appendix with common issues that we see when students submit papers. One particular important issue is not to use the underscore in bibtex labels. Sharelatex allows this, but the proceedings script we have does not allow this.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correcty, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

A.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

A.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

A.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

A.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

A.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

A.6 Character Errors

Erroneous use of quotation marks, i.e. use "quotes" , instead of " "

To emphasize a word, use *emphasize* and not "quote"

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬃgure and not a figure in text you copied from a text that has the fi combined as a single character

A.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

A.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named "images"

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

List of Figures

1 HDFS Architecture [4] 5
2 HDFS Read [1, Ch. 3, p. 38] 6
3 HDFS Write [1, Ch. 3, p. 39] 7
4 YARN Architecture [3] 8

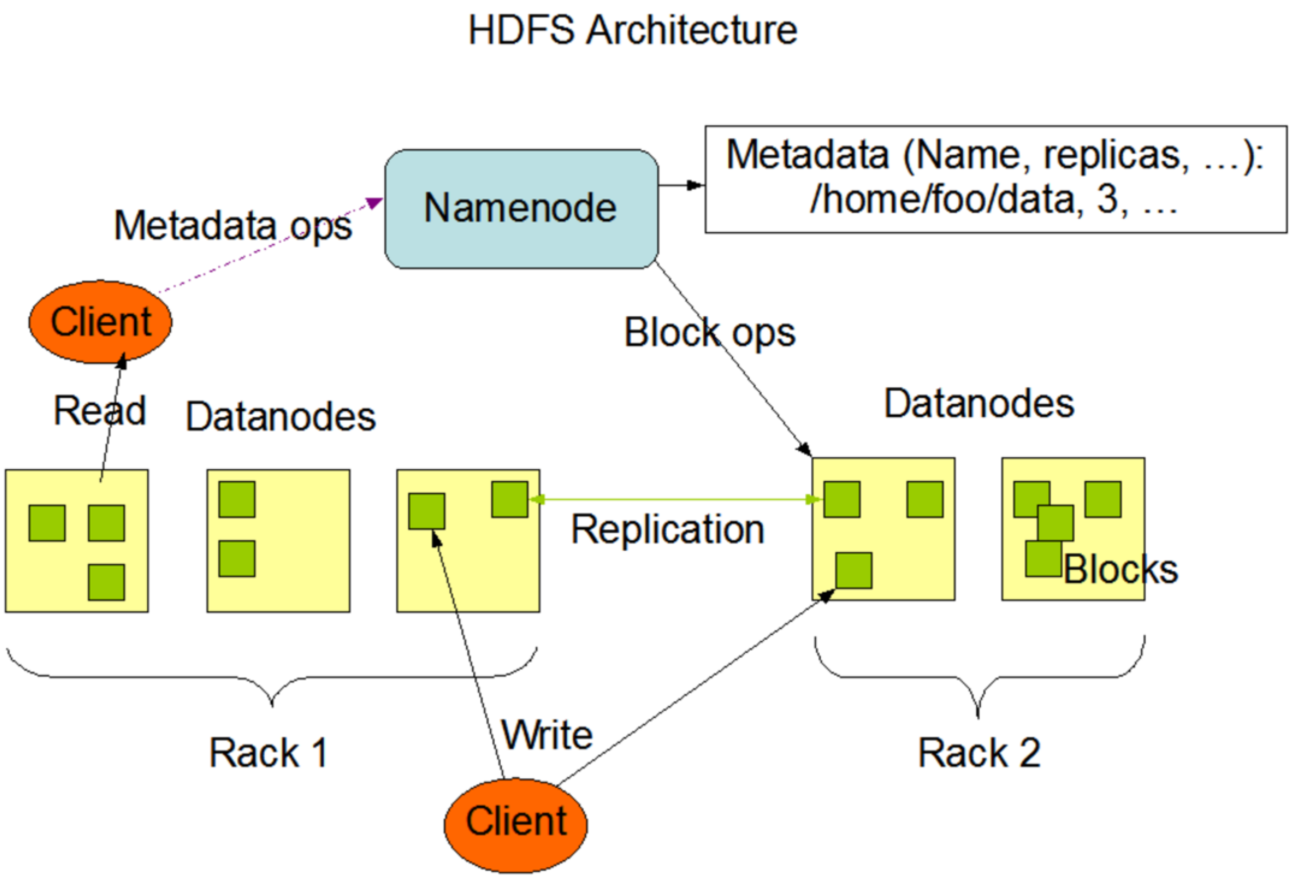


Figure 1: HDFS Architecture [4]

5

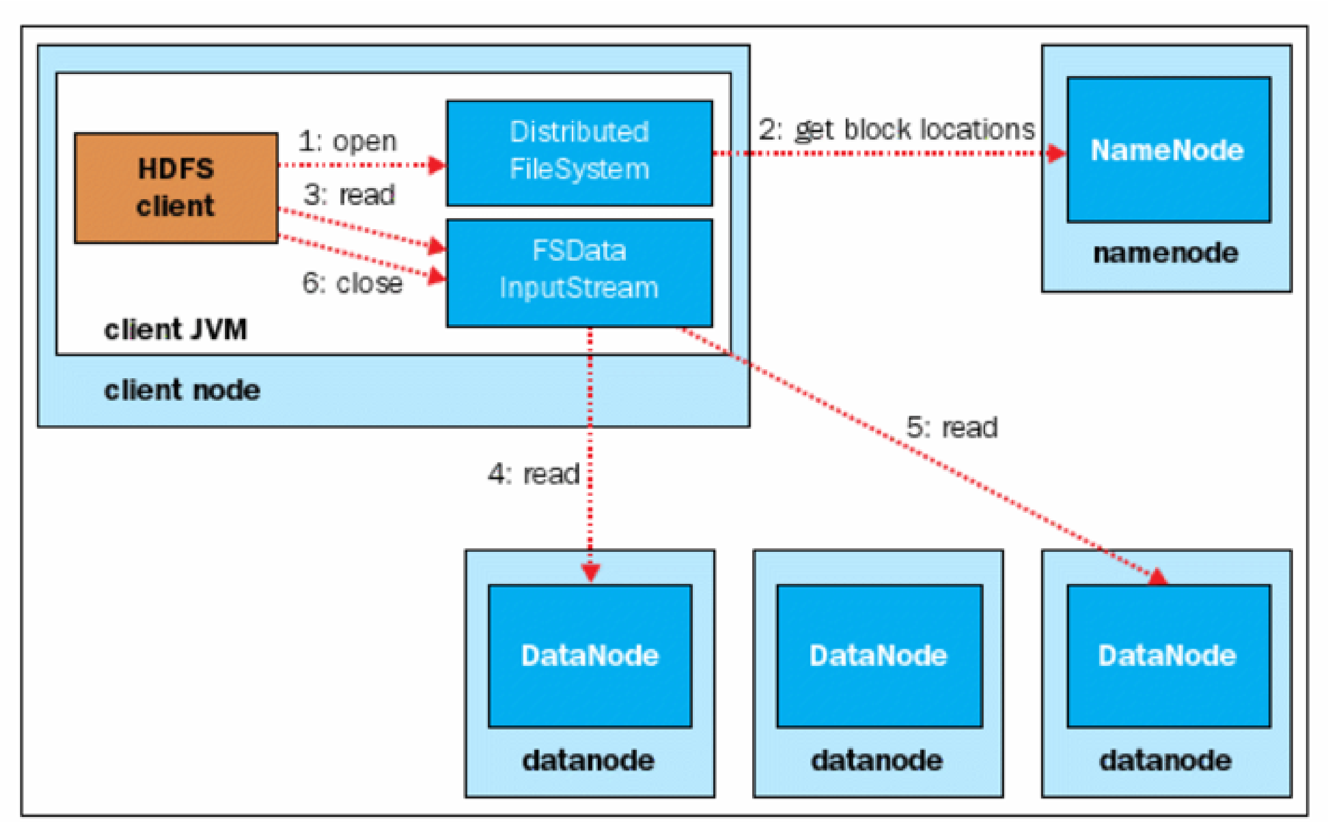


Figure 2: HDFS Read [1, Ch. 3, p. 38]

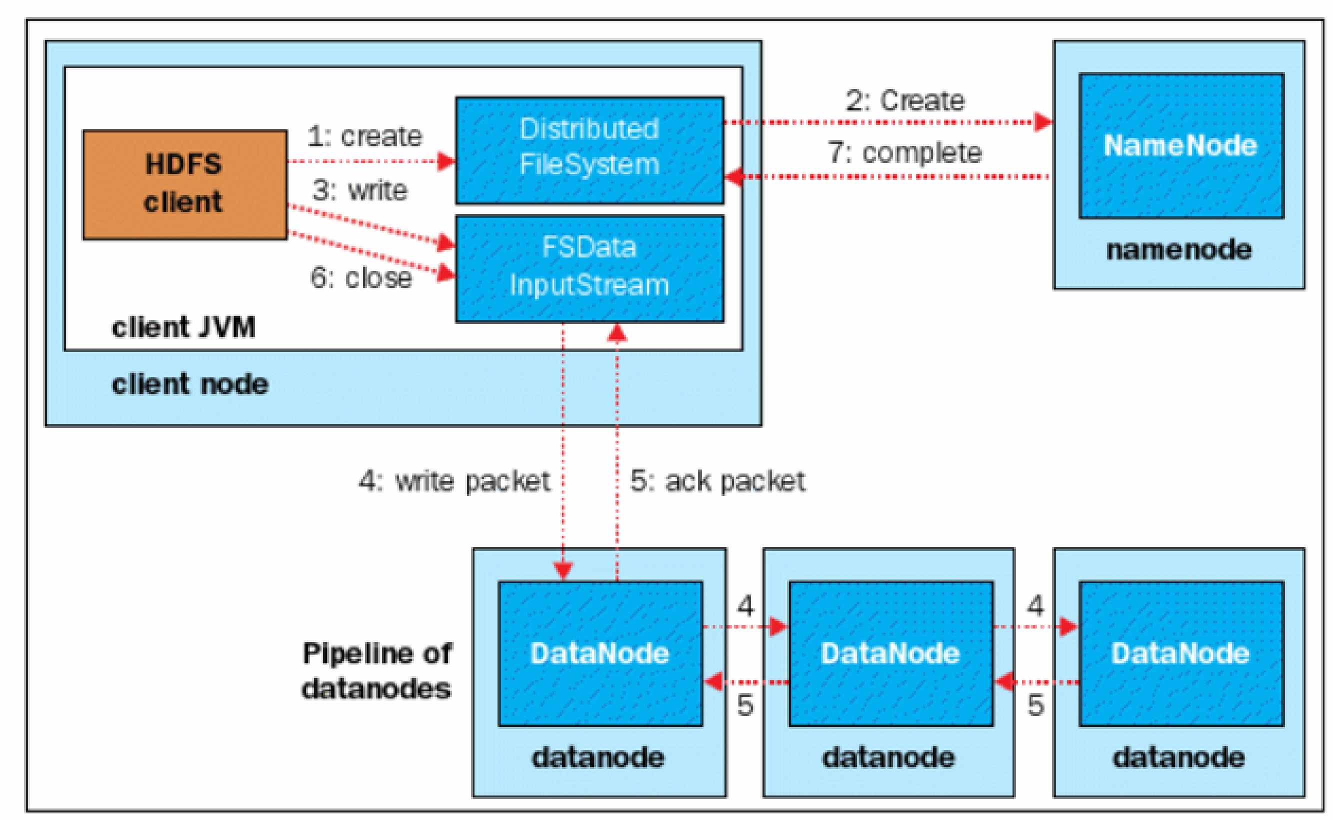


Figure 3: HDFS Write [1, Ch. 3, p. 39]

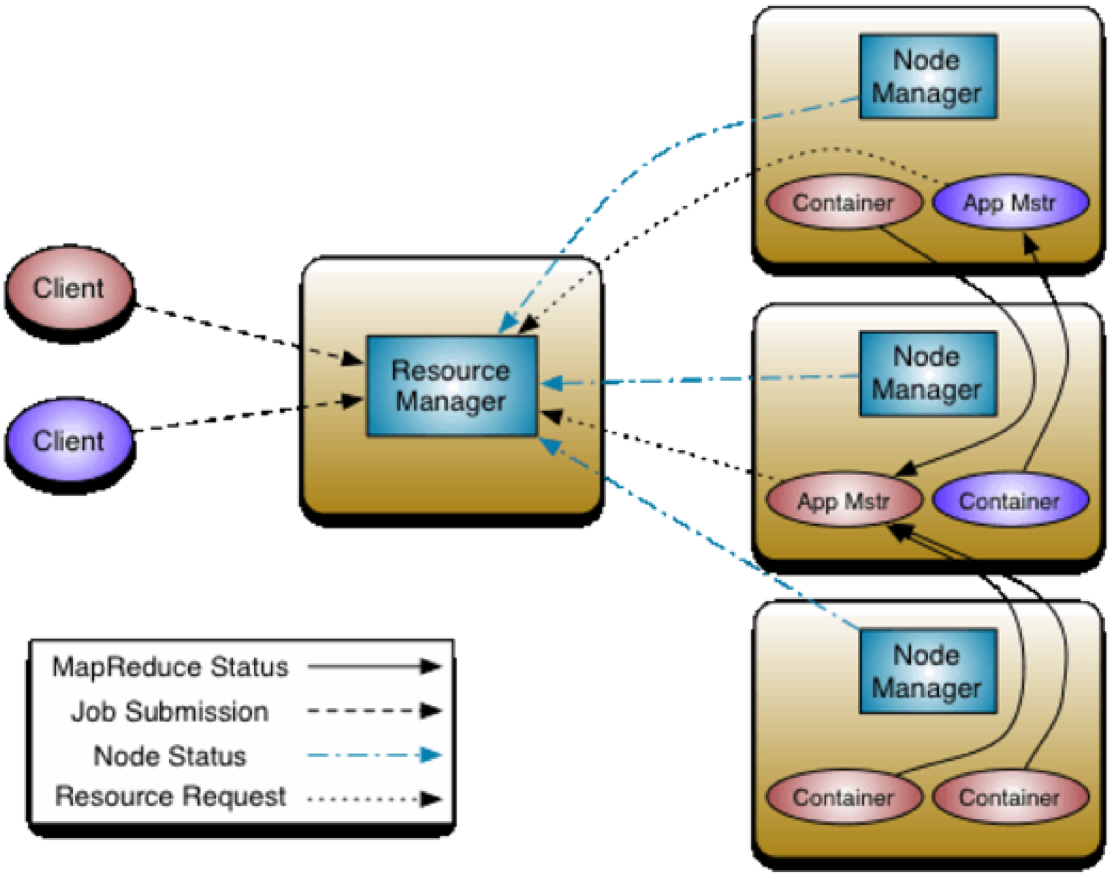


Figure 4: YARN Architecture [3]

