

Crime Search Engine & Heat Map

Anish, Sunkara
YaoChing, Peng

Agenda

- i Introduction
- i Project model
- i Nutch Search Engine
 - § Plug-ins
 - § Interface
- i Source
 - § Precise Data
- i Text Operation
- i User Interface
 - § Heat map
- i Conclusion

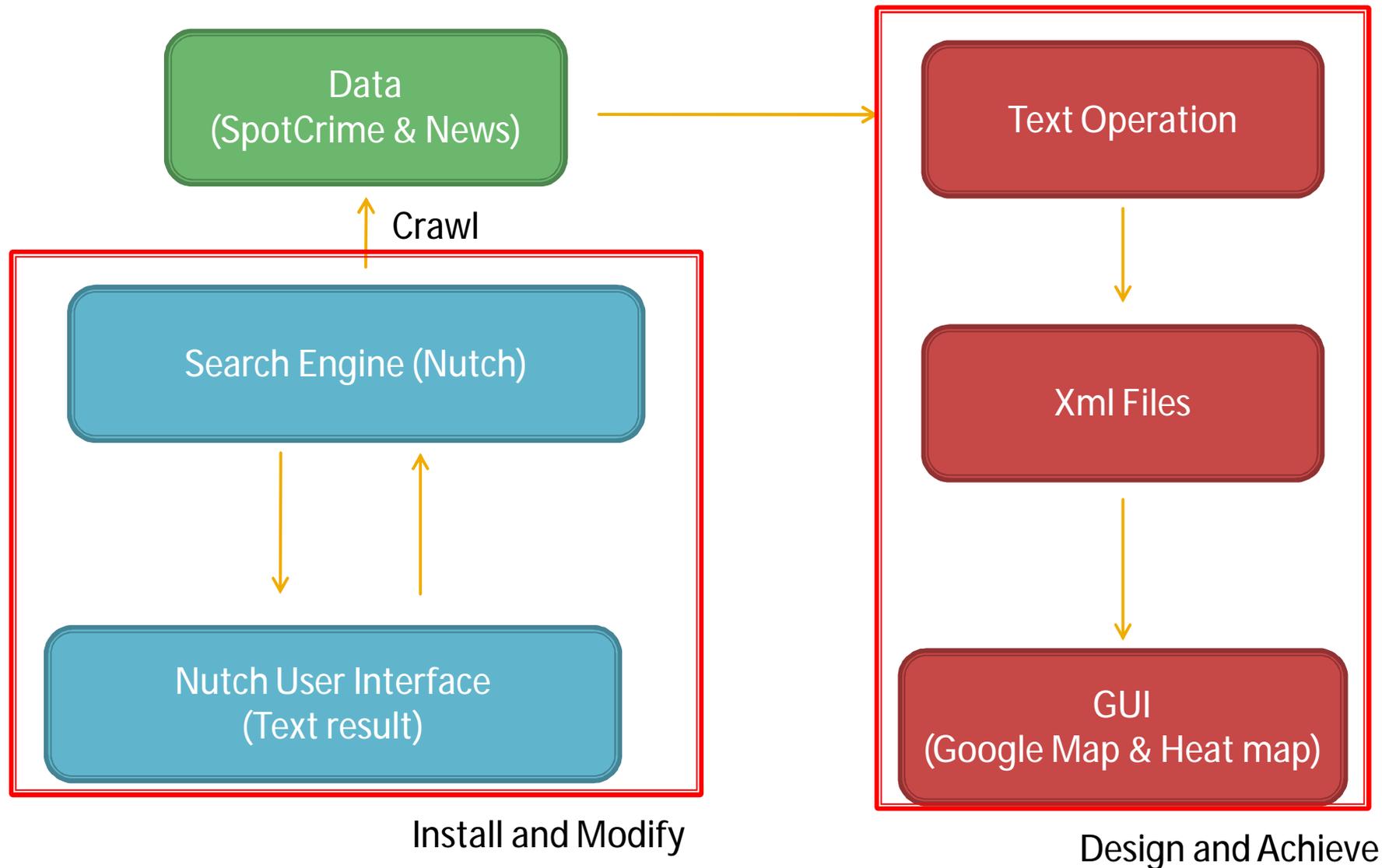
Introduction

Crime analysis involves data about crimes to enable law enforcement to better prevent crimes. This project is an attempt to incorporate data mining techniques like cluster discovery, detection, and co-location mining on crime data. A dataset based on police blotter from major cities in U.S. is used for this project.

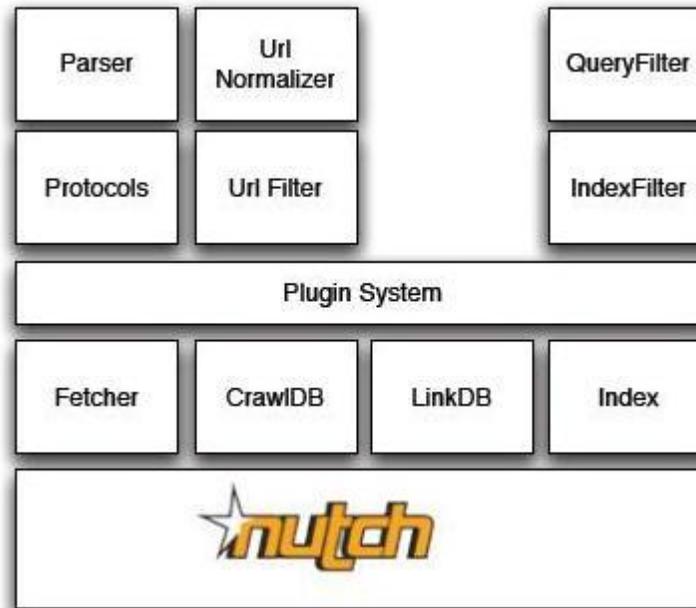
Mining of crime data provides spatial and temporal information about crime patterns.

This project provides a visual interface to display results. Combining spatial and temporal with significance of information.

