

# DETECTING ENVIRONMENTAL DISASTERS IN DIGITAL NEWS ARCHIVES

A regional flood database case study

---

Amelia Yzaguirre, Michael Smit, Robert Warren  
Dalhousie University, Halifax, Nova Scotia

October 29, 2015

2015 IEEE International Big Data Conference  
Big Data in the Geosciences workshop

# CONTRIBUTORS



Initiated by Nova Scotia Environment's Climate Change  
Adaptation Project

# PROJECT DETAILS

## NS Environment's Climate Change Adaptation Project

- History of floods in Nova Scotia
- Build a database economically
- Web-based visualization for users

## The Chronicle Herald (data source)

- Two decades of digitized newspaper stories (1992 - Present Day)
- 2 million articles in XML format
- International, regional, and local news coverage

BACKGROUND

## Text mining and GIR

- Goal: convert unstructured text into “interesting” data
- Tools: Natural language processing (NLP), analytical methods, gazetteers
- Typical tasks: text categorization, sentiment analysis, document summarization, semantic similarity

# FRAMING THE PROBLEM

## RELEVANT AND INTERESTING INFORMATION

... No major traffic accidents were reported in the Valley, but the storm resulted in hundreds of cancellations of church services, Sunday schools, bingos and flea markets. Moncton got the worst of the snow, with 52 centimetres reported by late Sunday afternoon. But along the Atlantic coast, in eastern Nova Scotia counties and in Cape Breton, residents had to contend with freezing rain and rain. Most of those areas received five to 10 centimetres of snow before the precipitation changed to rain. The mess left catch basins clogged, flooding streets in many communities. "We got a lot of rain, a lot of flooding in the streets," said Sgt. Barry Gordon of Cape Breton regional police in Sydney. "Public works . . . have been out early clearing catch basins and manhole covers that are ice-covered." Sgt. Gordon said with the snow being followed by rain, some roads were extremely slippery. "You couldn't even stand up on some of them." Yarmouth got off relatively easy, with six centimetres of snow followed by 1.8 millimetres of rain. The Halifax area received ...

**Figure:** Sample text from a news article.



## RELEVANT AND INTERESTING INFORMATION

... No major traffic accidents were reported in the Valley, but the storm resulted in hundreds of cancellations of church services, Sunday schools, bingos and flea markets. Moncton got the worst of the snow, with 52 centimetres reported by late Sunday afternoon. But along the Atlantic coast, in eastern Nova Scotia counties and in Cape Breton, residents had to contend with freezing rain and rain. Most of those areas received five to 10 centimetres of snow before the precipitation changed to rain. The mess left catch basins clogged, flooding streets in many communities. "We got a lot of rain, a lot of flooding in the streets," said Sgt. Barry Gordon of Cape Breton regional police in Sydney. "Public works . . . have been out early clearing catch basins and manhole covers that are ice-covered." Sgt. Gordon said with the snow being followed by rain, some roads were extremely slippery. "You couldn't even stand up on some of them." Yarmouth got off relatively easy, with six centimetres of snow followed by 1.8 millimetres of rain. The Halifax area received ...

**Figure:** Extracting relevant details for our database.

## Opportunities:

- Working with proprietary data
- Large corpus of articles <sup>1</sup>
- Extracting interesting details or summaries from each article

---

<sup>1</sup>2 million newspaper articles, 6.54 GB

# METHODS

## “Off-the-shelf” text mining tools

- Build a dictionary using NLP <sup>2</sup>
- Rule-based filtering using NLP and ML <sup>3</sup>
- Location tagging using Named Entity Recognition and high-precision lists of named entities (*gazetteer*) <sup>4</sup>

