Insert Your Title Here∗

Insert Subtitle Here

Siqi Liu  
 Computer Science  
 University of Waterloo  
 Waterloo Ontario Canada  
 sq2liu@uwaterloo.ca

Zibai Wang  
 Computer Science  
 University of Waterloo  
Waterloo Ontario Canada  
 zibai.wang@uwaterloo.ca

ABSTRACT

Our project is a Flink application written in java that analyzes Twitter Streaming data including HashTag count, Tweets Geographic data and Retweets percentages based on the given search terms, hashtags and geodata as input, transfer the data into ElasticSearch and further visualize using Kibana.

KEYWORDS

Flink, Twitter, Streaming, ElasticSearch, Kibana

Methodology

The updated template, user manuals, samples, and required fonts, all are available at the URL <https://www.acm.org/publications/proceedings-template>. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

 (1)

**Continuation part of Paragraph Text** The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**



The **DisplayFormulaUnnum** style is applied only in case of an unnumbered equation. An unnumbered display equation never contains an equation number to its right, and this unique property distinguishes it from a numbered equation.



Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

**Theorem/Proof/Lemma.** Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

Evaluation

In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

1. In a Word 2010 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. Select the **Alt Txt** option from the left-side panel options.
4. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

Below are steps to place alt-txt value in **MS Word 2013/2016**. To add alternative text to a picture in Word 2013/2016, follow these steps:

1. In a Word 2013/2016 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. In the settings at the right side of the window, click on the "Layout & Properties" icon (3rd option).
4. Expand **Alt Txt** option.
5. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

*1.1.1 Heading Level 3.* Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

*1.1.1.1 Heading Level 4.*Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

**Lesson Learnt:**

1. Always assume there will be missing data.  
Some attributes in the Tweet JSON are Nullable and our code needed to handle it. Lack of handling Nullable data has caused some trouble for us

2. Always use the correct version, and read the correct versions’ documentation. Initially, we installed the latest ElasticSearch version (7.4.2) although we did see the Flink Elasticsearch Connector was only tested on ES version 6. This mistake resulted in a few issues, for example, the Mapping process changed dramatically from v6 to v7 which caused Flink Elasticsearch Connector no longer working with ES 7. After we changed our ES version to v6, things started to run smoothly.

3. End-user - do we care about #Trump vs #trump? Need to use lower cases. User inputs should be preprocessed in order to produce a more accurate result for End-user.

4. Trade-off between WindowSize and size of the index in ES

5. Before diving into the task, understand if the result will be meaningful

* 1. WordCount - doesn’t really mean much since a lot are filler words. Unless we can filter out, the result doesn’t matter
  2. FavouriteCount - doesn’t make much sense either since normally the tweet will not be liked by someone else in such a short window.

6. Irregularities in data (emojis, pictures, user mentions)

**Further Improvements:**

1. Dockerize:

Since all the frameworks(Flink, ElasticSearch, Kibana) that we used in our project can be dockerized, we can build our project into a docker app and publish it on the docker hub. As Docker can be used in a wide variety of Operating Systems, any user who is interested in using our project can pull it from the docker hub and get it up and running in minutes, without the need to install each component separated.

2. Set-up parallel tasks to handle each task separately

3. See if we can flush indices periodically to save space in ES

REFERENCES

[1] <https://learning.oreilly.com/library/view/stream-processing-with/9781491974285/>

[2] <https://medium.com/@chandanbaranwal/spark-streaming-vs-flink-vs-storm-vs-kafka-streams-vs-samza-choose-your-stream-processing-91ea3f04675b>

[3] <https://ci.apache.org/projects/flink/flink-docs-stable/dev/connectors/twitter.html>

[4] <https://ci.apache.org/projects/flink/flink-docs-stable/dev/connectors/elasticsearch.html>

[5] <https://www.tutorialspoint.com/apache_flink/apache_flink_batch_realtime_processing.htm>

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00