Project model



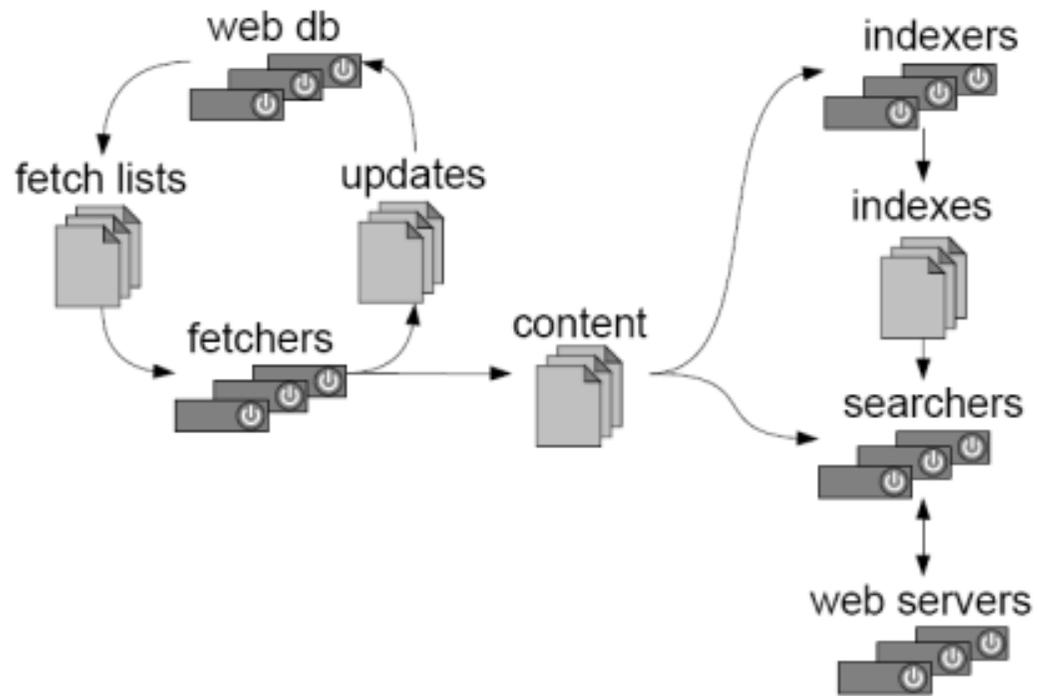
Search Engine



Search Engine



Nutch Architecture



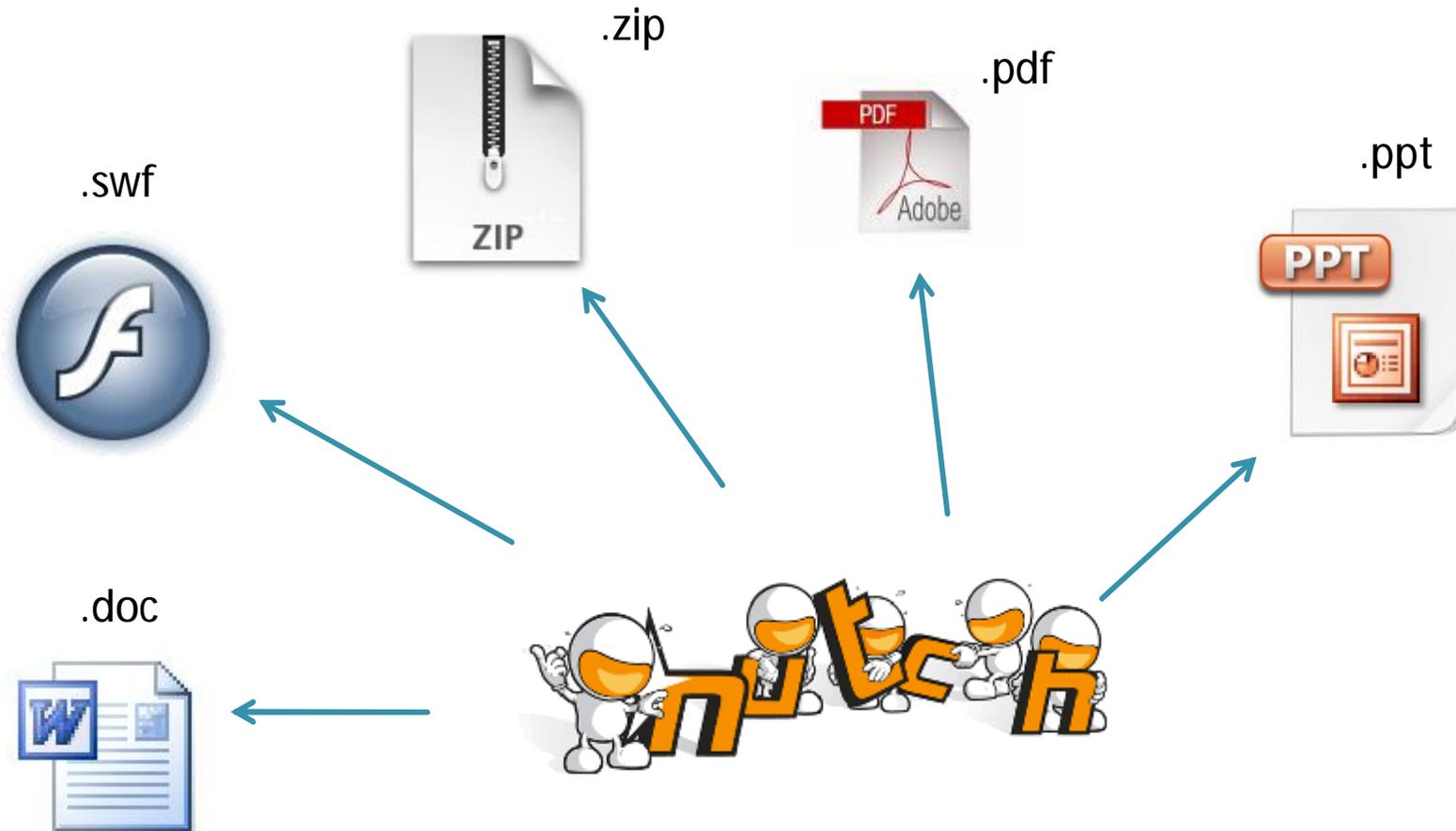
Search Engine



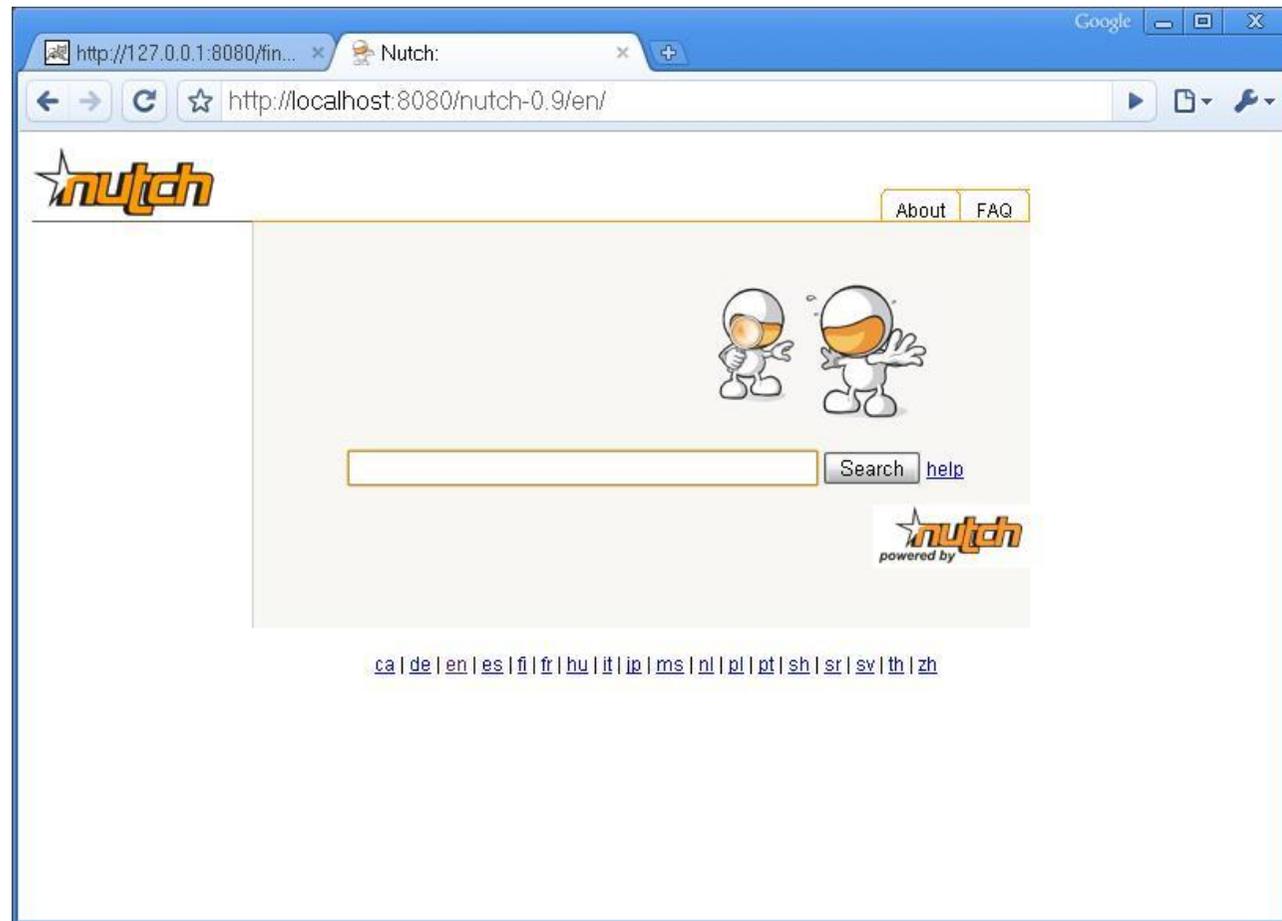
```
Logan@IBM-A92D08AF4 /cygdrive/d/Tomcat/nutch/bin
$ sh nutch crawl urls -dir 0503 -threads 2 -depth 3 -topN 50
LOGAN./cygdrive/d/Tomcat/nutch/conf:/cygdrive/c/Program Files/Java/jdk1.6.0_12/1
ib/tools.jar:/cygdrive/d/Tomcat/nutch/nutch-0.9.jar
crawl started in: 0503
rootUrlDir = urls
threads = 2
depth = 3
topN = 50
Injector: starting
Injector: crawlDb: 0503/crawlDb
Injector: urlDir: urls
Injector: Converting injected urls to crawl db entries.
Injector: Merging injected urls into crawl db.
Injector: done
Generator: Selecting best-scoring urls due for fetch.
Generator: starting
Generator: segment: 0503/segments/20090503142310
Generator: filtering: false
Generator: topN: 50
Generator: jobtracker is 'local', generating exactly one partition.
Generator: Partitioning selected urls by host, for politeness.
Generator: done.
Fetcher: starting
Fetcher: segment: 0503/segments/20090503142310
Fetcher: threads: 2
fetching http://www.spotcrime.com/
Fetcher: done
CrawlDb update: starting
CrawlDb update: db: 0503/crawlDb
CrawlDb update: segments: [0503/segments/20090503142310]
CrawlDb update: additions allowed: true
CrawlDb update: URL normalizing: true
CrawlDb update: URL filtering: true
CrawlDb update: Merging segment data into db.
CrawlDb update: done
Generator: Selecting best-scoring urls due for fetch.
Generator: starting
Generator: segment: 0503/segments/20090503142319
Generator: filtering: false
Generator: topN: 50
Generator: jobtracker is 'local', generating exactly one partition.
```