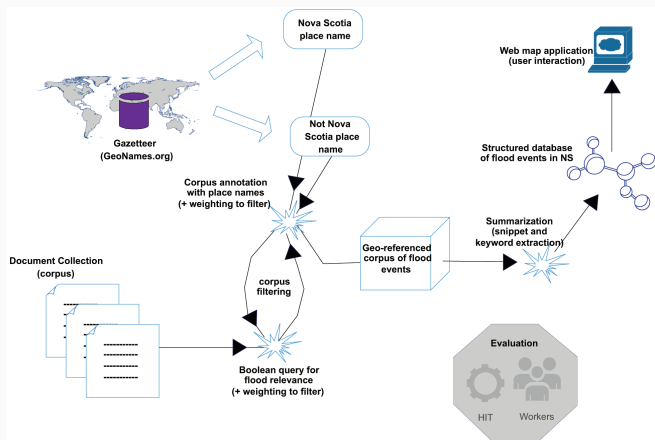
---

<sup>2</sup>Natural Language Toolkit - <http://www.nltk.org/>

<sup>3</sup>scikit-learn - <http://scikit-learn.org/>

<sup>4</sup>Geonames - <http://www.geonames.org/>

# SYSTEM ARCHITECTURE



Set a seed snippet.

The weekend's heavy rain caused a sewer backup which flooded the basements of an apartment building and about 10 area homes on Perron Street.

## Sentences that contain dictionary words.

“... Wild winds overnight Saturday caused a power outage in about 22,500 Halifax households. ... A wire fastened to the pole broke free in the wind, connected with another wire and shorted out, cutting power to homes in Clayton Park, Fairview and Spryfield. Heavy rain that accompanied the high winds also caused problems for some metro residents by flooding their basements. ... A store clerk said customers had been coming in all morning, complaining of basements flooded by cracks in foundations, roof leaks or sewer backups. Several people also had frozen pipes in their homes burst. ...”

Most similar to seed snippet.

“... Wild winds overnight Saturday caused a power outage in about 22,500 Halifax households. ... A wire fastened to the pole broke free in the wind, connected with another wire and shorted out, cutting power to homes in Clayton Park, Fairview and Spryfield. Heavy rain that accompanied the high winds also caused problems for some metro residents by flooding their basements. ... A store clerk said customers had been coming in all morning, complaining of basements flooded by cracks in foundations, roof leaks or sewer backups. Several people also had frozen pipes in their homes burst. ...”



RANDOM SEEDS CULTIVATE VARIETY.

Odie.

“Odie the puppy forlornly stared out over the flooded front lawn he used to romp through, perhaps trying to figure out where it went.”

# EVALUATION

# EVALUATION VIA AMAZON'S MECHANICAL TURK

## Thematic and spatial extraction evaluation

Event Comparator	Accuracy (%)	Geo-location Comparator	Accuracy (%)
Positive	92%	MTurk	52%
Non-negative	95%	Manual ( $n = 20$ )	60%
		MTurk (consistent)	64%

Location extraction is difficult to assess even for human readers.

# THE WEB APPLICATION

EARLY STAGE

# INTERACTIVE MAP

Nova Scotia Flood Events from 1992 to 2015

Home New Window

Discover past flood events in Nova Scotia using any of the fields below.\*

Search around an address (find me)

Enter an address or an intersection

within

Year of flood events: 1992 - 2015 (selected)

1992 2015

Causes (e.g. ice jam, snow, hurricane)

enter a flood cause

Seasons

Winter

Spring

Summer

Autumn

On: Nova Scotia Counties

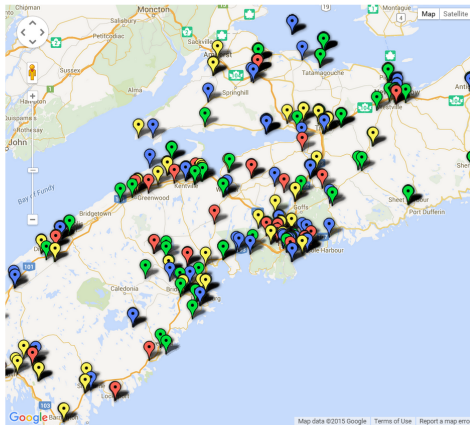
Off

Search

Reset

List view

575 news articles found. Select "List view" to display all stories.