List of Tables

1 My caption 10

Table 1: My caption

1	2	3
4	5	6
7	8	9

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--empty year in www-hadoop
(There was 1 warning)

bibtext _ label error
============================================


bibtext space label error
============================================


bibtext comma label error
============================================


latex report
============================================

        [2017-11-04 09.45.53] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
p.2   L96   : 't:mytable' undefined
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
The anchor of a bookmark and its parent's must not be the same. Added a new anchor.
There were undefined references.
Typesetting of "report.tex" completed in 1.5s.

================================================================================
Compliance Report
================================================================================
```

```
name: Sushant Athaley
hid:  302
paper1: Nov 3 2017 100%
paper2: 30%

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

10
wc 302 paper2 10 1325 report.tex
wc 302 paper2 10 2017 report.pdf
wc 302 paper2 10 214 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

55: Hadoop Distributed File System (HDFS) is the default distributed
    file system provided by the Hadoop. HDFS serves as storage
    mechanism in the Hadoop framework. HDFC specifically designed to
    process large data set and run on low cost hardware. It is high
    fault tolerant which contains mechanism for quick fault detection
    and auto recovery. HDFS is designed to port across heterogeneous
    hardware and software platform. It does data computation on same
    node instead of moving data to the server which is faster as well
    as avoid network congestion. It provides scalability by adding or
```

```
removing nodes in the HDFS cluster and can support hundreds of nodes in single cluster \cite{www-hdfs-arch}. Figure \ref{f:hdfs-arch} shows HDFS architecture.
56: \begin{figure}[!ht]
57: \centering\includegraphics[width=\columnwidth]{images/hdfsArch.PNG}
58: \caption{HDFS Architecture \cite{www-hdfs-arch}}\label{f:hdfs-arch}
61: HDFS is based on master/slave architecture where NameNode is the master server and DataNodes are the slave nodes. There can be only one NameNode server which manages file system name space and all read write requests. NameNode doesn't store any data but contains all the meta-data about files and DataNodes. DataNode contains actual data and they can be multiple in numbers usually one per node. DataNodes are responsible for the create, delete, replicate of the datablocks on the node as per the instruction by the NameNode. DataNode also sends block-report to NameNode which has list of all blocks on the DataNode. DataNode sends heartbeat message to NameNode which helps in identifying the failure nodes. If heartbeat is not received by NameNode in specified interval then that DataNode is marked as dead and NameNode usage different DataNode. Figure \ref{f:hdfs-read} and \ref{f:hdfs-write} depicts read and write in HDFS respectively.
63: \begin{figure}[!ht]
64: \centering\includegraphics[width=\columnwidth]{images/hdfsRead.PNG}
65: \caption{HDFS Read \cite[Ch.\ 3, p. 38]{AchariShiva2015HE}}\label{f:hdfs-read}
68: \begin{figure}[!ht]
69: \centering\includegraphics[width=\columnwidth]{images/hdfsWrite.PNG}
70: \caption{HDFS Write \cite[Ch.\ 3, p. 39]{AchariShiva2015HE}}\label{f:hdfs-write}
76: Figure \ref{f:yarn-arch} shows YARN architecture.
77: \begin{figure}[!ht]
78: \centering\includegraphics[width=\columnwidth]{images/yarnArch.PNG}
79: \caption{YARN Architecture \cite{www-apache-yarn}}\label{f:yarn-arch}
97: or generate them by hand while using the provided template in Table\ref{t:mytable}. Not ethat
100: \begin{table}[htb]
103: \label{t:mytabble}

figures 4
tables 1
```

```
includegraphics 4
labels 5
refs 4
floats 5

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
False : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--empty year in www-hadoop
(There was 1 warning)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex
```

```
find ""
--------------------------------------------------------------------------------

passed: True

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Big Data Applications in Historical Studies

Neil Eliason
Indiana Unversity
Anderson, Indiana

ABSTRACT

KEYWORDS

i523, HID 312, Big Data, History

1 INTRODUCTION

1.1 Big Data

Big data attention and success stories. Driven by More's Law.

Big data to date can claim numerous victories in a variety of fields, and promises more. Businesses such as Facebook and Netflix have built corporate empires off of the insights gathered from their big data, and physicists and biologists are learning what makes up the universe and ourselves via big data [1].

Despite all this, the concept itself is rather nebulously defined. A rough description of is data with quantitative factors that require specialized techniques to utilize. The most commonly referenced big data factors are volume (amount of data), variety (number of data source types), and velocity (rate of data collection or input) known as the three vs. While this definition is generally accepted, its application varies based upon the industry or field of study and often changes with developments in information technology [5].

The focus on big data arises partially from the phenomenon of data storage capabilities growing at a faster rate than data processing. This creates a situation where data can be economically stored, but not as economically processed, requiring specialized analytic techniques. As big data progresses through the storage, cleaning, analysis, and interpretation stages of the data life cycle, specialized approaches are required [1].

DIKW Data lifecycle

1.2 History of History

The historian's labor has involved interacting with voluminous and varied data for centuries. Before computers, this process involved searching physical archives for relevant data, and manually copying and organizing it into useful information to be analyzed. Though this method can deliver deep insights, some data sets are too big to be studied in a manual fashion [7].

Around the mid-twentieth century, computers had become sufficiently powerful and usable for historians to begin using them to process larger amounts of information. This facilitated a change towards a more quantitative approach and a focus by some from tracing the rise and fall of political or ideological forces, to developing a more complete understanding of mundane topics, such as the family or economics.

Now as archives become digitized and accessible via the internet, the quantity of data available leads to the appeal to big data analytic methods [4]. The potential of unlocking significant connections and developing big picture historical insights at the scale of the growing digital archives of the world is alluring. This hope has driven the labor of many researchers towards developing more big data informed research methods and has directed funds of many institutions towards investments in data infrastructure. However, many are also concerned that the promises of big data are at best optimistic, and at worst hiding potential pitfalls to the historical process [7].

1.3 Thesis

Big Data Analytics have the potential to provide new insights to the field of historical studies. However, their application will differ due to the nature of historical data, and they will serve as an additional tool for the historian, rather than the only tool.

2 BIG DATA IN HISTORICAL STUDIES

2.1 Data Sources

Source types

Methods to get information from data

Methods different from streaming data It could be argued that the seeds of big data history have long laid dormant in archives and libraries, waiting to be germinated by sufficient computational capabilities to process them. As big data analytics mature, pressure develops to make more data available for analysis by digitizing more archival material. This is evidenced not only by the familiar repositories of e-books, but also by archives containing millions of pages from newspapers [7] centuries of letters [4], and

Sources for big data research consist not only of the content of documents in an archive, but also the bibliographical records.While originally designed to allow individual works to be located in an archive, historians have began to study the bibliographical data themselves, an approach called distant reading. By looking at the data about a document, rather than the document's content, societal or intellectual trends can be identified across large scale factors such as time or geography in a more comprehensive way. This approach has elicited some criticism that collections of bibliographical data are not complete enough to derive such large-scale conclusions. Still, considerable interest exist in targeting these data sets for historical analysis [10].

However, the data from these sources differs from that of other fields which utilize big data analytics. Historical data is not streaming the way that social media or smartphone sensors are. It is data which has already been collected, organized, and often times analyzed for a purpose defined by people from a different time and different needs/constraints from ourselves. This creates data sets which are difficult to compare and often require considerable cleaning and reworking to be used in a larger framework. [4].

2.2 Analytics for Big Historical Data

Analytic Techniques

Due to the natural reliance on documents in historical studies, text analytic techniques are the primary set of big data approach utilized by historians. Text analytics is broad category of related algorithms and statistical techniques, such as artificial intelligence, machine learning, and natural language processing that attempt to extract specific information from the text and identify patterns and relationships within the body of data [7].

Artificial intelligence is "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings"[2]. In the context of historical research, this would include tasks such as extracting relevant content from sources, identifying relationships within the data. A specific type of artificial intelligence is machine learning, which consists of programs which change their actions autonomously in response external input. Their ability to adapt allows them to do decision-making tasks, and thus can search through data sources in a more intelligent way to find relevant data [1]. Natural language processing is another artificial intelligence technique, which aims to create programs that can take human language, and make it machine readable [9]. Historians can use such programs connect archival documents to more complex analytic algorithms.

In order to interpret the results of big data analysis, visualization is critical. This is a challenge, as the large scale of the data makes striking a balance between a sufficiently big picture perspective without losing relevant details difficult. Many approaches attempt to utilize high resolution approaches to avoid losing important information [1]. This process is especially challenging in historical studies, as the data is often incomplete and may have inconsistencies which prevent assuming a uniform set of data. For this reason, historians often use visualizations to identify qualitative, rather than quantitative relationships in the data, to inform further inquiry [4].

2.3 Software Packages and Resources for Big Data History

A variety of software packages have been utilized to assist the process of translating raw data into historical insights, such as such as Tableau, Gephi, R, and ArcGIS. However, a limitation of these tools is their their quantitative focus, which tends to exclude more qualitative approaches [4].

Some software has been developed to provide a more qualitative visualization tool set for researchers.For example Stanford University developed a software package called Palladio, designed to visualize connections in large scale historical data. Their approach focused on visualizations that encouraged exploring data, rather than creating statistical statements about it. Examples of this would be mapping connections between historical actors over geography or creating a visualization of the social network of a particular figure in history. They do not create statistical arguments, rather they give a framework for understanding how the data are connected [6].

Another tool with a qualitative visualization focus is WAHSP. It was developed to extract data from the National Library of the Netherlands' newspaper archive, but has been utilized on a number of other databases as well. It provides a number of useful analyses, such as word frequency cloud visualizations, detecting positive or negative sentiment related to certain terms, and Named Entity Recognition, which can identify people, places, events, etc. and then connect them into a relational or geographical framework. It also provides an interactive histogram where the resolution of the data can be adjusted to quickly move between a big picture and detailed data perspective. A derivative project is BILAND, which is a program that can perform many of the analyses of WAHSP, but applies them across two languages, Dutch and German for comparative cultural studies [7].

Along with these data intensive tools specifically designed for historical studies, there are also resources to help the historian learn some of these methods. For example, The Programming Historian website provides a wide range of tutorials and lessons on how to use digital tools in historical studies. At the time of this writing there were 67 lessons available organized by their target stage of research, including lessons on using R, Python, Java, and GitHub [8].

2.4 Insights from Big Historical Data

A number of studies have used these techniques to approach historical research from a big data perspective. Stanford's Mapping of the Republic of Letters project sought to map the social network of Enlightenment thinkers who actively corresponded with each other. This was accomplished by utilizing big data analytics on the metadata of these letters to see how these thinkers related temporally, geographically, socially. Through the research process, the need for more qualitative approaches to visualization was recognized, and eventually led to the development of the Palladio tool set.

Their analysis revealed a number of interesting points. By mapping the social network of John Locke, they supported previous scholarly contentions that the Enlightenment culture was not homogeneously connected, but was made up of a number of subcultures which had thin social connections. Also, by analyzing Benjamin Franklin's letters, they noted that despite his reputation as cross cultural traveler, the main hub of his correspondence was between Philadelphia and London, which were major centers of British culture. [4] .

The WAHSP tool has also been used to analyze large data-sets. One researcher used the tool to study attitudes found towards drugs in the early 20th century newspapers. It found by using the word cloud analysis tool, that before 1924 that drugs such as heroin and opium were used in the context of health, but after 1924 they were associated with crime.