Plug-ins



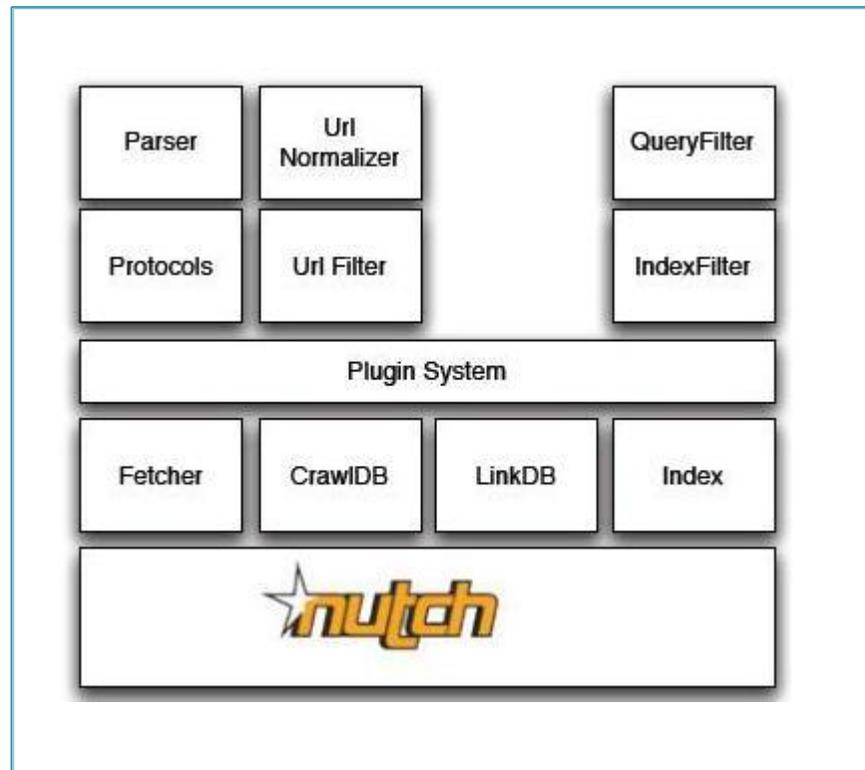
Interface



Interface

A screenshot of a web browser window showing search results for "SpotCrime". The browser's address bar contains the URL "http://localhost:8080/nutch-0.9/search.jsp?query=SpotCrime&lang=zh&hitsPerPage=100&hitsPerSite=0&clust...". The page header includes the Nutch logo and a search bar with the text "SpotCrime" and buttons for "Search" and "help". Below the search bar, the text "Hits 1-100 (out of about 1,021 total matching pages):" is displayed. The search results are listed as follows:

- SpotCrime - Know Your Neighborhood**
... New? Signup Free! Forgot password? **SpotCrime.com** Disclaimer: The data ... from its original source. Neither **SpotCrime.com** nor our data sources ...
<http://www.spotcrime.com/> (cached) (explain) (anchors)
- Register Your Account and Setup a Crime Alert**
... New? Signup Free! Forgot password? **SpotCrime.com** Disclaimer: The data ... from its original source. Neither **SpotCrime.com** nor our data sources ...
<http://www.spotcrime.com/account.php> (cached) (explain) (anchors)
- Contact Us - Provide Your Feedback**
... New? Signup Free! Forgot password? **SpotCrime.com** Disclaimer: The data ... from its original source. Neither **SpotCrime.com** nor our data sources ...
<http://www.spotcrime.com/contact.php> (cached) (explain) (anchors)
- Help**
... New? Signup Free! Forgot password? **SpotCrime.com** Disclaimer: The data ... from its original source. Neither **SpotCrime.com** nor our data sources ...
<http://www.spotcrime.com/help.php> (cached) (explain) (anchors)
- http://www.spotcrime.com/search.php**
... A Crime | Sign Up Search **SpotCrime.com** Disclaimer: The data ... from its original source. Neither **SpotCrime.com** nor our data sources ...
<http://www.spotcrime.com/search.php> (cached) (explain) (anchors)
- Password Recovery**
... New? Signup Free! Forgot password? **SpotCrime.com** Disclaimer: The data ... from its original source. Neither **SpotCrime.com** nor our



Crawl \longleftrightarrow Data?

Source(1/3)

Original idea:

crawl news websites

The screenshot shows the top portion of the Washington Post website. At the top is the masthead with the newspaper's name in a large, stylized font. To the right of the masthead, it says "TODAY'S NEWSPAPER" and provides links for "Subscribe" and "PostPoints". Below the masthead is a horizontal navigation bar with tabs for "NEWS", "POLITICS", "OPINIONS", "BUSINESS", "LOCAL", "SPORTS", "ARTS & LIVING", and "GOING OUT GUIDE". Under the "LOCAL" tab, there are sub-links for "Metro News", "Education", "Columns & Blogs", "District Maryland", and "Virginia". To the right of these are links for "Weather", "Traffic", "Crime", "Obituarie", "Local Ex", and "Commur". A search bar is located below the navigation bar. The date "Sunday, May 3, 2009" is displayed in red. Below the navigation bar, there is a "More News" section with a dropdown menu for "Select RSS Method (may require registr...)". The main content area features several news headlines, each with a date and time in parentheses:

- 2 Students Charged With Plot to Kill Principal, Spark Explosion at Md. School:** [Montgomery High School, Principal Were Allegedly Targeted](#) (Post, April 30, 2009)
- U.S. DISTRICT COURT:** [Woman Pleads Guilty in Greenbelt to Money Laundering in Mortgage Fraud Scam](#) (Post, April 29, 2009)
- Fan of Va. Tech Shooter Pleads Guilty to Sending Threatening E-Mail to Students** (Post, April 29, 2009)
- U.S. NAVAL ACADEMY:** [Midshipman Sentenced to 15 Months for Theft](#) (Post, April 29, 2009)
- Md. Man Arrested in Alleged Meat Cleaver Attack** (Post, April 28, 2009; 3:25 PM)
- Five Charged in Federal Mortgage Fraud Probe:** [U.S. Charges 5; Many Victims Are From Pr. George's](#) (Post, April 28, 2009)
- LOUDOUN COUNTY CRIME:** [Loudoun Prosecutor Considers Seeking Death Penalty in Walker's Slaying](#) (Post, April 28, 2009)
- Man Sentenced to 20 Years in Baby's Beating Death** (Post, April 28, 2009)
- 1 Killed, 3 Wounded in NW Shooting** ([washingtonpost.com](#), April 27, 2009; 6:52 AM)

Source(2/3)

The Washington Post TODAY'S NEWSPAPER
Subscribe | PostPoints

Advertisement

NEWS POLITICS OPINIONS BUSINESS LOCAL SPORTS ARTS & LIVING GOING OUT GUIDE

SEARCH: go washingtonpost.com Web : Results by Google™ | Se

washingtonpost.com > Metro > Maryland > Montgomery

» THIS STORY: [READ +](#) | [Comments](#)

2 Students Charged With Explosion, Slaying Plot

Montgomery High School, Principal Were Allegedly Targeted

By Dan Morse and Aaron C. Davis
Washington Post Staff Writers
Thursday, April 30, 2009