# INTERACTIVE MAP

Nova Scotia Flood Events from 1992 to 2015

Home New Window

Discover past flood events in Nova Scotia using any of the fields below.\*

Search around an address (find me)

Enter an address or an intersection

within

Year of flood events: 1992 - 2015 (selected)

1992 2015

Causes (e.g. ice jam, snow, hurricane)

enter a flood cause

Seasons

- Winter
- Spring
- Summer
- Autumn

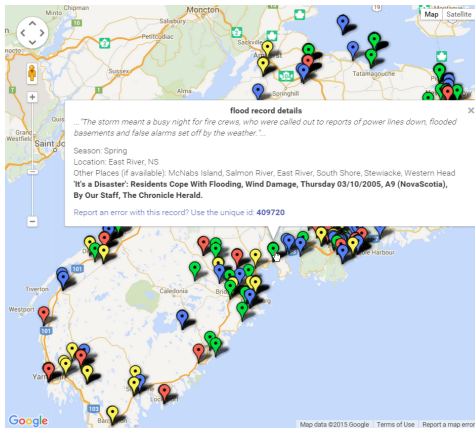
- On: Nova Scotia Counties
- Off

Search

Reset

List view

575 news articles found. Select "List view" to display all stories.



# INTERACTIVE MAP

Nova Scotia Flood Events from 1992 to 2015

Home [New Window](#)

Discover past flood events in Nova Scotia using any of the fields below.\*

Search around an address (find me)

Argyle Street, Halifax, NS, Canada

within 2 km

Year of flood events: 2005 - 2014 (selected)

1992 2015

Causes (e.g. ice jam, snow, hurricane)

enter a flood cause

Seasons

- Winter
- Spring
- Summer
- Autumn

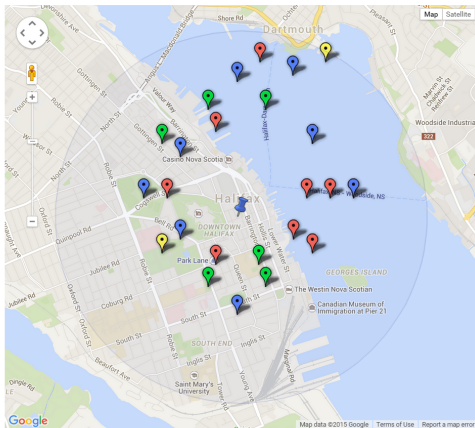
- On: Nova Scotia Counties
- Off

Search

Reset

List view

25 news articles found. Select "List view" to display all stories.



\* Raw data source.



# INTERACTIVE MAP

## Nova Scotia Flood Events from 1992 to 2015

[Home](#)[New Window](#)

Discover past flood events in Nova Scotia using any of the fields below.\*

10 records per page

### Search around an address (find me)

within

Year of flood events: 1992 - 2014 (selected)

1992 2015

Causes (e.g. ice jam, snow, hurricane)

### Seasons

 Winter Spring Summer Autumn On: Nova Scotia Counties Off

26 news articles found. Select "List view" to display all stories.