The related tool BILAND was used to study

3 POTENTIAL ISSUES

Opportunistic research: Data driven, not question driven [3], [4] Over-hyped Gaps in data sources Improperly formated data[4]

4 CONCLUSION

ACKNOWLEDGMENTS

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

REFERENCES

[1] C.L. Philip Chen and Chun-Yang Zhang. 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences* 275, Supplement C (2014), 314 – 347. https://doi.org/10.1016/j.ins.2014.01.015

[2] B.J. Copeland. 2017. artificial intelligence (AI). Webpage. (01 2017). https://www.britannica.com/technology/artificial-intelligence

[3] Malte C. Ebach, Michaelis S. Michael, Wendy S. Shaw, James Goff, Daniel J. Murphy, and Slade Matthews. 2016. Big data and the historical sciences: A critique. *Geoforum* 71, Supplement C (2016), 1 – 4. https://doi.org/10.1016/j.geoforum.2016.02.020

[4] Dan Edelstein, Paula Findlen, Giovanna Ceserani, Caroline Winterer, and Nicole Coleman. 2017. Historical Research in a Digital Age: Reflections from the Mapping the Republic of Letters ProjectHistorical Research in a Digital Age. *The American Historical Review* 2 (2017), 400. http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsoaf&AN=edsoaf.a29ec0ac934f1257030b477fa5986b1cff6def96&site=eds-live&scope=site

[5] Amir Gandomi and Murtaza Haider. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* 35, 2 (2015), 137 – 144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

[6] Stanford Humanities and Design. 2017. Palladio. Visualize complex historical data with ease. webpage. (2017). http://hdlab.stanford.edu/palladio/about/

[7] Eijnatten Joris van, Pieters Toine, and Verheul Jaap. 2013. Big Data for Global History: The Transformative Promise of Digital Humanities. *BMGN: Low Countries Historical Review, Vol 128, Iss 4, Pp 55-77 (2013)* 4 (2013), 55. http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsdoj&AN=edsdoj.6259f58bab47404485225cd4776fcf48&site=eds-live&scope=site

[8] Editorial Board of the Programming Historian. 2017. About the Programming Historian. Website. (10 2017). https://programminghistorian.org/about

[9] Technopedia. 2017. Natural Language Processing (NLP). Webpage. (2017). https://www.techopedia.com/definition/653/natural-language-processing-nlp

[10] Sandra1 Tuppen, Stephen2 Rose, and Loukia Drosopoulou. 2016. LIBRARY CATALOGUE RECORDS AS A RESEARCH RESOURCE: INTRODUCING 'A BIG DATA HISTORY OF MUSIC'. *Fontes Artis Musicae* 63, 2 (2016), 67 – 88. http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=llf&AN=114128249&site=eds-live&scope=site

5 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

5.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

5.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

5.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

5.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use *&* but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

`\section{Introduction} and NOT \section{INTRODUCTION}`

5.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

5.6 Character Errors

Erroneous use of quotation marks, i.e. use “quotes” , instead of " "

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

If you see a ﬃgure and not a figure in text you copied from a text that has the ﬁ combined as a single character

5.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

5.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named "images"

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use textwidth as a parameter for includegraphics

Figures should be reasonably sized and often you just need to add columnwidth

e.g.

`/includegraphics[width=\columnwidth]{images/myimage.pdf}`

re

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--unusual to have number, but no volume, for digitalrepublicletters
Warning--unusual to have number, but no volume, for bdglobalhist
(There were 2 warnings)

bibtext _ label error
============================================


bibtext space label error
============================================


bibtext comma label error
============================================


latex report
============================================

        [2017-11-04 09.46.05] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Typesetting of "report.tex" completed in 1.2s.
../README.yml
  8:81      error    line too long (86 > 80 characters)  (line-length)
  9:81      error    line too long (89 > 80 characters)  (line-length)
  10:81     error    line too long (95 > 80 characters)  (line-length)


================================================================================
Compliance Report
================================================================================
```

```
name: Neil Eliason
hid:  312
paper1: Review on 3 Nov 2017
paper2: 15%
project: not yet started

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

4
wc 312 paper2 4 1767 report.tex
wc 312 paper2 4 2605 report.pdf
wc 312 paper2 4 1058 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------


figures 0
tables 0
includegraphics 0
labels 0
refs 0
floats 0
```

```
True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------



label errors



bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Warning--unusual to have number, but no volume, for digitalrepublicletters
Warning--unusual to have number, but no volume, for bdglobalhist
(There were 2 warnings)

bibtex_empty_fields
--------------------------------------------------------------------------------



entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------
```

```
passed: True

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Advancements in Drone Technology for the U.S. Military

Peter Russell
Indiana University
petrusse@iu.edu

ABSTRACT

Technological breakthroughs in military technology have put the U.S. in a new a chapter of warfare. These advancements have become realizations of what was at one time only deemed possible in science fiction, such as autonomous decision making and weaponization of drones. These innovations provide unique advantages to the leaders of the technology and are too lucrative to ignore. However, in coming years as these technologies push the boundaries, decisions will need to be made in how much control military leaders are willing to give to their new mechanic allies and whether they should be passive, as they have been in the past, or active participants on the battlefield.

KEYWORDS

i523, HID 334, Drone Technology, Big Data, Military Technology

1 INTRODUCTION

Among the many industries being transformed by the Big Data movement, few carry more consequences than the changes being experienced in the U.S. military due to the direct impact on human life. It has been argued that the current changes could affect warfare and diplomatic landscape on the same scale that nuclear weapons did [1].

Traditionally, drones can be categorized as a re-usable, autonomous vehicle either in the air or on the ground. In the air, these are known as "Unmanned Aerial Vehicles", or UAVs, and on the ground, "Unmanned Ground Vehicles", or UGVs. More recently, this has expanded to USVs for "Unmanned Surface Vehicles" and UUVs for "Unmanned Underwater Vehicles." Our focus will remain primarily on the former two types, UAVs and UGVs since these have been the most long-standing types.

For most citizens, UAVs remain the more well known of the two and have garnered more attention recently for their proposed commercial and consumer uses. Unsurprisingly, this has created rapid growth in the industry. In 2017 alone, the $6 billion industry for these uses is expected to grow by 35%, roughly matching its 2016 growth [9]. While these market segments are growing quickly, they remain in their nascent stages and dwarfed by drone spending in the Department of Defense (DoD). To put the difference of size in perspective, for the FY2018 budget, the DoD requested nearly $7 billion *for the year* in drone spending, which is the highest since FY2013 and $3.3 billion than previously estimated four years ago [10]. This annual spending is no anomaly. Currently, the DoD is responsible for nearly 90% of the spending in the UAV market [7].

This continued investment in the drone program demonstrates a clear vote of confidence in the advancement of this technology and its impact on military operations.

2 STRATEGIC ADVANTAGE WITH DRONES

One facet in the complexity of military planning is finding the path that gives the highest probability of success with the lowest possible risk of casualties. In this light, the stakes of the problem that technology is trying to solve with Big Data could not be higher. Leaders are constantly trying to solve a constrained optimization problem and when it comes to this overarching problem of mission success, Big Data can be utilized to help the decision maker solve the smaller sub-problems that comprise it, such as where to set up surveillance, where and when to pursue the enemy, how to carry out reconnaissance and how to disarm hazardous obstacles along the way.

2.1 Surveillance

The RQ-4 Global Hawk is currently America's most expensive surveillance drone, projected to cost nearly $428 million in 2018 [10]. Having flown missions for nearly 15 years now, the total cost of this drone is over $14 billion [6]. This drone provides a great case study in the evolution of military drone capabilities for its long track record, which stands at over 200,000 flight hours [11].

Initially, the RQ-4 provided imagery through its equipped sensors, which were for landscape topography (synthetic aperature radar), navigation (electro-optical sensors) and heat signatures (infrared sensors). Later models have been equipped with features that allow the antennae to move on its own for an improved signal and a radar tracking system that allows its operators to zero in a target from its surroundings (moving target indication).

The advancement of sensors provides leaders with high resolution photos for strategic planning. In its current capability, images can be obtained with up to a 1 foot resolution and targeting precision within a 60 foot radius from its maximum height of 65,000 feet [21]. However, it should be noted, the revolutionary aspect in drones does not necessarily come entirely from its sensors. The U-2, which was first introduced in 1955 and famously downed over the Russian border during the Cold War, can be retrofitted with these sensors. The stark difference modern aerial surveillance has with its predecessors like the U-2 is that the current technology allows the device to be unmanned. As a result, if a situation were to occur like with the U-2 where the plane is downed, there is no pilot to capture and a diplomatic crisis around hostage negotiation is avoided. Along these lines, a similar situation played out in late 2016 when China captured an underwater U.S. drone with no major diplomatic ramifications. Additionally, unmanned drones allow for flight times that would likely push beyond the boundaries of human focus and endurance since the RQ-4 can sustain flights in excess of 30 hours [19].

2.2 Swarms

One of the most exciting applications of drone technology revolves around drone swarms. Spending in this category, broadly defined as "Autonomy, Teaming and Swarms" has doubled in the last four years [10]. With UGV spending stagnant over this period, this program is now receiving twice as much funding, but is still only a small fraction of the largest program, unmanned aircrafts, at 10% of that spending. The rapid growth in this program will undoubtedly continue given the revolutionary nature of these swarms as it relates to warfare.

The public has been aware of this new technology since early 2016, but its development has been underway since at least 2014 [13]. The program, however, did not gain mass attention until a 60 Minutes special aired in early 2017 introducing the Perdix drone and demonstrating a mock swarm mission comprised of 103 Perdix drones acting as a single unit [15].

The Perdix drone looks similar to a toy airplane, weighing only a pound and with a wingspan of 6.5 inches. This simple design reflects the expendability of each drone, which is one of the swarms major advantages. In the swarm, there is no lead drone in the swarm and therefore, no single vulnerability to attack if one of the drones was taken down by the enemy [5]. As a result, each drone is designed to work with other drones of the same type as a single unit to achieve a given mission objective and fill in any gaps if drones are no longer functional. These drone swarms are intended to be able to scan large areas very quickly, provide electronic jamming against the enemy, create a wide communication area for ground troops or confuse enemy radar [16].

In the 60 Minutes demonstration, these Perdix drones were dropped from F-18 jets at the speed of sound, aggregated together and collectively scanned an area, entirely on their own. The innovation in computing and Big Data allows the swarm to exist since no human either individually or as part of a team could make the calculations that these drones are making collectively to achieve their mission.

At the moment, while also being a means of surveillance like the RQ-4, these swarms are not a replacement for these traditional drones, nor does that seem to be the end goal. These Perdix drones have a flight time of only 20 minutes currently and are flown at a relatively low-altitude. This compares with with the RQ-4, which is considered a HALE, or High Altitude Long Endurance drone. Additionally, the RQ-4 requires a team of nearly a dozen while the swarm is given a directive by an operator on its objective and requires no human intervention [8]. Lastly, with a unit cost of $235 million per unit, the RQ-4 holds an economic liability with any enemy attack that the Perdix does not at only $30,000 per unit.

Eventually, these swarm drones are expected to have the capability to be aggregated together by the thousands and carry out overwhelming and confusing attacks on enemies. It has been properly described as the "difference between a wolf pack and just little wolves[8]."

2.3 Disarmament and Detection

Of these two drone segments, UAVs remain by far the larger of the two with spending on UAVs nearly 20x that of UGVs[10]. To date, UGVs have been responsible for aiding ground troops in their mission. While this could come in the form of reconnaissance or in helping carry heavy loads, explosive detection has arguably been the most important impact for their ability to screen areas for improvised explosive devices (IEDs) along paths that ground troops must travel to complete their mission.

In comparison to aerial innovations, UGV development has developed at a slower pace when it comes to full autonomy. This is largely due to the nature of challenges a ground drone faces in navigation versus flying. Namely, how to deal with uneven terrain and unpredictable obstacles [14]. Nonetheless, user operated UGVs have proven to be a tremendous advantage as it relates to disarmament and detection.

To circumvent the endless and unique possible situations a UGV could be faced with, the military been innovative in the way these UGVs are deployed instead to avoid these hurdles. For example, soldiers can now throw a five pound UGV from a height of up to 15 feet to begin a reconnaissance or bomb detection mission [18]. This allows them to be thrown on top of a roof or into openings that humans might not be able to fit. These robots are equipped with video cameras and various sensors to relay information about the landscape back to the operator.

3 RECENT DEVELOPMENTS

In the field of surveillance, one of the newest drones being pursued is the Zephyr 8, a solar powered drone that can fly for 45 *days* continuously. This flight time allows the drone to be launched in the U.S. and reach destinations like Afghanistan on its own, but perhaps even more incredible is the amount of data this drone can produce. Specifically, it flies at a height of nearly 12.5 miles in the sky, far exceeding the height needed to see the curvature of the Earth, but can still take pictures at the precision of 6-inch resolution. This height allows surveillance of 386 square miles and coupled with this resolution, this becomes a large data set very quickly [3]. One of the newest developments in drone technology by the U.S. military does not categorize as a UAV or UGV, but instead as a USV, for Unmanned Surface Vehicle. These are autonomous boats with the most famous example to date being the Sea Hunter, which was introduced in 2016. This massive vessel, with the length of 132 feet and 135 tons, was built to track diesel submarines and detect mines [17]. It is a major innovation for the U.S. military for its range, which is 12,000 miles on a single tank of gas, and its economic savings, which is 2% of what a traditional ship costs to operate daily. [20] [4]. Or, framed differently, the U.S. military can operate 50 of these Sea Hunter ships for the same cost as one traditional ship. This has proven to be a Big Data and computational marvel as the ship operates autonomously through 36 computers running 50 million lines of code [15].