Two Montgomery County teenagers have been charged with arson and conspiracy to commit murder in an alleged plot to kill the principal at their White Oak high

TOOLBOX
AAA Resize Print E-mail
b Yahoo! Buzz Save/Share +

COMMENT

washingtonpost.com readers have posted

Problem: **multiple**
spatial and temporal terms

Source(3/3)

Automatically Identifying Localizable Queries

-Michael J. Welch, Junghoo Cho (SIGIR 2008)

Decision tree **could** solve multiple term problems

Query ID	Base Query	Location Tag
10005397	county florida animal shelter	city:lee
10005397	— county animal shelter	city:florida
10005397	— county animal shelter	state:florida
10005397	florida animal shelter	county:lee county
10005397	— animal shelter	city:florida
10005397	— animal shelter	state:florida
10005397	lee county animal shelter	city:florida
10005397	— county animal shelter	city:lee
10005397	— animal shelter	county:lee county
10005397	lee county animal shelter	state:florida

Table 1: Base Query Generation

Localization Ratio

$$r_i = \frac{Q_i(L)}{Q_i + Q_i(L)}$$

Combined DT, SVM, and NN using a majority voting scheme

Precise Data(1/3)

Local Police Blotter:

May 1, 2009

Police blotter

Police reports in Baltimore city and county

Assault/arrest A man, 22, was arrested Tuesday morning and charged with beating and kicking a woman, 21, during an incident in her home in the 900 block of Leadenhall St.

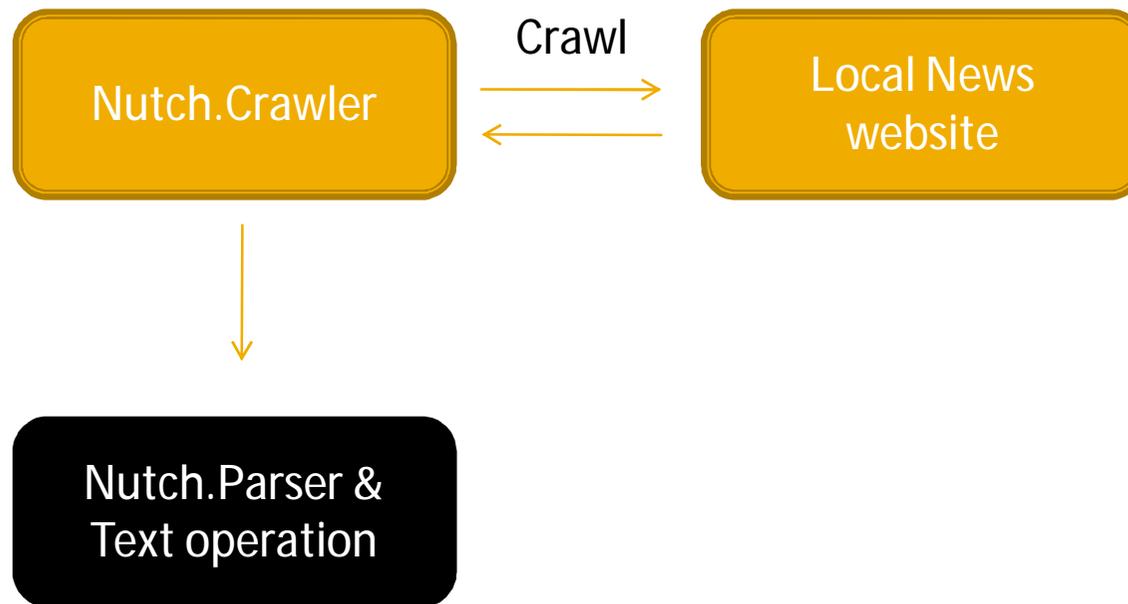
Charged with aggravated assault was Terrell Jones of the 1100 block of Harlem Ave. The woman sustained minor injuries but did not require hospital treatment.

Theft A 1995 Dodge Charger with tags 88N F23 was parked in the 300 block of Washburn Ave. between Monday and Tuesday when someone removed both tags.

Burglary Jewelry of unknown value was stolen Tuesday from a house in the 200 block of S. Stricker St. by someone who entered through a rear second-floor bathroom window.

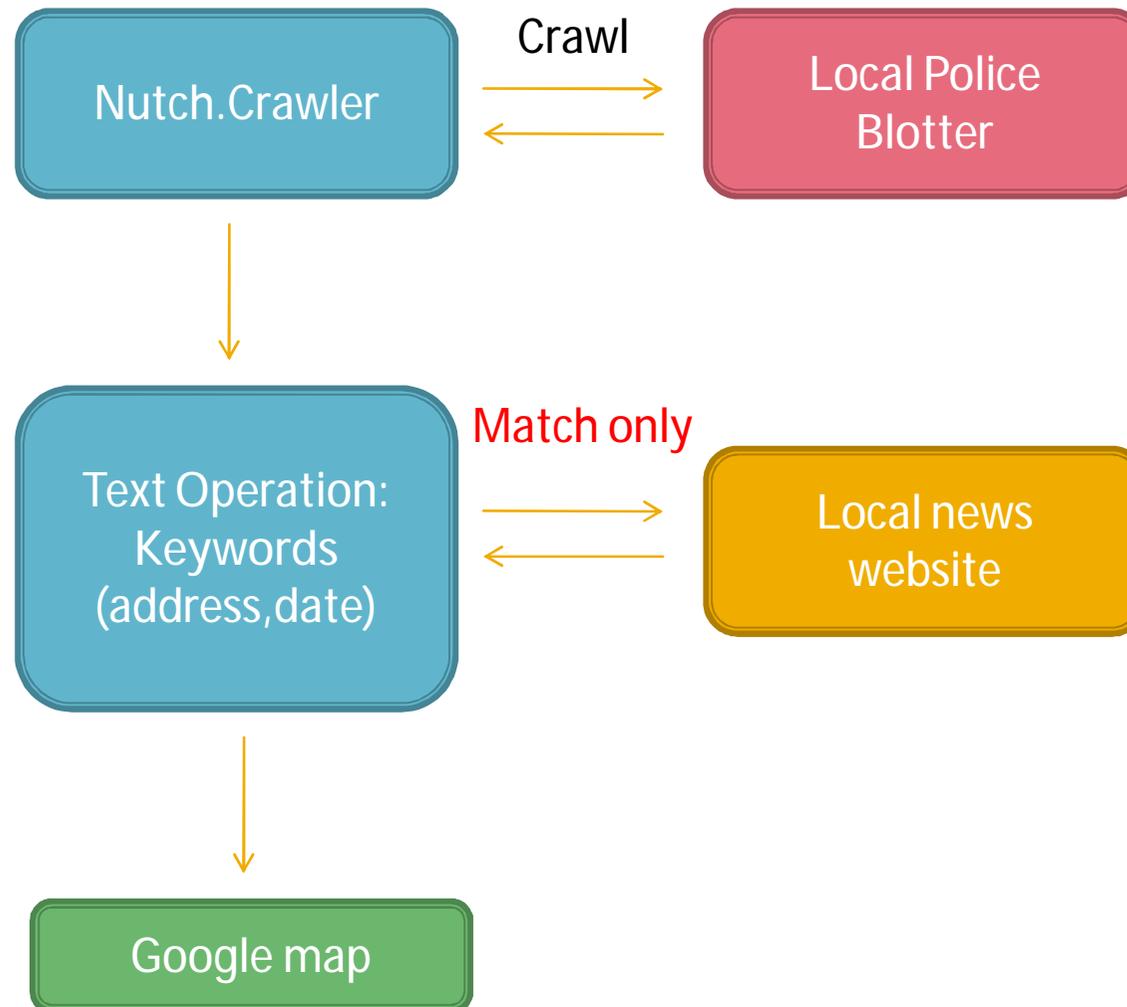
Precise Data(2/3)