Year	Location	Story Snippet	Other Places	Citation
1992	Ingonish, NS	Water from the ice-blocked Clyburn Brook continues to feed a snow-covered lake, threatening to flood a road and isolate more than 3,000 people in northern Cape Breton.	Victoria County, Ingonish	Ice Removal to Begin at C.B. River, Friday 02/07/1992, -- (News), By Julie Zatzman and Catherine Ann Fuller, The Chronicle Herald.
1992	Truro, NS	The rain, compounded by melting snowbanks, turned several major streets and highways into rivers, leaving dozens of motorists with stalled or flooded vehicles.	Truro	Flooding Closes Roads, Monday 02/17/1992, -- (News), By Dale Madill Staff Reporter, The Chronicle Herald.
1993	Halifax, NS	If it wasn't record-setting, the rainfall was still enough to keep metro works crews scrambling Saturday, unclogging drains to clear flooding from streets that in a few places were covered by more than a half-metre of water.	Halifax	Well, It Coulda Snowed! Record Warm Weather Saves Shovelers, Monday 02/15/1993, A5 (News), By Our Staff, The Chronicle Herald.
1994	Debert, NS	Heavy rain and snow melting in the Cobequid Mountains on Thursday caused the Debert River to overflow and flood the park, where camper trailers and motor homes are stored.	Cobequid Mountains, Debert	Trailers Hard Hit by Debert River Flood, Saturday 04/09/1994, A3 (News), By Mary Ann Archibald, The Chronicle Herald.
1994	Lower Debert, NS	Rain and melting snow created a blockage in the river, causing it to overflow, said witnesses.	Glenholme	Trailers Damaged in Glenholme Flood, Friday 04/08/1994, -- (News), The Chronicle Herald.
1994	West Brook,	The flooding in the brook was caused basically by rain and snow melt, and we have since	West Brook, Island Lake	County Man Complains of Flooding, Tuesday

# INTERACTIVE MAP

Nova Scotia Flood Events from 1992 to 2015

Home New Window

Discover past flood events in Nova Scotia using any of the fields below.\*

Search around an address (find me)

Enter an address or an intersection

within 50 km

Year of flood events: 1992 - 2015 (selected)

1992

2015

Causes (e.g. ice jam, snow, hurricane)

Odle

Seasons

Winter

Spring

Summer

Autumn

On: Nova Scotia Counties

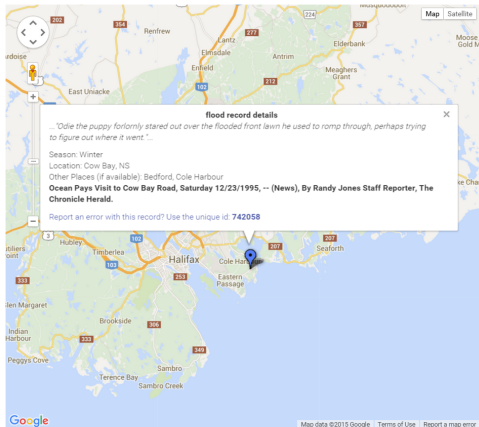
Off

Search

Reset

List view

1 news article found. Select "List view" to display all stories.



\* Raw data source.

## Current Focus

- Integrate multiple “high-res” gazetteers into geo-tagging
- Include other regional datasets, study other types of events
- Evaluation methods
- ??

THANKS!

# ACKNOWLEDGEMENTS

Thank you, and to the organizers of the Big Data in the Geosciences workshop.

## Contributors

- Nova Scotia Environment
- The Chronicle Herald
- The School of Information Management and The Institute for Big Data Analytics at Dalhousie University

## Questions later?

email: [ameliay@mathstat.dal.ca](mailto:ameliay@mathstat.dal.ca)

Site demo: <http://nsfloodhistory.management.dal.ca/>

### Text mining and GIR

- Topic Modeling - Blei, D. (2012). “Introduction to Probabilistic Topic Models”
- List of text mining software on Wikipedia
- Toponym Resolution - Leidner, J. (2007) “Toponym Resolution in Text”
- GIR System example - Adams, B., McKenzie, G., & Gahegan, M. (2015) “Frankenplace: Interactive Thematic Mapping for Ad Hoc Exploratory Searching”

Main site is built with HTML, CSS, and Javascript with the data stored in Google Fusion Tables.

### Technologies used:

- jQuery
- Google Maps API V3
- Google Fusion Tables

Create mobile-friendly interactive maps.

- Leaflet - <http://leafletjs.com/index.html>
- Neatline - <http://neatline.org/>
- CartoDB - <https://cartodb.com/>