Real-time Mining of Tweet and Police-Report Streams for Forensic Applications

Laksh Gupta, Farnoush Banaei-Kashani, Cyrus Shahabi
{lgupta, *banaeika, shahabi} @usc.edu*

iCampus/MediaQ iWatch ✓ CT

Introduction

•The research focuses on the real-time mining of the events from tweets and police records.
•Provides better assistance in solving crime as a part of an intelligent surveillance system.
•Uses data from Twitter and Department of Public Safety (DPS).
•Extracts entities from unstructured text using rule-based approach.

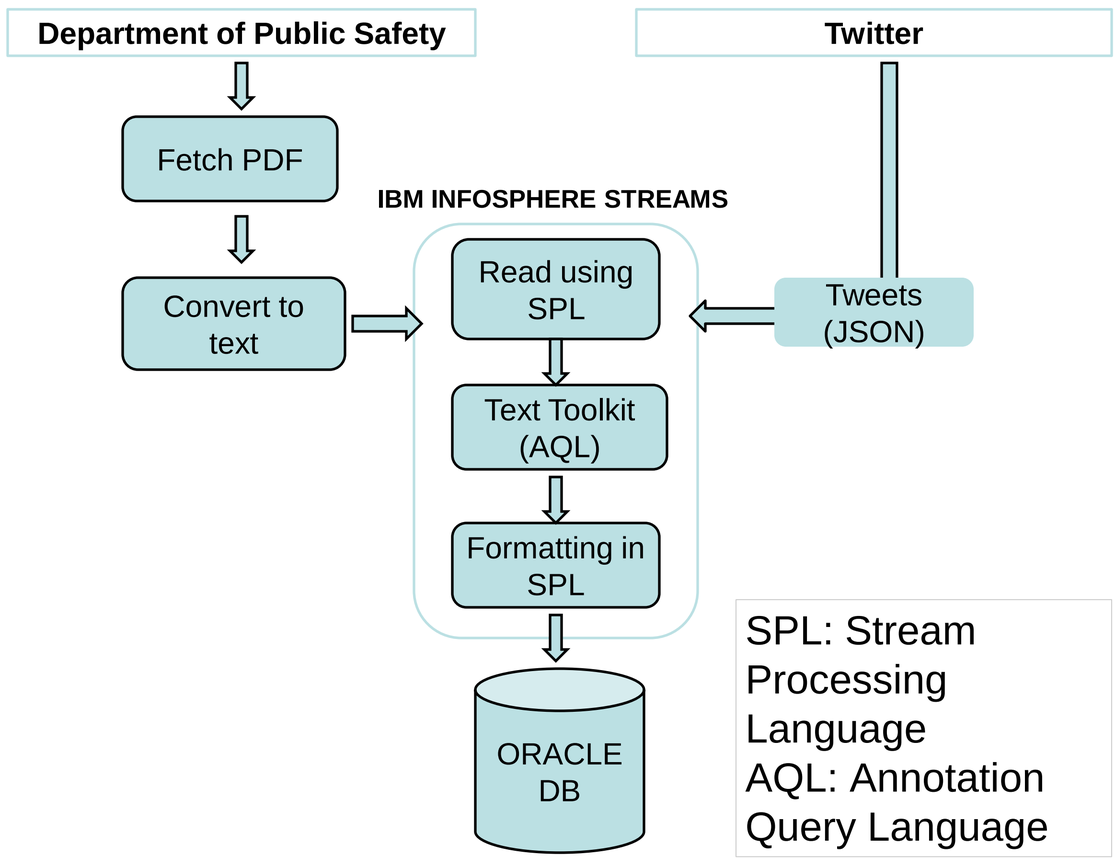
Motivation

•Data for analysis can come from various sources such as video and sensors as well as from Structured/Unstructured text.
•Police records provide more specific and detailed information for identifying crime patterns.
- type of event
- person (culprit/victim)
- time of event, etc.

•Twitter has proved to be one of the fastest medium of receiving and exchanging information. Many organizations are involved with Twitter data analysis. In our case it may provide information:
- person (culprit/victim)
- location
- description, etc

•Combining information from these sources and performing real-time analysis could prove valuable in providing better information in-hand for forensic investigation.

System Architecture



SPL: Stream Processing Language
AQL: Annotation Query Language

Capability

Real Time Processing of Text

IBM InfoSphere Streams enables us to perform real-time analysis using Text Toolkit.

1. **Police Records from DPS**

Reported: 2/11/13 - 9:23 am **Location:** RONALD TUTOR CAMPUS CENTER **Report #:** 1300606

Occurred: 2/8/13 - 9:00 am to 2/11/13 - 9:00 am **Disposition:** Open

Incident: PROPERTY Missing Property

Summary: A student reported her bicycle missing.

A student reported her bicycle missing.

- Victim: group 2 in
(<D.determiner>)(<N.noun>{1,3})(<V.verb>)
- Object: group 2 in
(<P.possesive>)(<N.noun>{1,3})(<V.verb>)

2. Tweets

{"created_at": "Wed Feb 06 09:12:08 +0000 2013", "id": 299083039781290000, "id_str":"299083039781289984", "text": "Someone is stalking me", "source": "<a href …….. "user": {"id": 88206644, "id_str": "88206644", "name": "Ebutecastle ", "screen_name":"ebutecastle“ ……

Someone is stalking me.

- Culprit: group 1 in
(<P1.pronoun>)(‘is stalking’)(<P2.pronoun>)
- Victim: group 2 in
('stalked'|'stalking’) (<P.pronoun>)

Related Research

•Laura Chiticariu, Rajasekar Krishnamurthy, Yunyao Li, Sriram Raghavan, Frederick R Reiss, Shivakumar Vaithyanathan: SystemT: an algebraic approach to declarative information extraction. ACL 2010.

•Rui Li, Kin Hou Lei, Ravi Khadiwala, Kevin Chen-Chuan Chang: TEDAS: A Twitter-based Event Detection and Analysis System. ICDE 2012

Conclusion and Future Work

•Rule based approach works well with the standard English language as used in police records, F1 score around 79%

•Improve Twitter analysis
•Extract more attributes.