Original idea

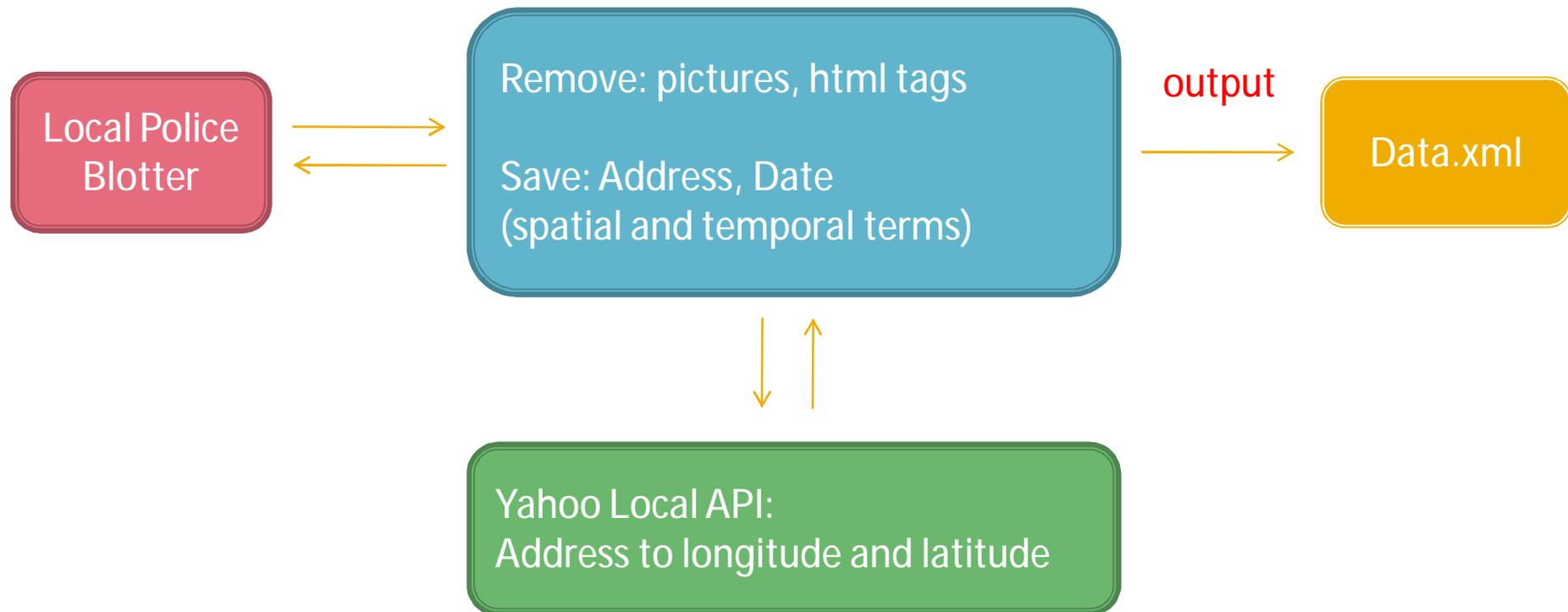


Precise Data(3/3)

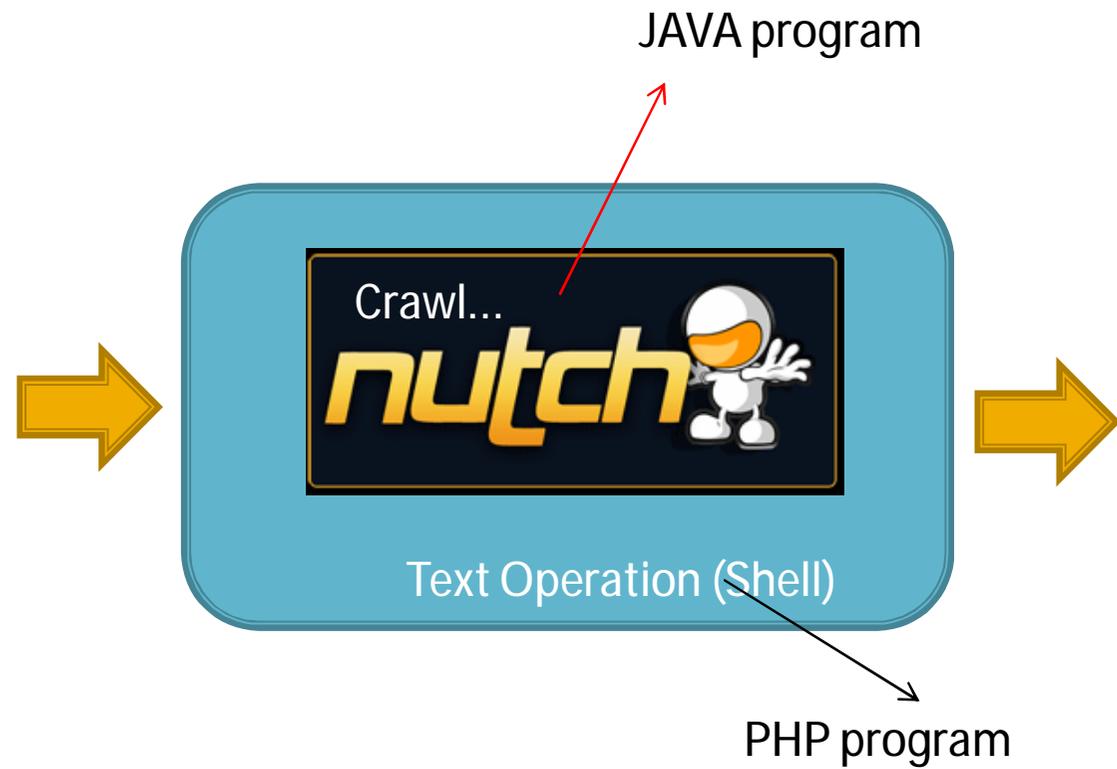
Current idea



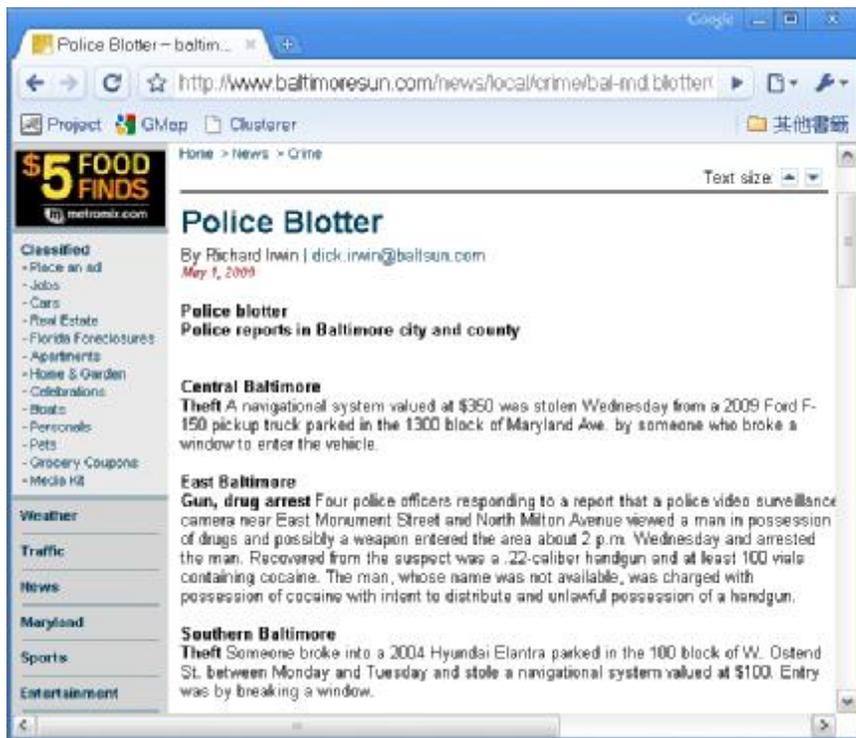
Text Operation(1/3)



Text Operation(2/3)