4 FUTURE DEVELOPMENTS

One of the aspirational areas of future drone development for the military is in the field of Micro Air Vehicles (MAVs), which as the name implies, are extremely small UAVs, such as the size of a small bird. Even within that area, there is a growing interest in Nano Air Vehicles, which could be the size of an insect. The future of this technology is for troops to gain intelligence on areas that would be either too dangerous or physically impossible to enter.

One of the more well-known MAVs is the Black Hornet Nano. The drone measures 4 inches in length and is an inch wide with the weight of just a half an ounce, or the weight of 3 pieces of paper. This drone has three cameras and can fly for 20 minutes non-stop. Interestingly, the drone is designed to stream video back to its operators to avoid the risk of footage being compromised if it were stored locally. While this is all extremely impressive, future developments are pushing to make these MAVs even smaller. However, the smaller and lighter these MAVs become, the harder they become for a user to control. The reason being is that the smaller they are, the more sensitive they become to cross winds, the more difficult they are to equip with navigation sensors and the smaller field of vision the camera has. However, the inability to be detected by enemies is a tremendous advantage and to circumvent these piloting issues, work is being done to make them fully autonomous, potentially even as a swarm.

5 INTEGRATION OF DRONE TECHNOLOGIES

One of the beautiful aspects of technological innovation are the synergies created. For the U.S. military, these synergies in the context of the Big Data movement are opening new possibilities with difficult questions that will eventually have to be answered. An example of this is how or if drones should be weaponized, even if their decision making is superior to humans.

Without Big Data this debate could never take place. For example, one of the highest resolution drone surveillance cameras in 2014, ARGUS-IS, disclosed some, but not all, of its features as some parts remained classified. It was equipped with a 1.8 billion megapixel camera that could monitor 10 square miles and store all of this information, which works out to be 6 petabytes of data daily [2].

This information accumulation allows greater monitoring of potential targets. Namely, if a known target is tagged and tracked, pictures can be taken at different angles and stored in a database. This ability to accumulate a massive amount of data improves the accuracy of facial recognition. One demonstration showed how a low-altitude drone could be coupled with a UGV and USV against an enemy. The low-altitude and UGV would work in conjunction with each other to carry out a reconnaissance and scan an area and once a match of the target has been found, communicate this to the USV in a different location to fire the weapon systems on the target [15].

While this chain of events is currently possible, which is to attack an enemy with no human interaction, there is a difficult ethical choice to be made in how, or if, these drones will be integrated with respect to weaponization. Even if computers are able to make better decisions on facial recognition, which recent evidence suggests that they can, there remains a large reluctance to open this potential Pandora's box as a new type of warfare [12].

6 CONCLUSION

Adoption of drone and autonomous technology has become the modern arms race and the U.S. has shown itself willing to push to the forefront of these new technologies. This new arms race is unlike the nuclear arms race in that there is no clear first mover, or innovator, advantage. Instead, in the era of Big Data, as shown in the use example of these technologies, the operator that is best able to use the vast amount of information available to them simultaneously will hold the advantage. The U.S. is making promising steps towards this end and will face new, difficult choices in how to integrate these innovations.

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski and the Associate Instructors for their support and suggestions in exploring this topic.

REFERENCES

[1] Greg Allen and Taniel Chan. 2017. *Artificial Intelligence and National Security*. Technical Report. Harvard Kennedy School - Belfer Center for Science and International Affairs, 79 JFK Street, Cambridge, MA 02138.

[2] Sebastian Anthony. 2013. DARPA shows off 1.8-gigapixel surveillance drone, can spot a terrorist from 20,000 feet. Website. (01 2013). http://www.extremetech.com/extreme/146909-darpa-shows-off-1-8-gigapixel-surveillance-drone-can-spot-a-terrorist-from-20000-feet

[3] Allison Barrie. 2015. 'Star Trek'-style surveillance drone for the US military. Website. (09 2015). http://www.foxnews.com/tech/2016/09/15/star-trek-style-surveillance-drone-for-us-military.html

[4] Richard A. Burgess. 2015. ACTUV Sea Trials Set for Early 2016. Website. (11 2015). http://science.dodlive.mil/2015/11/09/actuv-sea-trials-set-for-early-2016/

[5] Jamie Condliffe. 2017. *A 100-Drone Swarm, Dropped from Jets, Plans Its Own Moves.* resreport. MIT Technology Review.

[6] Deagel. 2017. RQ-4A Global Hawk. Website. (04 2017). http://www.deagel.com/Support-Aircraft/RQ-4A-Global-Hawk_a000556001.aspx

[7] The Economist. 2017. *Taking Flight.* resreport. The Economist: Technology Quarterly.

[8] Emily Feng and Charles Clover. 2017. Drone swarms vs conventional arms: Chinafis military debate. Website. (08 2017). https://www.ft.com/content/302fc14a-66ef-11e7-8526-7b38dcaef614?mhq5j=e7

[9] Gartner. 2017. Gartner Says Almost 3 Million Personal and Commercial Drones Will Be Shipped in 2017. Press Release. (02 2017). https://www.gartner.com/newsroom/id/3602317

[10] Dan Gettinger. 2017. *Drones in the Defense Budget.* resreport. Center for the Study of Drones, Bard College.

[11] Northrop Grumman. 2017. Global Hawk. Website. (2017).

[12] Derrick Harris. 2015. Google: Our new system for recognizing faces is the best one ever. Website. (03 2015). http://fortune.com/2015/03/17/google-facenet-artificial-intelligence/

[13] Dan Lamothe. 2016. Watch Perdix, the secretive Pentagon program dropping tiny drones from jets. Website. (03 2016). https://www.washingtonpost.com/news/checkpoint/wp/2016/03/08/watch-perdix-the-secretive-pentagon-program-dropping-tiny-drones-from-jets/?utm_term=.0a44c6311045

[14] John Markoff. 2013. Military Lags in Push for Robotic Ground Vehicles. Website. (09 2013). http://www.nytimes.com/2013/09/24/science/military-lags-in-push-for-robotic-ground-vehicles.html

[15] 60 Minutes. 2017. New generation of drones set to revolutionize warfare. Television. (01 2017). Correspondent David Martin.

[16] Kyle Mizokami. 2017. The Pentagon's Autonomous Swarming Drones Are the Most Unsettling Thing You'll See Today. Website. (01 2017). http://www.popularmechanics.com/military/aviation/a24675/pentagon-autonomous-swarming-drones/

[17] Kris Osborn. 2017. Navy sub-hunting drone ship goes on offense. Website. (01 2017). https://defensesystems.com/articles/2017/01/11/seahunter.aspx

[18] Caroline Reese. 2017. Endeavor Robotics to Provide U.S. Government with Throwable UGV. Website. (09 2017). http://www.unmannedsystemstechnology.com/2017/09/endeavor-robotics-provide-u-s-government-throwable-ugv/

[19] Tyler Rogoway. 2014. Why The USAF's Massive $10 Billion Global Hawk UAV Is Worth The Money. Website. (09 2014). https://foxtrotalpha.jalopnik.com/why-the-usafs-massive-10-billion-global-hawk-uav-was-w-1629932000

[20] Adam Stone. 2016. ACTUV on track for Navy success story. Website. (12 2016). https://www.c4isrnet.com/unmanned/uas/2016/12/21/actuv-on-track-for-navy-success-story/

[21] Patrick W. Watson. 2017. U.S. Military May Soon Deploy Millions Of Drones, Which Presents A Big Investment Opportunity. Website. (08 2017). https://www.forbes.com/sites/patrickwwatson/2017/08/02/u-s-military-may-soon-deploy-millions-of-drones-which-presents-a-big-investment-opportunity/#4068439334a8

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib

bibtext _ label error
============================================


bibtext space label error
============================================


bibtext comma label error
============================================


latex report
============================================

        [2017-11-04 09.46.19] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
Missing character: ""
Missing character: ""
Typesetting of "report.tex" completed in 0.9s.
../README.yml
  14:81     error    line too long (81 > 80 characters)  (line-length)
  15:81     error    line too long (83 > 80 characters)  (line-length)
  16:81     error    line too long (85 > 80 characters)  (line-length)
  17:81     error    line too long (84 > 80 characters)  (line-length)
  18:81     error    line too long (95 > 80 characters)  (line-length)
  19:81     error    line too long (87 > 80 characters)  (line-length)
  20:81     error    line too long (81 > 80 characters)  (line-length)
  21:81     error    line too long (81 > 80 characters)  (line-length)
  22:81     error    line too long (88 > 80 characters)  (line-length)
  46:81     error    line too long (842 > 80 characters)  (line-length)


================================================================================
Compliance Report
================================================================================
```

```
name: Peter Russell
hid:  334
paper1: Oct 28 17 100%
paper2: 90%
project: in progress

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

3
wc 334 paper2 3 2816 report.tex
wc 334 paper2 3 2906 report.pdf
wc 334 paper2 3 655 report.bib

find "
--------------------------------------------------------------------------------

26: Traditionally, drones can be categorized as a re-usable,
    autonomous vehicle either in the air or on the ground. In the air,
    these are known as ``Unmanned Aerial Vehicles", or UAVs, and on
    the ground, ``Unmanned Ground Vehicles", or UGVs. More recently,
    this has expanded to USVs for ``Unmanned Surface Vehicles" and
    UUVs for ``Unmanned Underwater Vehicles." Our focus will remain
    primarily on the former two types, UAVs and UGVs since these have
    been the most long-standing types.

59: Eventually, these swarm drones are expected to have the capability
    to be aggregated together by the thousands and carry out
    overwhelming and confusing attacks on enemies. It has been
    properly described as the ``difference between a wolf pack and
    just little wolves\cite{ftswarm}."

passed: False

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------
```

```
4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------


figures 0
tables 0
includegraphics 0
labels 0
refs 0
floats 0

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
```

```
Database file #1: report.bib

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

passed: True

ascii
--------------------------------------------------------------------------------


================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Big Health Data from Wearable Electronic Sensors (WES) and the Treatment of Opioid Addiction

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

ABSTRACT

Wearable electronic sensors (WES) generate to collect vital health data in the treatment of opioid addiction.

KEYWORDS

Big Data Applications, Health Analytics, Wearable Sensors, i535, HID335

1 INTRODUCTION

In the increasingly connected digital age, personal electronic devices are generating huge volumes of data with important applications for health analytics. Wearable electronic sensors (i.e., *wearables*) and fitness monitors (e.g, FitBit, iWatch) can record our movements and vital physiological measures such as heart rate, temperature, and blood pressure [7]. Consumers are using wearables to self-monitor stress and hypertension, and wearable sensors can be used to help track recovery following medical procedures such as surgery [2]. The development of personalized health care models are also enabling individuals to self-monitor and manage their own health in partnership with care providers. This paper explores approaches to using personal electronic devices and wearable sensors for the treatment of addiction disorders and the prevention of drug overdose. Past research has shown that *Mobile Health* platforms have been used to address prescription medication abuse in several ways: (a) monitor patient health conditions at any time and remotely, (b) monitor medication consumption, and (c) connect patients with health care providers and treatment services [19]. The following review of the literature shows that wireless digital technologies and smartphone applications are effective at providing health data in real time and can assist patients in recovery to resist physical cravings, prevent relapse, and access treatment support. Mobile applications can play an important role in addressing the opioid epideic by supplementing traditional approaches to addiction treatment and recovery.

1.1 The Opioid Epidemic: Medication Abuse and Addiction

The abuse of prescription opioid medication in the U.S. has become a major health crisis that the Department of Health and Human Services (HHS) has described as an epidemic [20]. Approximately 2 million Americans were dependent on or abused prescription opioids (e.g., oxycodone, hydrocodone) in 2014 [9]. Overdose deaths from prescription opioids has quadrupled since 1999, resulting in more than 180,000 deaths between 1999 to 2015. Figure 1 shows that the dramatic increase in overdose deaths in the U.S. between 2000 and 2016 are from synthetic opioids (other than methadone), natural and semi-synthetic opioids, and heroin [14]. Of the estimated 64,000 drug overdose deaths in 2015, over 20,000 were from fentanyl and other synthetic opioid analogs. Public health agencies are implementing comprehensive efforts to address four major risk areas of prescription opioid abuse, overdoses, and deaths: (i) Increasing knowledge of opioid abuse and improving decisions among medication prescribers, (ii) Reducing inappropriate access to opioids, (iii) Increasing effective overdose treatment, (iv) Providing substance-abuse treatment to persons addicted to opioids. The nature of the opioid epidemic is complex, and to understand how technological interventions can play a role in mitigating the crisis, it is necessary to consider the nature of addiction and approaches to treatment.

[Figure 1 about here.]

1.1.1 *Drug Addiction and Treatment.*

For millions of people struggling with substance abuse and dependency in the U.S., addiction and relapse are chronic health conditions [4]. Drug addiction has many similar characteristics to other chronic medical illnesses; however, there are unique challenges to the treatment of addiction illnesses. For example, drug addicted patients undergo intense detoxification in rehabilitation treatment programs, but then are released back into the same environment associated with their drug use. The lack of continuity in the treatment of addiction disorders, leaves addicts in recovery at high risk for relapse into substance use and abuse. Second, individuals with severe addiction disorders end up at emergency rooms for care following acute intoxication, often after law enforcement interventions. Emergency personal are capable with crisis interventions for drug overdose, but lack resources to evaluate severe addiction disorders or provide follow-up. Furthermore, addicted individuals seeking treatment often relapse at night or on weekends when treatment centers are not open. Various theories of addiction and relapse have been proposed. According to the classical conditioning model, situational cues or events can elicit a motivational state underlying relapse to drug use. A slightly more complex model suggests that addictive behavior can be reinstated after extinction of dependency by exposure to drugs, drug-related cues, or environmental stressors [15]. Understanding that a user's affective response to cues in the environmental can lead to relapse and drug use are key to developing strategies for prevention and treatment.

1.2 Technology-Based Interventions for Addiction Treatment

Technology-based interventions have been used for drug addiction assessment, treatment, prevention and recovery [12]. In terms of assessment, data about individuals substance use can be obtained from mobile cell phone reporting outside of treatment settings.

Web-based approaches to treatment have been implemented online to improve behavioral and psychosocial functioning for addicted individuals in recovery [13]. For example, the Therapeutic Education System (TES) is a self-directed, web-based interactive treatment program consisted of 65 training modules that focuses cognitive-behavioral skills, psychosocial functioning (family/social relations). This online approach helped to increase access to treatment for individuals in rural areas, and included an optional contingency management module. A computer based Training in Cognitive Behavioral Therapy (CBT) program was found to enhance treatment outcomes when provided in conjunction with traditional substance abuse treatment, and helped improve coping skills and decision-making skills [6]. In evaluating the effectiveness of mobile applications for addiction treatment, several questions remain to be answered: First, if mobile applications primarily are regarded primarily as supplements to traditional therapeutic treatment, can their effectiveness be measured independently from the approach used in treatment? Second, over what time period period can the benefits of mobile applications be observed? Research evidence suggests that the benefits of mobile interventions may be limited to 12 or 15 weeks [16]. However, it is unclear whether individuals struggling from addiction would continue to use mobile treatment applications in the long terms, beyond a limited course of treatment.

1.2.1 Mobile-Based Applications. Mobile based applications have been used for monitoring and treatment of substance abuse and addiction disorders for several decades [4]. Early applications included the use of electronic pagers (i.e., beepers) for experience sampling with paper-based assessments that generated data about daily life behavior and experiences [16]. In the 1990s, programmable personal digital assistants (e.g., palm-pilot) enabled collection of data electronically, and subsequent mobile research tools facilitated the collection of information about psychological factors (e.g., daily stressors, emotional states, thoughts) and other variables related to addiction (e.g., craving, contextual cues, actual substance use). Assessments performed several times throughout the day (commonly, every 2 to 4 hours) allowed for analysis of the daily fluctuations of these symptoms and features. Historically, addiction research has faced some unique challenges that the use of mobile technologies may help to overcome. Methodological aspects of traditional research using retrospective, cross-sectional, or longitudinal assessments (over periods of weeks, months, or years) have been problematic for investigating risk factors including behaviors and symptoms (severe physiological cravings, withdrawal, and substance use) that can span a relatively short time. An additional factor is the co-morbidity, or co-occurence, of substance use disorders (SUDs) with other psychological disorders, such as anxiety and mood disorders. For example, the "self-medicationfifi model has commonly been used to explain the association between alcohol abuse is used as an effort by an individual to reduce or cope with a high degree of anxiety (or depression). It has also been challenging for researchers to capture the role of environmental or contextual cues (e.g., people, places, things) associated with substance abuse and addiction, which can trigger relapse for individuals in recovery.

Smartphone Applications. Continued care is an important ingredient for recovery from addiction that involves monitoring, outreach, planning, case management, and social support [11]. Smartphone applications can help individuals in recovery to monitor cravings at critical points in daily life, track contextual cues associated with substance use, and provide outreach to support services. A team of researchers at the University of Wisconsin evaluated the effectiveness of a smartphone application called Addiction Comprehensive Health Enhancement Support System (A-CHESS), designed to provide recovery support patients leaving residential alcohol treatment center [10]. A-CHESS provided anytime, anywhere access to support services in audio-visual format, GPS monitoring and warnings for risky locations, and communication with counselors. Over an 8-month period and 4 month follow-up, patients who used the A-CHESS intervention reported fewer risky drinking days, on average, per month than patients in a comparable control group. The findings provide evidence that the smartphone intervention was effective at treating a critical behavioral measure for treatment of alcohol use disorder (AUD). The methods described in this study could be extended by re-purposing built-in smartphone sensors to record physiological measures related to opioid usage, and communicate data to health care providers or treatment specialists to initiate interventions for opioid addiction [11].

[Figure 2 about here.]

1.3 Medication Adherence and Abuse Monitoring System

Mobile health applications can be used to monitor medication adherence and as an advanced warning system for potential abuse of prescription medication [18]. Medication abuse can consist of higher medication dosages or rapid escalation of a prescribed dosage, and the general goal of a prediction model is to analyze patient data for sudden changes in medication consumption. Figure 2 illustrates several steps in a process and decision support structure for a medication monitoring system, with adjustable parameters, such as the threshold for abuse (e.g., greater than N doses in X hours) [19]. A major challenge for measuring medication abuse is obtaining reliable information from potentially addicted individuals based on self report data. Ideally, information on medication consumption and adherence can be obtained from multiple sources. Addiction is a complex behavior that involves a variety of factors, including: demographics (e.g., age, gender), past history, comorbidity with other disorders, family support, social influence, employment status, and patient motivation. Figure 3 shows a model architecture of a system for monitoring potential abuse where dose information is provided via a smartphone application, relayed via wireless cellular network to analytic models that measure changes in medication consumption, relays reports to support treatment services for possible interventions, and to a smart medication box that dispenses medication. In order to function successfully a medication abuse monitoring system depends on the collection of reliable information, including data from wearable sensors that can directly measure physiological changes (e.g., heartrate, blood pressure, respiration, temperature) related to changes in medication usage. In the context of prescription opioid abuse, such as medication monitoring system could be very beneficial in anticipating opioid addiction and preventing overdose death.

[Figure 3 about here.]

1.4 Mobile Detection with Wearable Biosensors

Portable biosensors can provide a continuous stream of data on the timing, location, context, and duration of drug use by individuals in treatment. In a small pilot study, researchers used an Affectiva Q sensor to measure electrodermal activity (EDA), skin temperature, and acceleration (8 recordings per second), in a sample of N = 4 patients during the administration of opioid medication in an emergency room setting [5]. Table 1 provides a summary of the participant characteristics. The biosensor was worn on the wrist and is similar in size and dimensions to a wristwatch or fitbit health monitor. The results showed an increase in EDA associated with intravenous opioid injection that was detected by the biosensors. In addition, there was some indication that the physiological response to opioids varied according to individual drug tolerance. The findings provide evidence to support the use of wearable biodensors to detect drug use in real time, though in a relatively controlled environment. An important limitation of the study is the small sample size, which reduces the generalizability of the findings to a broader population. The authors also acknowledged that psychological or physiological stress can produce alterations in EDA, skin temperature, and acceleration, and therefore this could not be ruled out as an alternative explanation for the findings. The results are promising, however, and encourage future efforts to explore the effectiveness of wearable biosensors in the context of environments that individuals in recovery associate with drug use where relapse can occur.

[Table 1 about here.]

1.5 Emerging Sensor Technologies

Wearable wireless sensors have been used to study physiological responses, activity, and social behavior in non-human primates, in the form of a fitted vest, and using a mobile phone with blue tooth protocol to collect data in real time. Figure 4 shows sample ambulatory data from a rhesus macaque recorded from a wearable wireless sensor for 11 hours inside a large group primate cage [8]. Data was recorded on a custom Android software application, which captured measures of EDA, heart rate (HR), temperature, and acceleration. The goals of this study were to measure associations between physiological measures and social behavior in primates; however, this practical application of sensor technology demonstrated a system that was relatively low-cost, highly portable, scalable, and simple to use. Future research could explore the development of a similar system modified for use with humans to collect data on physiological measures from addicted individuals in naturalistic settings.

[Figure 4 about here.]

1.5.1 LoRa Backscatter: Enabling Ubiquitous Connectivity. Emerging technologies, such as longe range (LoRa) backscatter, have the potential to extend the boundaries of wireless connectivity. Existing radio technologies (e.g., WIFI, ZigBee, SigFox, LTE-M) provide reliable long range coverage, but consume energy and would be costly to expand to large scale implementation; however, LoRa backscatter is a smaller, low-cost, low-power alternative with extended range between an RF source and receiver of approximately 475 meters (i.e., yards) [17]. Table 1 shows the sensitivity and supported data rates for different communication technologies and feasibility of different power sources. LoRa backscatter (LoRaB) performs best in terms of sensitivity (-149 dBm), supports bit rates of 18 pbs to 37.5 kbps, provides whole home coverage, and is the only option capable of being powered by button cell, tiny solar cell, or printed battery. LoRa backscatter uses chirp spread activation (CSS) that can synthesize continuous frequency modulated chirps; a limitation is that backscatter is drowned out by noise and the RF source. The LoRa backscatter system was tested in various deployments: across three floors of a 4800 square- foot house, a single floor of 13,000 square foot office building, and on a one-acre farm. Figure 5 shows the layout of the house with the RF source (TX) on the second floor and receiver in the basement (RX); the plot shows the system achieved RSSI values greater -144 dBm, with reliable wireless coverage throughout the house, and rates sufficient for temperature sensors that transmit small packages. The system was also implemented in the form flexible epidermal patch sensor shown in Figure 6, that provided reliable connectivity across a 3,300 square foot atrium with RSSI greater than -132 dBm. Overall, LoRa backscatter provides a compact, energy-efficient, and affordable wireless transmission system that can be extended to scale at reasonable cost. This system could possibly transmission of biometric data from wearable sensors to capture health information from addicted individuals in treatment and recovery.

[Table 2 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

1.5.2 Graphene Electronic Tattoo sensors. Tattoo-like epidermal sensors are the thinnest electric biosensors made and can measure signals from the heart, muscles, and brain and allows for continuous ambulatory monitoring [3]. A team of researchers at the University of Texas at Austin designed the graphene electronic tattoo (GET) as a long term wearable sensor that can be directly laminated on human skin, and can remain function for several days with a liquid bandage cover [1]. GET has successfully been used to measure electrocardiographm (ECK) electromyogram (EMG) electrocephalograph (EEG), as well as skin temperature and skin hydration.

[Figure 7 about here.]

1.6 Can Technological Applications Reduce Opioid Addiction?

2 TABLES

In case you need to create tables, you can do this with online tools (if you do not mind sharing your data) such as google for them). They even allow you to manage tables as CSV.

Note that the caption is before the tabular environment.

3 CONCLUSION

Put here an conclusion. Conlcusions and abstracts must not have any citations in the section.

ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski, the Assistant Instructors, Juliette Zurick and others, and anonymous reviewers who helped to improve this report.

REFERENCES

[1] Shideh Kabiri Ameri, Rebecca Ho, Hongwoo Jang, Li Tao, Youhua Wang, Liu Wang, David M. Schnyer, Deji Akinwande, and Nanshu Lu. 2017. Graphene Electronic Tattoo Sensors. *ACS Nano* 11, 8 (2017), 7634–7641. https://doi.org/10.1021/acsnano.7b02182 arXiv:http://dx.doi.org/10.1021/acsnano.7b02182 PMID: 28719739.

[2] Louis Atallah, Gareth G. Jones, Raza Ali, Julian J. H. Leong, Benny Lo, and Guang-Zhong Yang. 2011. Observing Recovery from Knee-Replacement Surgery by Using Wearable Sensors. In *Proceedings of the 2011 International Conference on Body Sensor Networks (BSN '11)*. IEEE Computer Society, Washington, DC, USA, 29–34. https://doi.org/10.1109/BSN.2011.10

[3] Katherine Bourzac. 2017. Graphene Temporary Tattoo Tracks Vital Signs. online. (Jan. 2017). https://spectrum.ieee.org/nanoclast/semiconductors/nanotechnology/graphene-temporary-tattoo IEEE Spectrum.

[4] E.W. Boyer, D. Smelson, R. Fletcher, D Ziedonis, and Picard R. W. 2010. Wireless Technologies, Ubiquitous Computing and Mobile Health: Application to Drug Abuse Treatment and Compliance with HIV Therapies. *Journal of Medical Toxicology* 6, 2 (2010), 212–216. https://doi.org/doi:10.1007/s13181-010-0080-z

[5] 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.* 11, 1 (Oct. 2015), 73–79. https://doi.org/10.1007/s13181-014-0439-7.