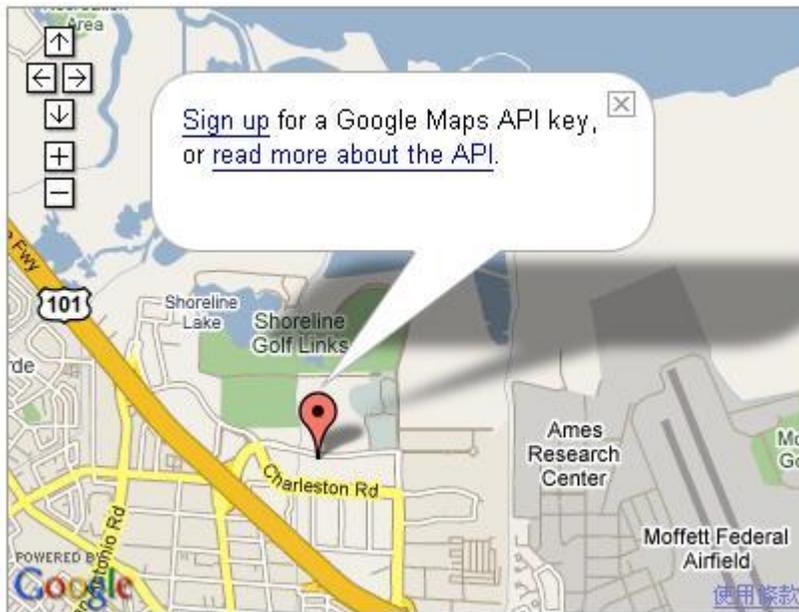
Text Operation(3/3)



```
<?xml version="1.0"?>
<entry>
  <title>M 0.5,</title>
  <georss:point>
    41.091120 -81.516089
  </georss:point>
  <category label="type"
    term="04/21/2009 01:00 AM
    Arrest Justin C. Richmond, 25,
    was driving a stolen Ford
    Escort when police attempted
    to stop him for a traffic
    violation. He led officers on a
    chase driving through"/>
</entry>
```

User Interface

Google API

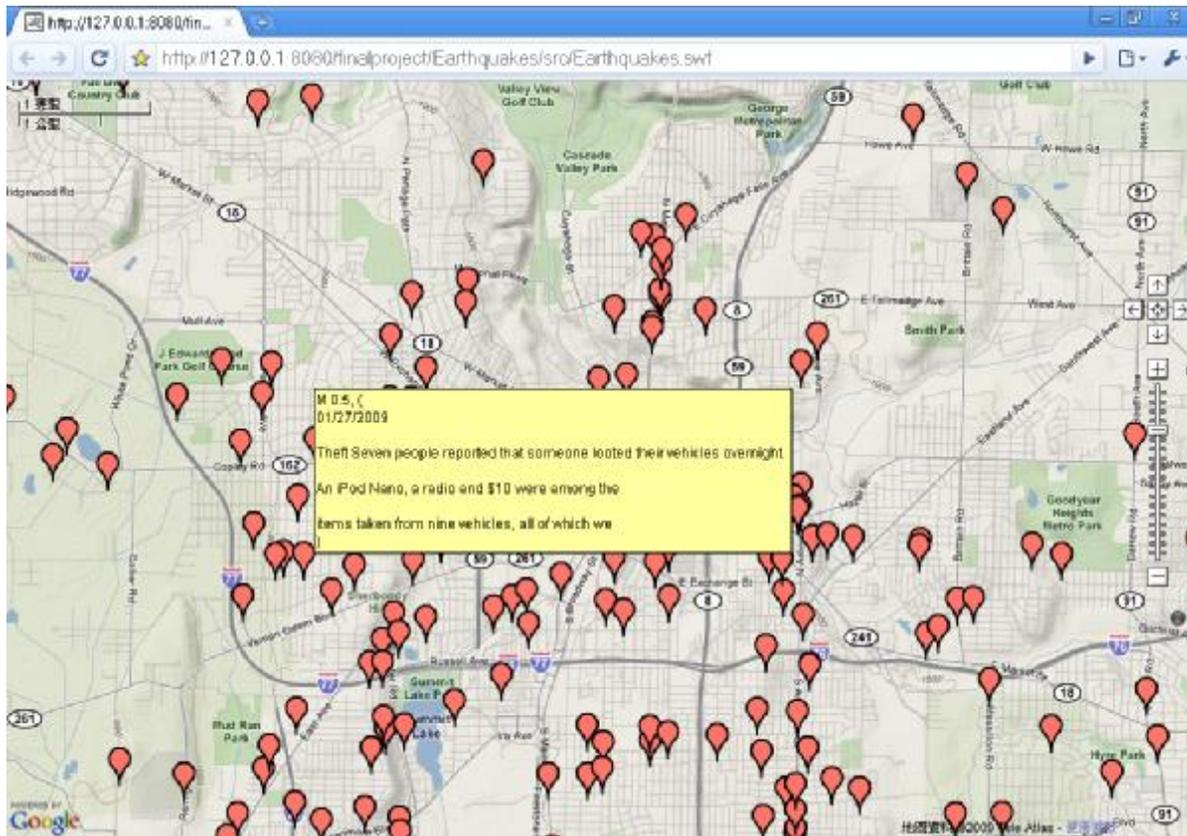


```
<script src="http://maps.google.com/maps?file=api&v=2&key=ABQIAA  
m51RBrvYughuyMxQ-i1QfUnH94QxW1a6N4U6MouMmBA"  
    type="text/javascript"></script>  
<script type="text/javascript">
```

```
function initialize() {  
  if (GBrowserIsCompatible()) {  
    var map = new GMap2(document.getElementById("map_canvas"));  
    map.setCenter(new GLatLng(37.4419, -122.1419), 13);  
  
    // Add 10 markers to the map at random locations  
    var bounds = map.getBounds();  
    var southWest = bounds.getSouthWest();  
    var northEast = bounds.getNorthEast();  
    var lngSpan = northEast.lng() - southWest.lng();  
    var latSpan = northEast.lat() - southWest.lat();  
    for (var i = 0; i < 10; i++) {  
      var point = new GLatLng(southWest.lat() + latSpan * Math.random(),  
                             southWest.lng() + lngSpan * Math.random());  
      map.addOverlay(new GMarker(point));  
    }  
  }  
}
```

```
</script>
```

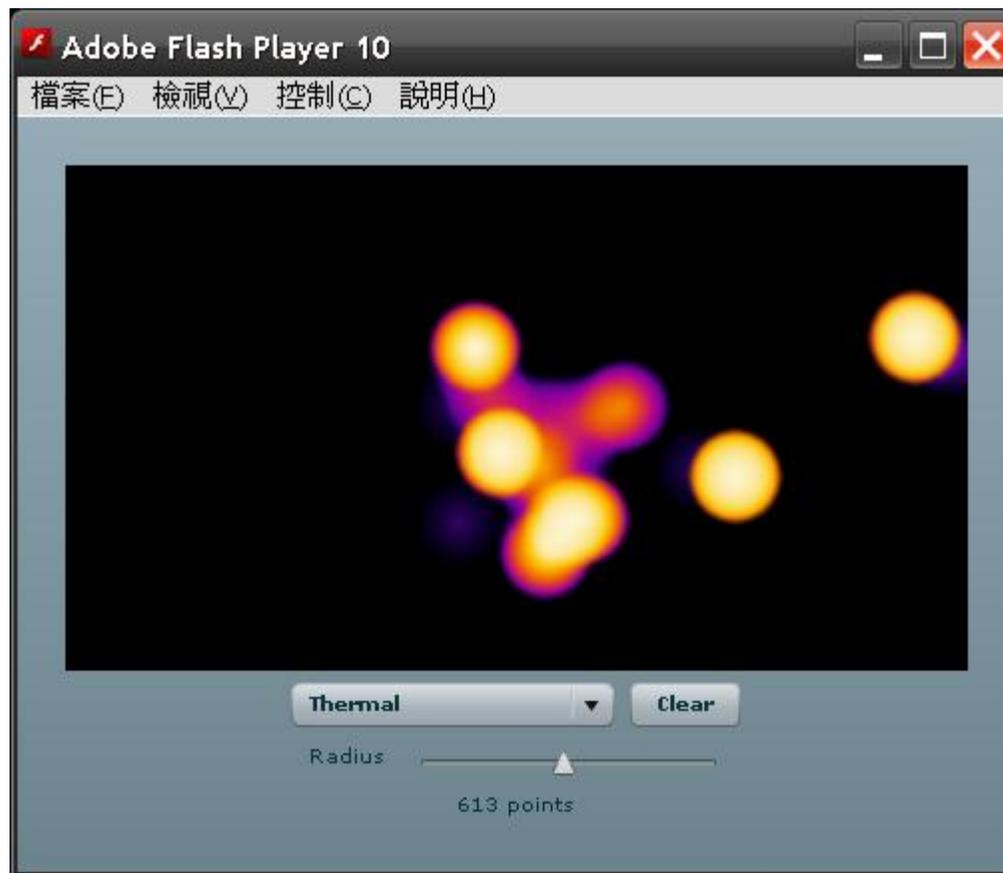
User Interface



Type	Weight
Arrest	0.5
Arson	1.0
Assault	0.6
Burglary	0.5
Robbery	0.7
Shooting	1.0
Theft	0.3
Vandalism	0.1
Other	0.5