[6] K.M. Carroll, S.A. Ball, S. Martino, and et al. 2008. Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. *Am J Psychiatry* 165, 7 (2008), 881fi?!8. https://doi.org/10.1176/appi.ajp.2008.07111835

[7] Melinda Gomez Michael Schwartz David Metcalf, Sharlin T.J. Milliard. 2016. Wearables and the Internet of Things for Health. *IEEE Pulse* (Oct. 2016). https://pulse.embs.org/september-2016/wearables-internet-of-things-iot-health/

[8] Richard Ribon Fletcher, Ken ichi Amemori, Matthew Goodwin, and Ann M. Graybiel. 2012. Wearable wireless sensor platform for studying autonomic activity and social behavior in non-human primates. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, Annual International Conference of the IEEE (Ed.). IEEE, IEEE, San Diego, CA, USA. https://doi.org/10.1109/EMBC.2012.6346855

[9] Centers for Disease Control and Prevention. 2017. Prescription Opioid Overdose Data. online. (Oct. 2017). https://www.cdc.gov/drugoverdose/data/overdose.html

[10] D.H. Gustafson, F.M. McTavish, M.-Y. Chih, A.K. Atwood, R.G. Johnson, M. Boyle, and M. ... Shah. 2014. A smartphone application to support recovery from alcoholism: A randomized controlled trial. *JAMA psychiatry.* 71, 5 (May 2014), 566–572. https://doi.org/10.1001/jamapsychiatry.2013.4642

[11] K. Johnson, A. Isham, D.V. Shah, and D.H. Gustafson. 2011. Potential Roles for New Communication Technologies in Treatment of Addiction. *Current psychiatry reports.* (2011). https://doi.org/10.1007/s11920-011-0218-y

[12] Lisa A. Marsch. 2012. Leveraging teachnology to enhance addiction treatment and recovery. *Journal of Addictive Diseases* 31, 3 (2012), 313–318. https://doi.org/10.1080/10550887.2012.694606

[13] L. A. Marsch and J. Dallery. 2012. Advances in the Psychosocial Treatment of Addiction: The Role of Technology in the Delivery of Evidence-Based Psychosocial Treatment. *The Psychiatric Clinics of North America* ;35(2):. doi:, 2 (2012), 481–493. https://doi.org/10.1016/j.psc.2012.03.009

[14] National Institute on Drug Abuse (NIDA). 2017. *Overdose Death Rates.* Summary. National Institutes of Health (NIH), Washington D.C. https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates

[15] Yavin Shaham, Uri Shalev, Lin Lu, Harriet de Wit, and Jane Stewart. 2003. The reinstatement model of drug relapse: history, methodology and major findings. *Psychopharmacology* 168, 1 (01 Jul 2003), 3–20. https://doi.org/10.1007/s00213-002-1224-x

[16] J. Swendsen. 2016. Contributions of mobile technologies to addiction research. *Dialogues Clinical Neuroscience* (2016). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4969708/

[17] Vamsi Talla, Mehrdad Hessar, Bryce Kellogg, Ali Najafi, Joshua R. Smith, and Shyamnath Gollakota. 2017. LoRa Backscatter: Enabling The Vision of Ubiquitous Connectivity. *CoRR* abs/1705.05953 (2017). http://arxiv.org/abs/1705.05953

[18] Upkar Varshney. 2013. Smart medication management system and multiple interventions for medication adherence. *Decision Support Systems* 55, 5 (May 2013), 538–551. https://doi.org/10.1016/j.dss.2012.10.011

[19] Upkar Varshney. 2014. Mobile Health: Medication Abuse and Addiction. In *Proceedings of the 4th ACM MobiHoc Workshop on Pervasive Wireless Healthcare (MobileHealth '14)*. ACM, New York, NY, USA, 37–42. https://doi.org/10.1145/2633651.2633656

[20] Nora D. Volkow, Thomas R. Frieden, Pamela S. Hyde, and Stephen S. Cha. 2014. Medication-Assisted Therapies: Tackling the Opioid-Overdose Epidemic. *New England Journal of Medicine* 370, 22 (2014), 2063–2066. https://doi.org/10.1056/NEJMp1402780 arXiv:http://dx.doi.org/10.1056/NEJMp1402780 PMID: 24758595.

When you submit the paper you need to address each of the items in the issues.tex file and verify that you have done them. Please do this only at the end once you have finished writing the paper. To d this cange TODO with DONE. However if you check something on with DONE, but we find you actually have not executed it correctly, you will receive point deductions. Thus it is important to do this correctly and not just 5 minutes before the deadline. It is better to do a late submission than doing the check in haste.

List of Figures

1 Drugs Involved in U.S. Overdose Deaths from 2000 to 2016, National Institute on Drug Addiction (NIDS) [14] 6
2 Process and Decision Support for Abuse Monitoring System [19] 7
3 Architecture for Abuse Monitoring System [19] 8
4 Sample Ambulatory Data from Rhesus Macaque Recorded on Wearable Sensor for 11+ hours Inside Large Primate Cage Facility [8] 9
5 Home Deployment of LoRa backscatter pakcets across 4,800 sq. ft. House Apread Across Three Floors [17] 10
6 LoRa Backscatter Epidermal Patch [17] 11
7 Graphene Electronic Tatoo Biosensor [1] 11

195

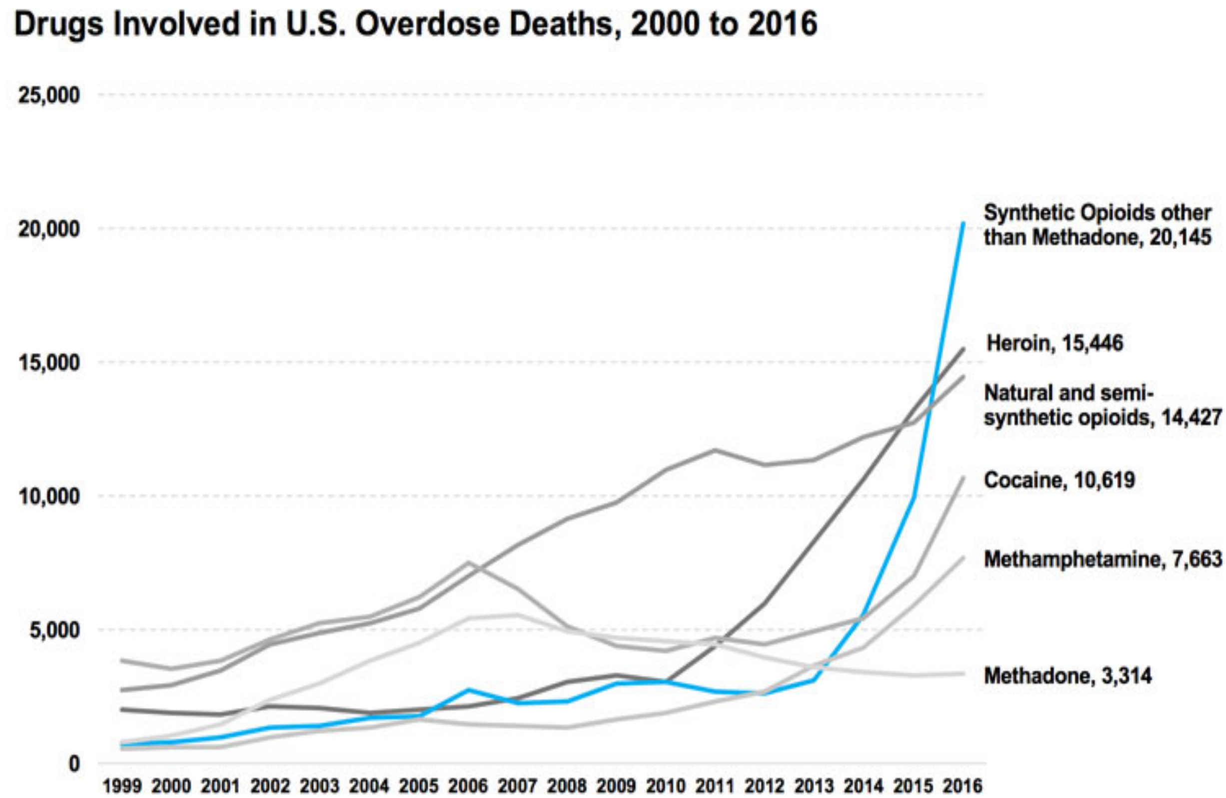
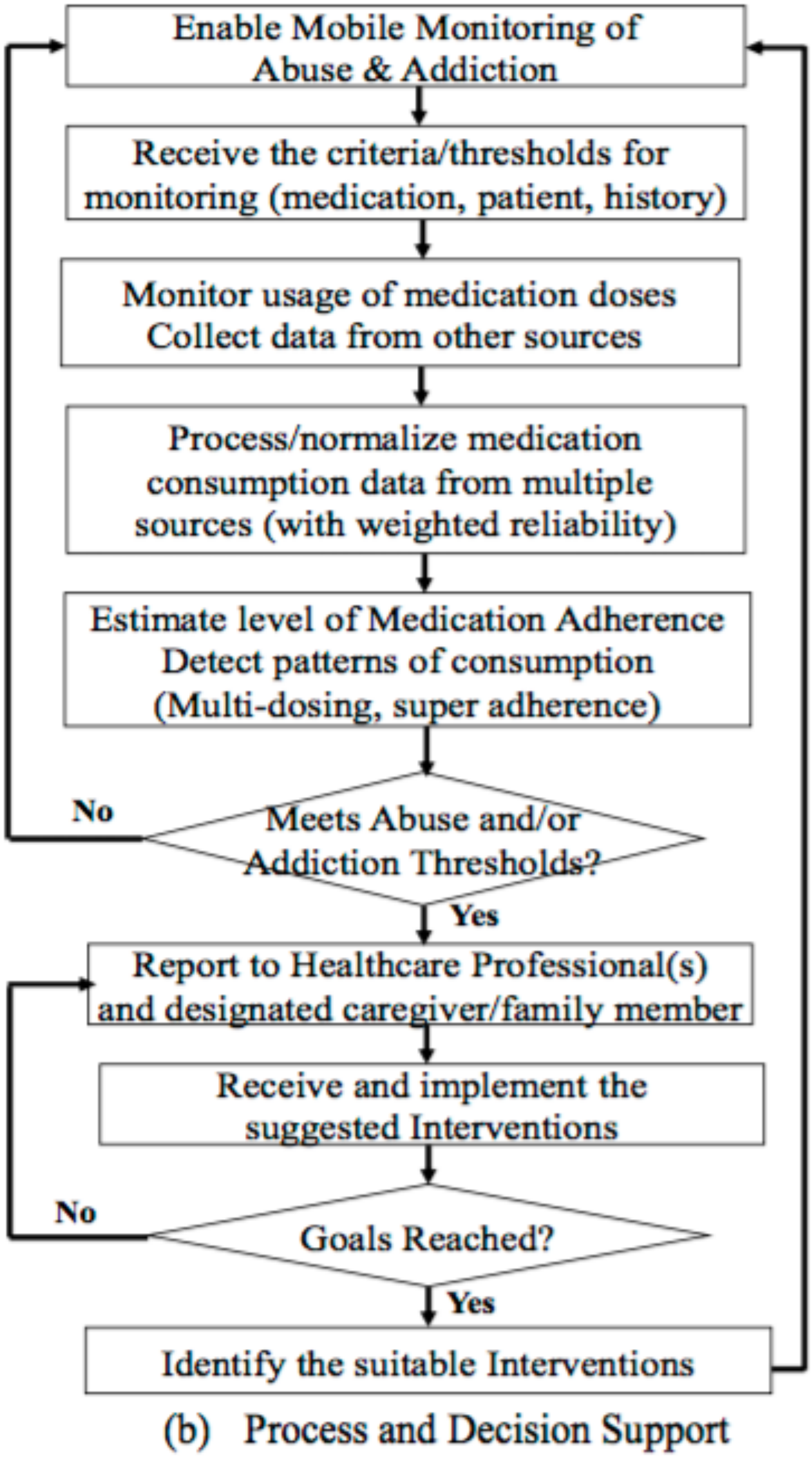


Figure 1: Drugs Involved in U.S. Overdose Deaths from 2000 to 2016, National Institute on Drug Addiction (NIDS) [14]



7

Figure 2: Process and Decision Support for Abuse Monitoring System [19]

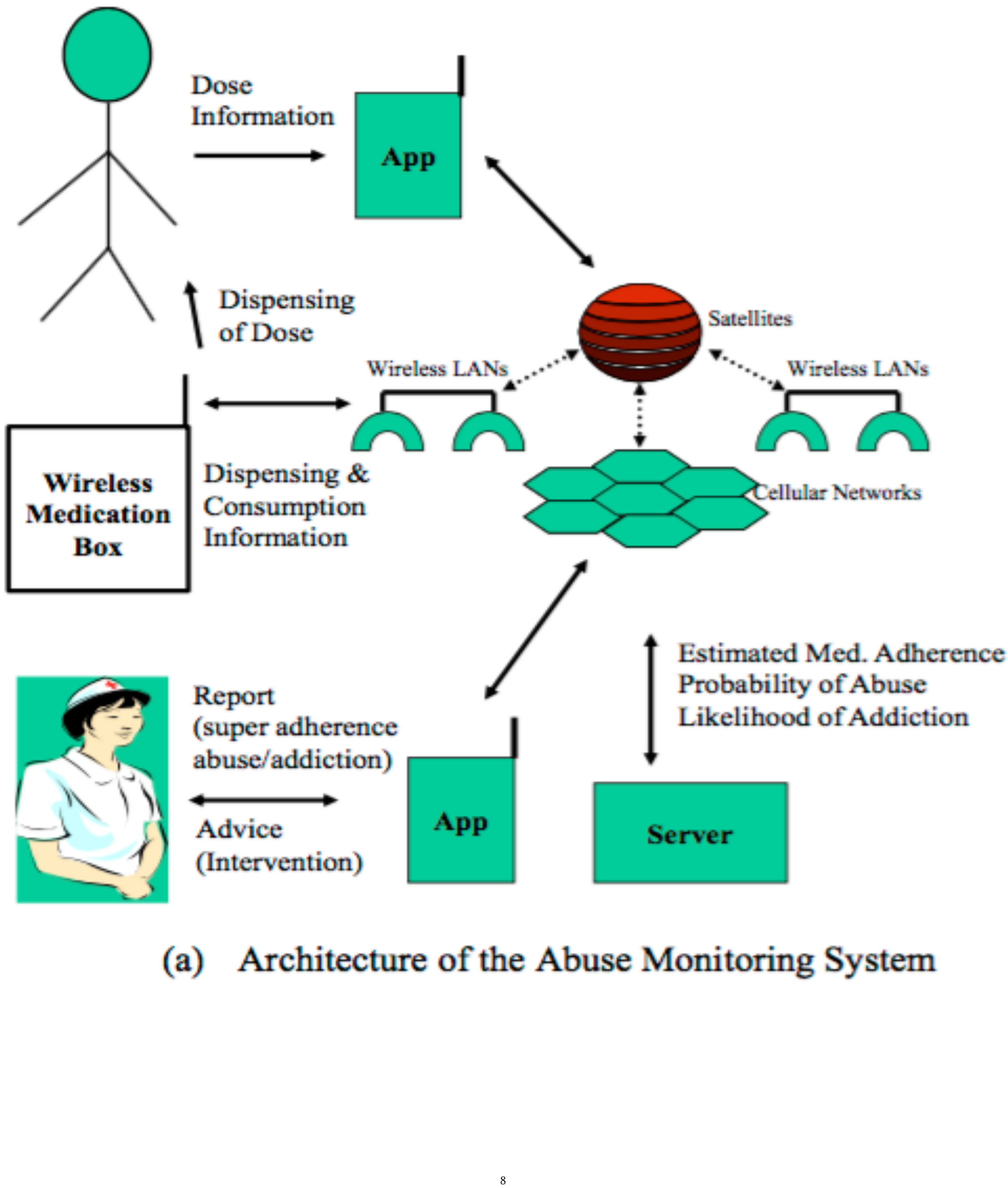


Figure 3: Architecture for Abuse Monitoring System [19]

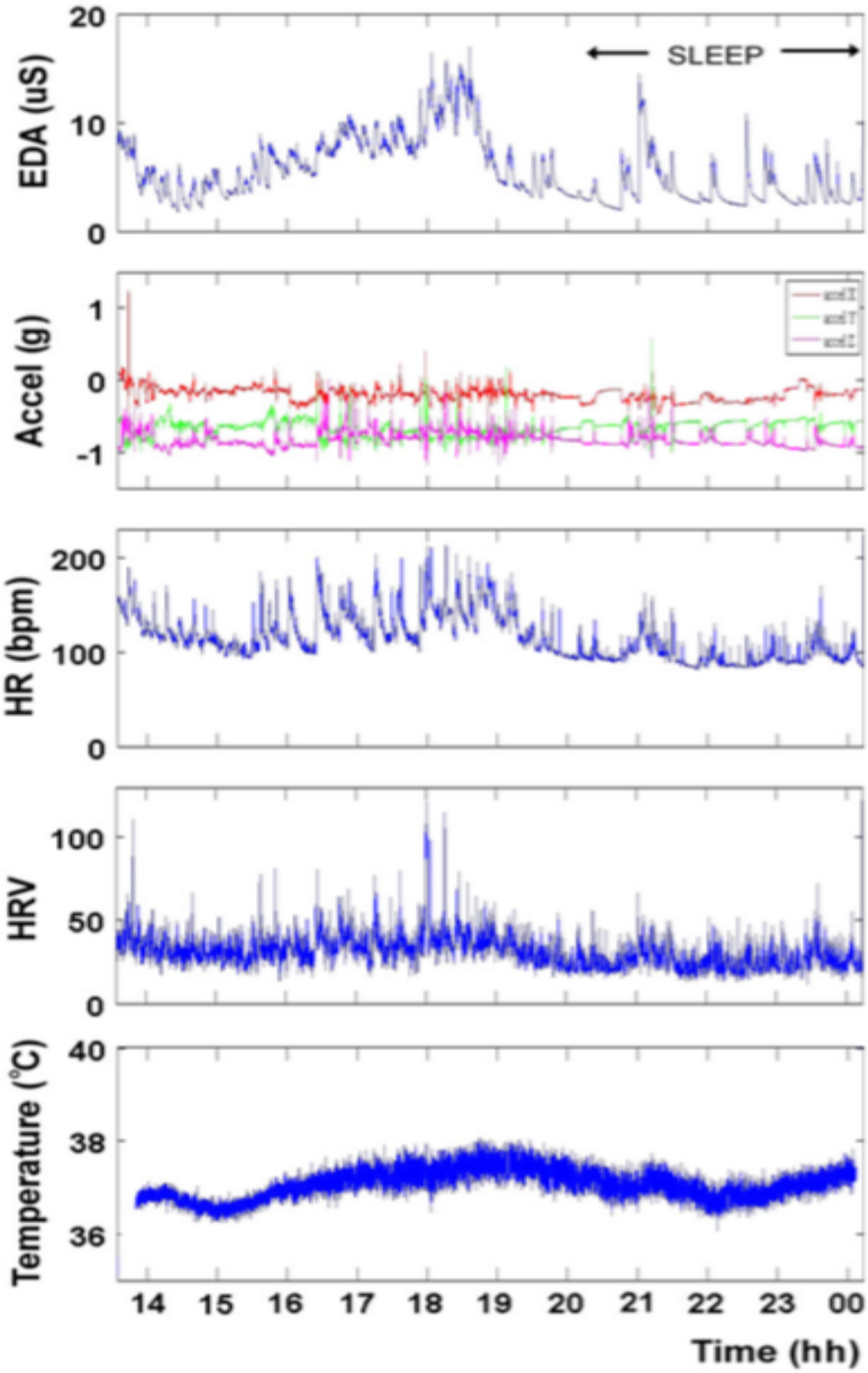
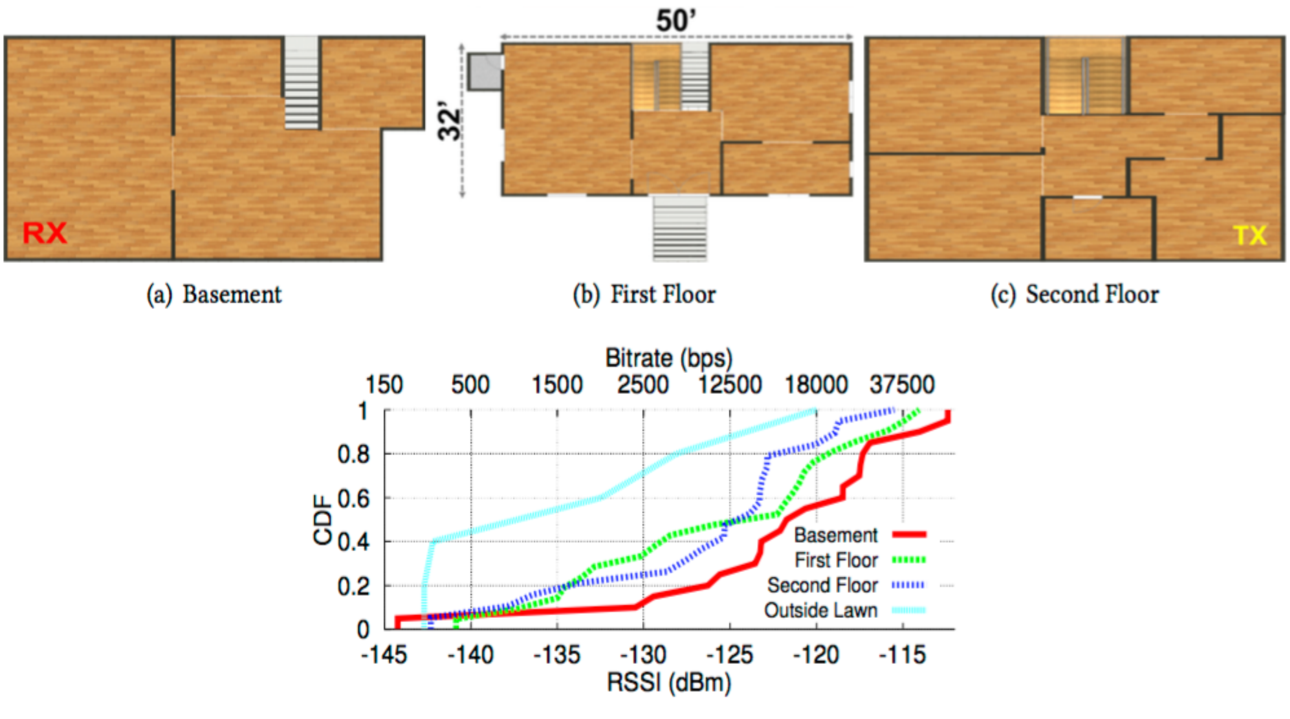


Figure 4: Sample Ambulatory Data from Rhesus Macaque Recorded on Wearable Sensor for 11+ hours Inside Large Primate Cage Facility [8]



(a) Basement (b) First Floor (c) Second Floor

(d) CDF

Figure 5: Home Deployment of LoRa backscatter pakcets across 4,800 sq. ft. House Apread Across Three Floors [17]

10

Figure 6: LoRa Backscatter Epidermal Patch [17]

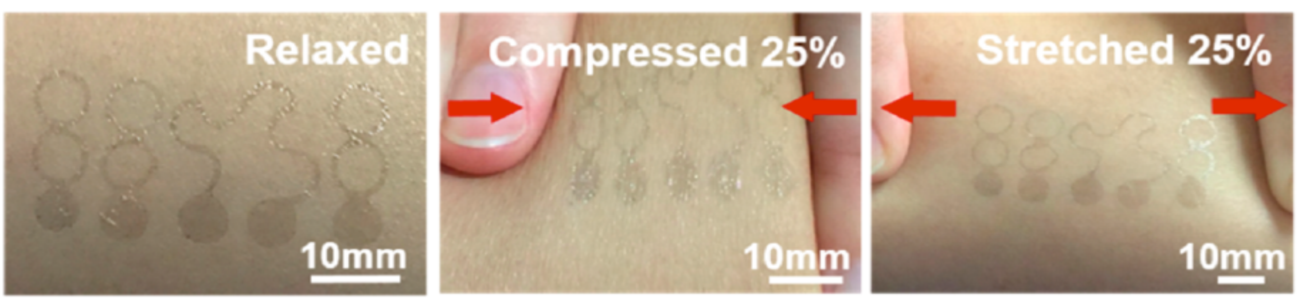


Figure 7: Graphene Electronic Tatoo Biosensor [1]

List of Tables

1 Summary of Participant Characteristics in Pilot study [5] 13
2 Comparison of Wireless Communication Technologies [17] 13

Table 1: Summary of Participant Characteristics in Pilot study [5]

Patient	Age	Gender	History of Use	Intervention	Pre-EDA	Post-EDA
1	82	Male	Opioid naive	4 mg morphine	4.5	60.0
2	47	Male	Recent short-term	1 mg hydromorphone	3.4	12.2
3	43	Female	Chronic opioid use	1 mg hydromorphone	0.2	0.2
4	72	Male	Chronic opioid use	4 mg morphine	0.9	1.6

Table 2: Comparison of Wireless Communication Technologies [17]

Technology	Sensitivity	Data Rate	Home Coverage	Button Cell	Tiny Solar Cell	Printed Battery
Wi-Fi (802.11 b/g)	-95 dBm	1-54 Mbps	yes	no	no	no
LoRa	-149 dBm	18 bps-37.5 kbps	yes	no	no	no
Bluetooth	-97 dBm	1-2 Mbps	no	no	no	no
SigFox	-126 dBm	100 bps	yes	no	no	no
Zigbee	-100 dBm	250 kpbs	yes	no	no	no
Passive Wi-Fi	-95 dBm	1-11 Mbps	no	yes	yes	yes
RFID	-85 dBm	40-640 kbps	no	yes	yes	yes
LoRA Backscatter	-149 dBm	18 bps-37.5 kpbs	yes	yes	yes	yes

```
bibtext report
===========================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3085 of file ACM-Reference-Format.bst
Name 4 in "Boyer, E.W. and Smelson, D. and Fletcher, R. and Ziedonis, D, and Picard R. W
while executing---line 3085 of file ACM-Reference-Format.bst
Name 4 in "Boyer, E.W. and Smelson, D. and Fletcher, R. and Ziedonis, D, and Picard R. W
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3131 of file ACM-Reference-Format.bst
Warning--numpages field, but no articleno or eid field, in atallah11
Name 4 in "Boyer, E.W. and Smelson, D. and Fletcher, R. and Ziedonis, D, and Picard R. W
while executing---line 3229 of file ACM-Reference-Format.bst
Name 4 in "Boyer, E.W. and Smelson, D. and Fletcher, R. and Ziedonis, D, and Picard R. W
while executing---line 3229 of file ACM-Reference-Format.bst
Warning--unrecognized DOI value [doi:10.1007/s13181-010-0080-z]
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3229 of file ACM-Reference-Format.bst
Warning--no number and no volume in metcalf16
Warning--page numbers missing in both pages and numpages fields in metcalf16
Warning--page numbers missing in both pages and numpages fields in fletcher12
Warning--no number and no volume in johnson11
Warning--page numbers missing in both pages and numpages fields in johnson11
Warning--no number and no volume in swedenson16
Warning--page numbers missing in both pages and numpages fields in swedenson16
Warning--page numbers missing in both pages and numpages fields in talla17
Warning--numpages field, but no articleno or eid field, in Varshney14
(There were 12 error messages)
make[2]: *** [bibtex] Error 2
```

latex report
===

```
        [2017-11-04 09.46.24] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
Label ‘tab:freq’ multiply defined.
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
Missing character: ""
bookmark level for unknown  defaults to 0.
The anchor of a bookmark and its parent’s must not be the same. Added a new anchor.
Float too large for page by 51.24545pt.
Float too large for page by 51.24545pt.
Float too large for page by 62.24545pt.
The anchor of a bookmark and its parent’s must not be the same. Added a new anchor.
There were multiply-defined labels.
Typesetting of "report.tex" completed in 1.0s.
../README.yml
  13:5      error    wrong indentation: expected 8 but found 4  (indentation)
  15:5      error    wrong indentation: expected 8 but found 4  (indentation)
  38:5      error    wrong indentation: expected 8 but found 4  (indentation)
  40:5      error    wrong indentation: expected 8 but found 4  (indentation)
  53:4      error    wrong indentation: expected 4 but found 3  (indentation)
  53:17     error    trailing spaces  (trailing-spaces)
  55:4      error    wrong indentation: expected 7 but found 3  (indentation)
  57:4      error    wrong indentation: expected 7 but found 3  (indentation)
  74:70     error    trailing spaces  (trailing-spaces)
  78:33     error    trailing spaces  (trailing-spaces)
  80:81     error    line too long (91 > 80 characters)  (line-length)
```