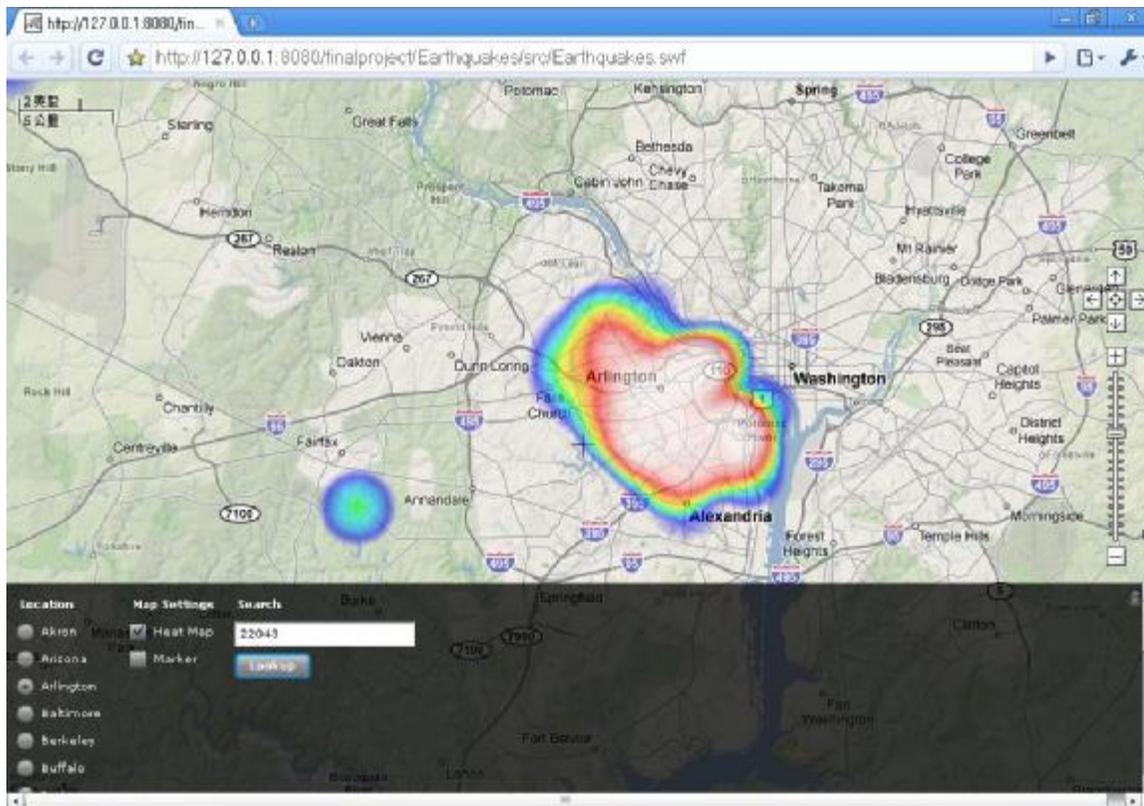
Heat Map

<http://michaelvandaniker.com/blog/2008/07/06/how-to-make-heat-maps-in-flex/>



Demo

Heat Map

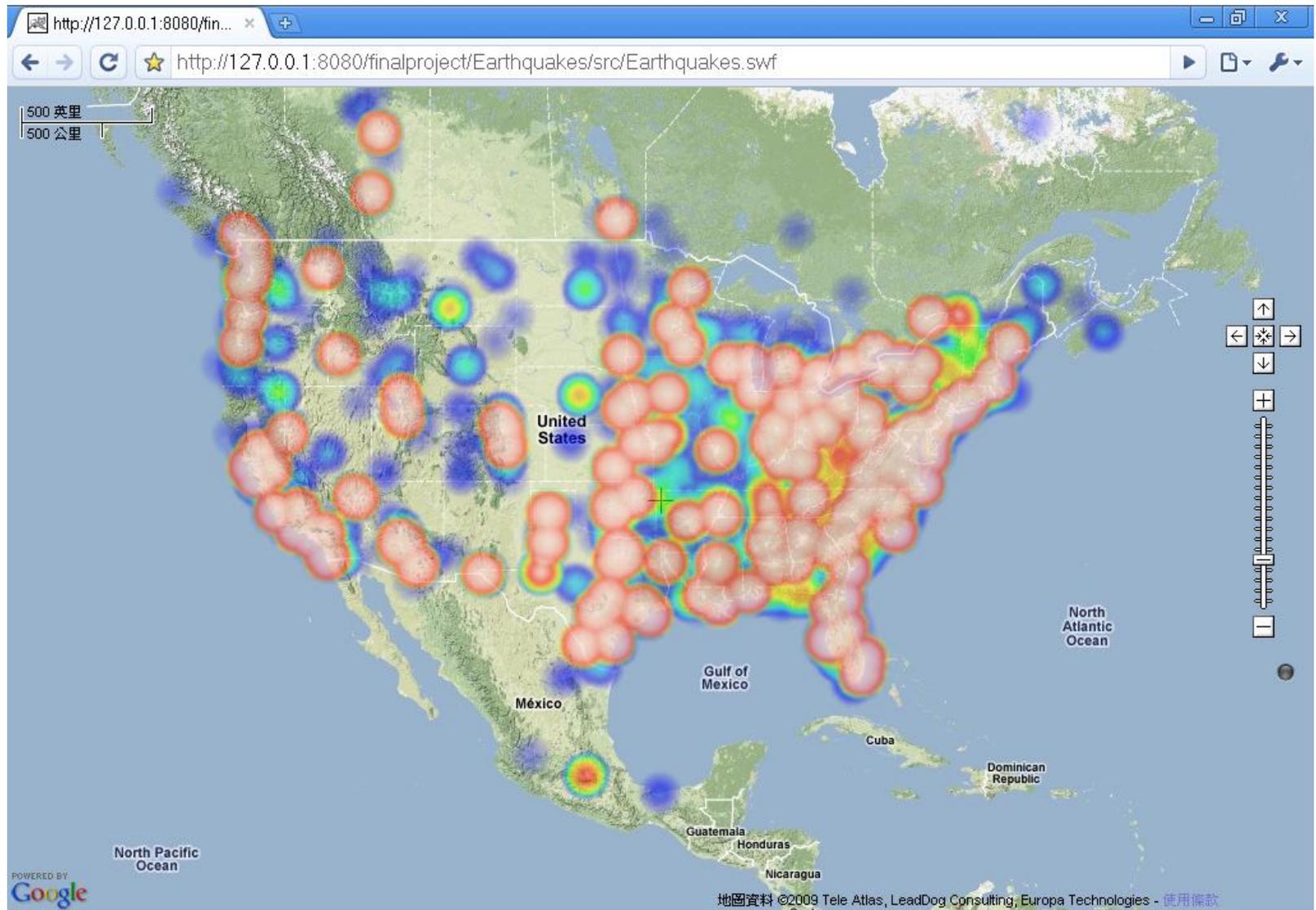


Type	Weight
Arrest	0.5
Arson	1.0
Assault	0.6
Burglary	0.5
Robbery	0.7
Shooting	1.0
Theft	0.3
Vandalism	0.1
Other	0.5

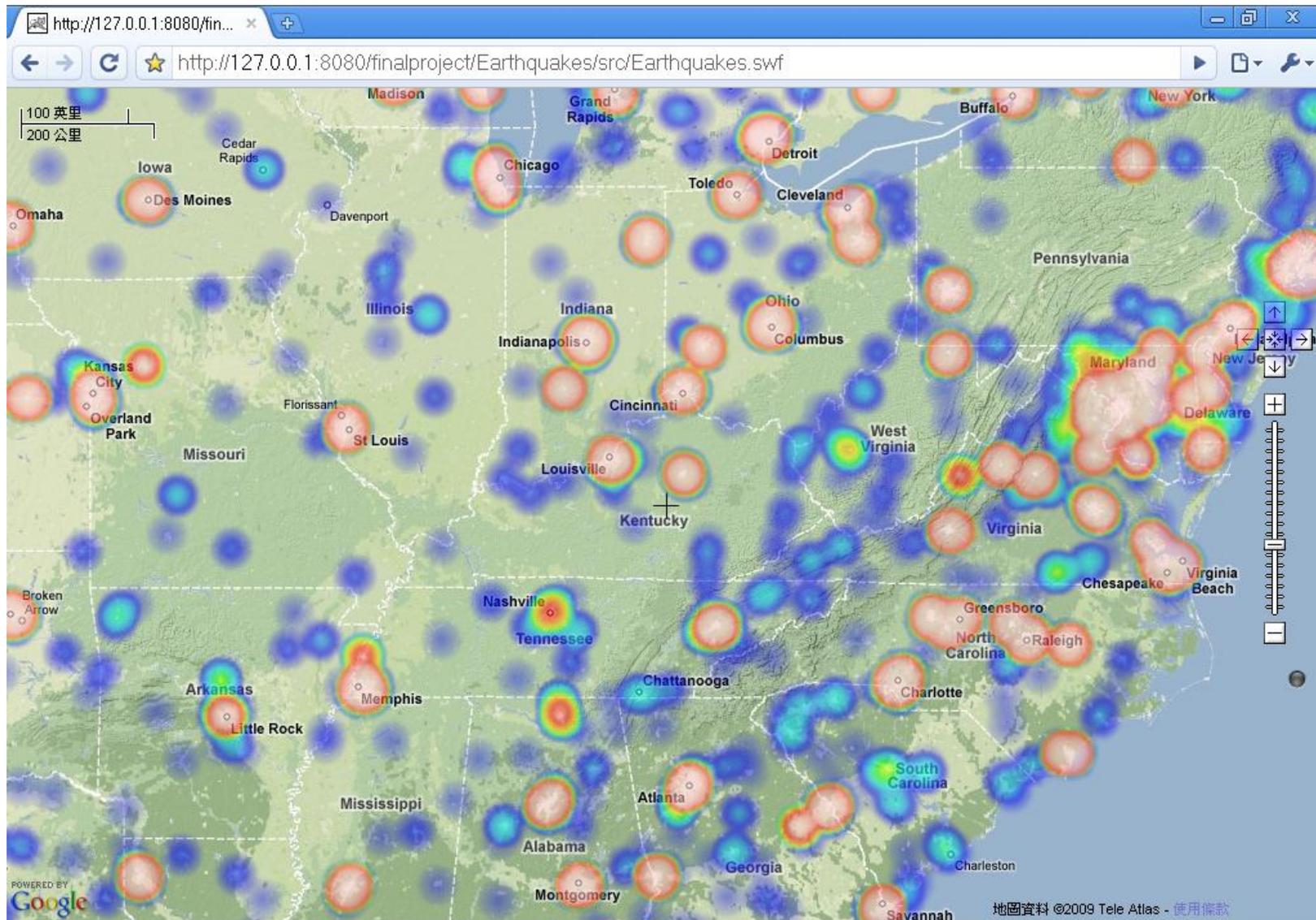
Heat Map

304 major cities in U.S.
Feb – March 2009
66,929 cases

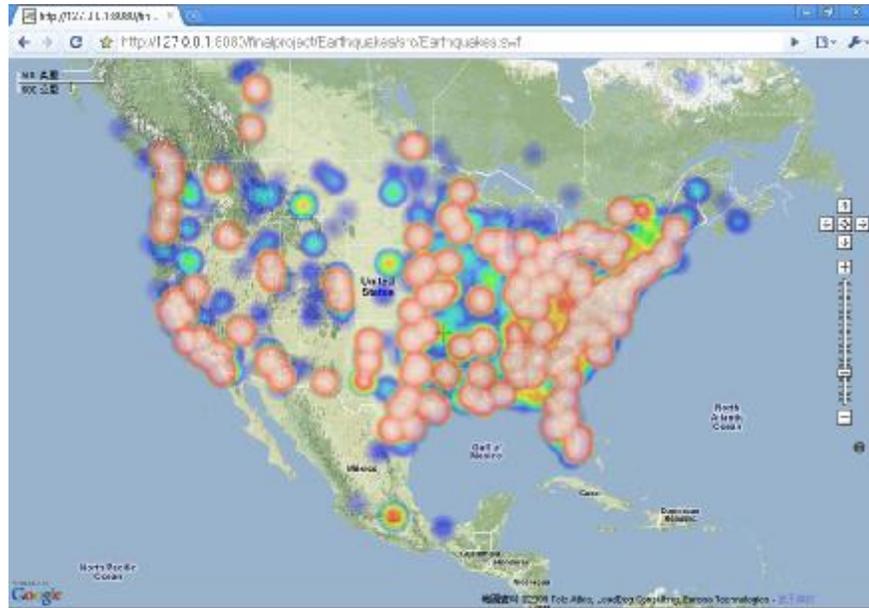
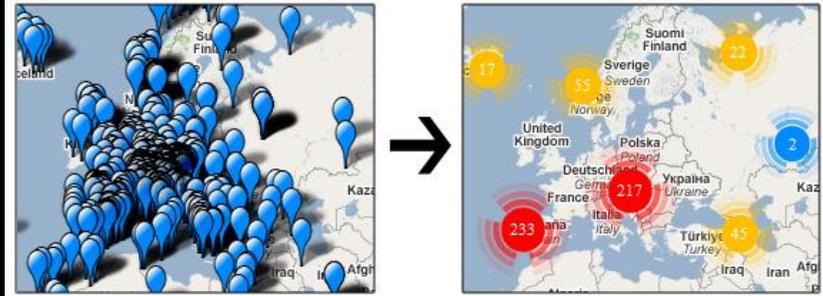
- Akron
- Albamarle
- Alexandria
- Allegany
- Amarillo
- American
- Anaheim
- Anchorage
- Annapolis
- Anne
- Arizona
- Arlington
- Arlington
- Atlanta
- Augusta
- Augusta
- Aurora
- Austin
- Bakersfield
- Baltimore
- Baltimore
- Baltimore
- Bangalore
- Baton
- Battle



Heat Map



Future work: Cluster



3 mins



2 secs

Future work2: Prediction

A Decision Model for Spatial Site Selection by Criminals

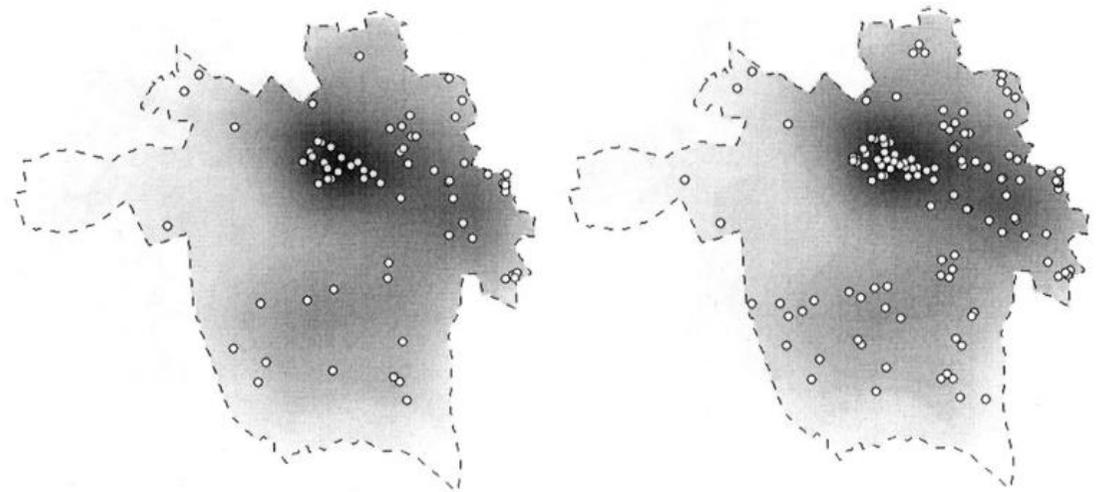
- Yifei Xue and Donald E. Brown, Fellow, IEEE 2002

A method based on space-time point process models.

It predicts future crime locations based on the features of **past crime data** and consider the **decision making** processes of the criminals.

$$P\{a_i | A_d, d\}$$

The probability that an individual decision maker will choose alternative from available choice set.



Prediction of hot spot model with crime incidents from 08/11/97 to 08/17/97 and 08/11/97 to 08/24/97.



Questions?
Thank you!