==
Compliance Report
==

name: Sean Shiverick
hid: 335
paper1: 10/25/17 100%
paper2: 90%
project: in progress

yamlcheck

```
wordcount
--------------------------------------------------------------------------------

13
wc 335 paper2 13 3108 report.tex
wc 335 paper2 13 3631 report.pdf
wc 335 paper2 13 1096 report.bib

find "
--------------------------------------------------------------------------------

passed: True

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------

75: \begin{figure}[!ht]
76: \centering\includegraphics[width=\columnwidth]{images/Figure1.pdf}
79: }\label{f:Figure1}
191: \begin{figure}[!ht]
192: \centering\includegraphics[width=\columnwidth]{images/Figure2.pdf
     }
195: }\label{f:Figure2}
228: \begin{figure}[!ht]
229: \centering\includegraphics[width=\columnwidth]{images/Figure3.pdf
     }
231: }\label{f:Figure3}
260: \begin{table}
262: \label{tab:freq}
293: \begin{figure}[!ht]
294: \centering\includegraphics[width=\columnwidth]{images/Figure4.pdf
     }
```

```
298: \label{f:Figure4}
332: \begin{table}
334: \label{tab:freq}
351: \begin{figure}[!ht]
352: \centering\includegraphics[width=\columnwidth]{images/Figure5.pdf
     }
355: }\label{f:Figure5}
358: \begin{figure}[!ht]
359: \centering\includegraphics[width=\columnwidth]{images/Figure6.pdf
     }
361: }\label{f:Figure6}
370: \begin{figure}[!ht]
371: \centering\includegraphics[width=\columnwidth]{images/Figure7.pdf
     }
373: }\label{f:Figure7}
381: %\begin{figure}[!ht]
382: %  \centering\includegraphics[width=\columnwidth]{images/Figure8.
     pdf}
384: %  }\label{f:Figure8}
398: %Table\ref{t:mytable}.
401: %\begin{table}[htb]
404: %\label{t:mytabble}

figures 8
tables 3
includegraphics 8
labels 11
refs 1
floats 11

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
False : check if all figures are refered to: (refs >= labels)

Label/ref check
60: more than 180,000 deaths between 1999 to 2015. Figure 1 shows that
    the dramatic
205: Figure 2 illustrates several steps in a process and decision
     support structure
214: status, and patient motivation. Figure 3 shows a model
     architecture of a system
241: medication in an emergency room setting \cite{carreiro15}. Table
     1 provides a
280: in real time. Figure 4 shows sample ambulatory data from a rhesus
     macaque
```

```
309: (i.e., yards) \cite{talla17}. Table 1 shows the sensitivity and
     supported data
319: one-acre farm. Figure 5 shows the layout of the house with the RF
     source (TX)
324: sensor shown in Figure 6, that provided reliable connectivity
     across a 3,300
passed: False -> labels or refs used wrong

When using figures use columnwidth
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
Database file #1: report.bib
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3085 of file ACM-Reference-Format.bst
Name 4 in "Boyer, E.W. and Smelson, D. and Fletcher, R. and Ziedonis, D, and Picard R. W
while executing---line 3085 of file ACM-Reference-Format.bst
Name 4 in "Boyer, E.W. and Smelson, D. and Fletcher, R. and Ziedonis, D, and Picard R. W
while executing---line 3085 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3131 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3131 of file ACM-Reference-Format.bst
Warning--numpages field, but no articleno or eid field, in atallah11
```

```
Name 4 in "Boyer, E.W. and Smelson, D. and Fletcher, R. and Ziedonis, D, and Picard R. W
while executing---line 3229 of file ACM-Reference-Format.bst
Name 4 in "Boyer, E.W. and Smelson, D. and Fletcher, R. and Ziedonis, D, and Picard R. W
while executing---line 3229 of file ACM-Reference-Format.bst
Warning--unrecognized DOI value [doi:10.1007/s13181-010-0080-z]
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3229 of file ACM-Reference-Format.bst
Too many commas in name 1 of "David Metcalf, Sharlin T.J. Milliard, Melinda Gomez, Micha
while executing---line 3229 of file ACM-Reference-Format.bst
Warning--no number and no volume in metcalf16
Warning--page numbers missing in both pages and numpages fields in metcalf16
Warning--page numbers missing in both pages and numpages fields in fletcher12
Warning--no number and no volume in johnson11
Warning--page numbers missing in both pages and numpages fields in johnson11
Warning--no number and no volume in swedenson16
Warning--page numbers missing in both pages and numpages fields in swedenson16
Warning--page numbers missing in both pages and numpages fields in talla17
Warning--numpages field, but no articleno or eid field, in Varshney14
(There were 12 error messages)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

passed: True

ascii
--------------------------------------------------------------------------------

non ascii found 8217
non ascii found 8217

================================================================================
The following tests are optional
================================================================================

Tip: newlines can often be replaced just by an empty line
```

```
find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```

Natural Language Processing (NLP) to analyze human speech data

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

Extracting meaningful information from large volumes of unstructured human language is a challenging big data problem. Automatic speech recognition (ASR) and natural language processing (NLP) based intelligent system can be used in several human machine interface applications both in consumer and industrial sector. Here describing the architecture, building blocks, performance and applications for such system that would use pre-developed ASR and NLP APIs.

KEYWORDS

i523, HID333, HID337, Natural Language Processing

1 INTRODUCTION

As voice becoming a common user interface, the need for accurate and intelligent speech recognition technologies is growing. In speech processing technology there are two main subtasks

- Speaker Recognition
- Speech Recognition

Although the performance of current speaker and speech recognition systems is far from perfect, these systems have already proven their usefulness in many applications.

2 SPEAKER RECOGNITION

Speaker identification is one of the important task in speech processing. Each person has a voice that is different from everyone else's.Speaker recognition is the process of identifying who is speaking by using acoustic features of speech.Speaker recognition has been applied mostly in security applications to control access. Current speaker recognition systems are not very accurate for large speaker populations.

3 NLP FOR SPEECH RECOGNITION

Speech recognition is the ability to identify spoken words. It is the process of converting speech into text. This process prepares the input data (speech) to be appropriate for Natural Language Processing(NLP). NLP is the processing of the text to understand the meaning of the text. It comes as the next step of speech recognition. *Machine learning* algorithms are used in conjunction with language models to recognize text in natural language processing systems, which may also employ speech models and hardware/software specialized to process and recognize speech.

Analyzing language for its meaning is a complex task. Modern speech recognition research began in the late 1950s with the advent of the digital computer. The 1960s saw advances in the automatic segmentation of speech into units of linguistic relevance like phonemes, syllables. And now with advancements in the field of Artificial Intelligence, neural networks have been used in many aspects of speech recognition such as phoneme classification,isolated word recognition,audiovisual speech recognition, audiovisual speaker recognition and speaker adaptation. In the context of Speech Recognition, NLP involves 4 basic steps

- **Morphological Analysis**: Morphological analysis is the identification, analysis, and description of the structure of a given languagefis root words,word boundaries, affixes, parts of speech, etc. The term Morpheme means the "minimal unit of meaning". For ex: if you take word "unhappiness" it has three morphemes each carrying its own meaning.
- **Syntactic Analysis**: Syntactic analysis is the process of analyzing a string of symbols in natural language conforming to the rules of a formal grammar.
- **Semantic Analysis**: Semantic analysis is the process of relating syntactic structures, from the levels of phrases, clauses, sentences and paragraphs to the level of the writing as a whole, to their language-independent meanings.
- **Pragmatic Analysis**: Pragmatic Analysis is how sentences are used in different situations and how use affects the interpretation of the sentence. Means what was said is reinterpreted as what it actually means.

NLP techniques are broadly categorized into

- **Rule based (knowledge driven)**: Rule based approach requires huge human effort to prepare the rules, parts of speech triggers etc. The best known parser with a rule base backbone is the RASP (Robust Accurate Statistical Parsing) system that combines rule-based grammar with a probabilistic parse selection model [12, 13].
- **Statistical based (data driven)**: Statistical/Data driven approaches treats natural language processing as a *machine learning* problem. They use supervised or unsupervised statistical machine learning algorithms. This method applies learning algorithm to a large body of previously translated text(large data) known as a parallel corpus.

  The main advantage of the statistical approach is its language Independence.Provided there are annotated data, the same algorithm can be used for learning rules or models for any language. The statistical approach is significantly leading in terms of accuracy against manually annotated corpora, as well as in overall number of statistical parsers compared to the number of rule-based parsers. Fast, cheap computing hardware, advances in processor speed, random access memory size, secondary storage, and grid computing making Statistical approach as popular choice. One example parser with his approach is MaltParser [128], a data-driven parser-generator for dependency parsing that supports several parsing algorithms and learning algorithms and allows user-defined feature models, consisting of arbitrary combinations of lexical features, part-of-speech features and dependency features.

4 CONCLUSION

ACKNOWLEDGMENTS

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

REFERENCES

```
bibtext report
============================================

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
I found no \citation commands---while reading file report.aux
Database file #1: report.bib
(There was 1 error message)
make[2]: *** [bibtex] Error 2

latex report
============================================

        [2017-11-04 09.46.35] pdflatex report.tex
This is pdfTeX, Version 3.14159265-2.6-1.40.17 (TeX Live 2016) (preloaded format=pdflate
Missing character: ""
Missing character: ""
Empty ‘thebibliography’ environment.
Typesetting of "report.tex" completed in 0.8s.
../README.yml
  16:16     error    trailing spaces  (trailing-spaces)


================================================================================
Compliance Report
================================================================================


name: Ashok Reddy Singam
hid:  337
paper1: Nov 01 17 100%
paper2: In Progress
project: not started

yamlcheck
--------------------------------------------------------------------------------


wordcount
--------------------------------------------------------------------------------

2
wc 337 paper2 2 813 report.tex
wc 337 paper2 2 765 report.pdf
```

```
wc 337 paper2 2 66 report.bib

find "
--------------------------------------------------------------------------------

57: Morphological analysis is the identification, analysis, and
    description of the structure of a given languages root words,word
    boundaries, affixes, parts of speech, etc. The term Morpheme means
    the "minimal unit of meaning". For ex: if you take word
    "unhappiness" it has three morphemes each carrying its own
    meaning.

passed: False

find footnote
--------------------------------------------------------------------------------

passed: True

find input{format/i523}
--------------------------------------------------------------------------------

4: \input{format/i523}

passed: True

floats
--------------------------------------------------------------------------------


figures 0
tables 0
includegraphics 0
labels 0
refs 0
floats 0

True : ref check passed: (refs >= figures + tables)
True : label check passed: (refs >= figures + tables)
True : include graphics passed: (figures >= includegraphics)
True : check if all figures are refered to: (refs >= labels)

Label/ref check
passed: True

When using figures use columnwidth
```

```
[width=1.0\columnwidth]
do not cahnge the number to a smaller fraction


find textwidth
--------------------------------------------------------------------------------

passed: True

bibtex
--------------------------------------------------------------------------------


label errors


bibtex errors

This is BibTeX, Version 0.99d (TeX Live 2016)
The top-level auxiliary file: report.aux
The style file: ACM-Reference-Format.bst
I found no \citation commands---while reading file report.aux
Database file #1: report.bib
(There was 1 error message)

bibtex_empty_fields
--------------------------------------------------------------------------------


entries in general should not be empty in bibtex


find ""
--------------------------------------------------------------------------------

passed: True

ascii
--------------------------------------------------------------------------------

non ascii found 8217

================================================================================
The following tests are optional
================================================================================
```

```
Tip: newlines can often be replaced just by an empty line

find newline
--------------------------------------------------------------------------------

passed: True
cites should have a space before \cite{} but not before the {

find cite {
--------------------------------------------------------------------------------

passed: True
```