

GlobeNow!



Ahmad Lashgar, PhD
Firmware Developer at Avigilon



GlobeNow!

- A feed per city listing events that happening right now
- Find every single event: small or large
- The project started in February 2015
 - Exploring Victoria, BC, Canada



Automatic Event Extraction

- Collect feed of latest posts from Twitter & Instagram:
 - Cache text, image, video, author, & URL
- Text/image sources used to extract events
 - Apply OCR to images to extract text
 - Process all the texts using NLP techniques



NLP Techniques

- Tokenization
 - Sanitize text and keep “important” words, called tokens
- Stemming & lemmatization
 - Reduce space of the input
- Grammar
 - Define rules for extracting dates



Practice

- Token types:
 - Event related: year, month, day, hour, minute, weekday
 - General: separator, undetected
- Grammar:
 - Month + Separator + Day => event on Month, day
 - Weekday => event on coming Weekday
 - +6



Example

“Official classes are starting next week! We will be hosting free trial classes for new members wanting to try us out AND for returning members too! . . . ~ On Jan 13 and Jan 15 classes will run from 6-8pm at the performance theatre on the second floor of the Nest. All levels are welcome! ...”

2 events extracted on:

- January 13
- January 15



Example (continued)

- Extracted event is available at Instagram:



GlobeNow! available at <https://ahmado.com/GlobeNow>

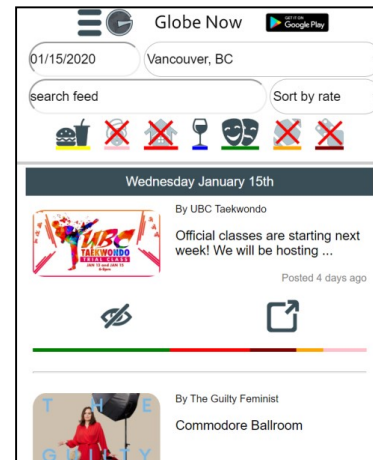
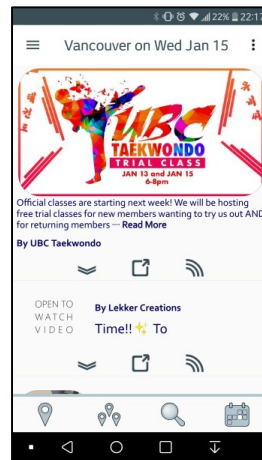
Organizing The Events

- Store future events only
 - Group by data
 - Group by city
 - Stored as a Json file
- List events of the city on a given day as a feed



Platforms: User Interfaces

- Website:
 - ahmado.com/GlobeNow
- Android App:
 - play.google.com/store/apps/details?id=myapps.globenow



GlobeNow! available at <https://ahmado.com/GlobeNow>

False Positives

- Example:
 - “January 15 going to be a great day.”
- Do we get false positives?
 - A lot!




How to address false positives?

- Filter words
 - E.g. exclude adult contents
- Rank events & drop low-ranked
 - An application of machine learning



How to rank events using ML?

- Simple linear model $Y = A * X + B$
 - X is a boolean vector indicating words of the text
 - A & B are trained using good and bad datasets
 - Y is the rank (floating point number)
- Training datasets:
 -  – High-rank: Twitter/Instagram pages that post events
 - Low-rank: Twitter/Instagram pages that do not post events



Text Topic

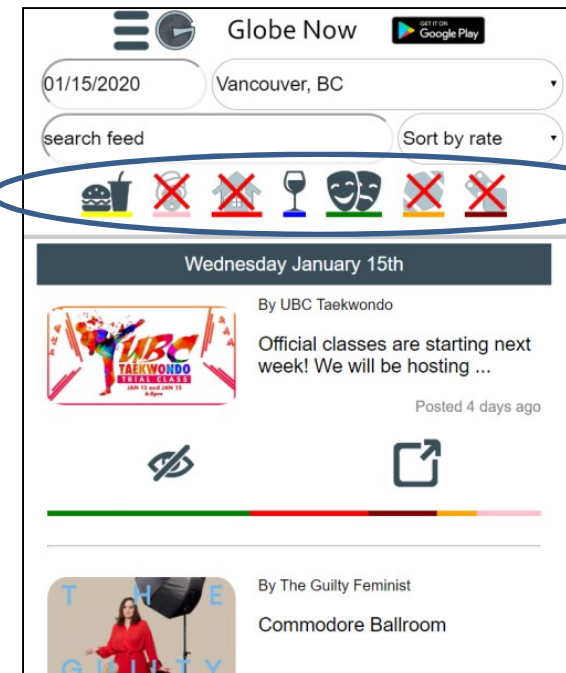
- Would be nice to categorize events
- Challenges:
 - Keyword-based approaches are not effective for long text
 - Hard to list all keywords of all topics without overlapping
 - How many topics is optimal?
- Solution:
 - Latent Dirichlet Allocation: an unsupervised topic modeling
 - After categorization, we label every set manually



Text Topic (continued)

```
- {
  status: "queued",
  title: "Time!!🌟 To",
  url: "https://www.instagram.com/p/B7BdezUBqtQ",
  media: "https://scontent-ort2-2.cdninstagram.com/v/t51.2885-15/e35/80695166",
  author: "24100838563",
  tweetid: "2216182151672933200",
  source: "Instagram",
  authorid: 40,
  text: "🌟Giveaway Time!!🌟 To celebrate the New Year we're giving you a ch
  Contest will close January 15, 2020 at 12:59pm MST. Winner will be contacte
  #yow",
  ml_rating: 0.9674237966537476,
  prettytext: "🌟Giveaway Time!!🌟 To celebrate the New Year we're giving yo
  only. Contest will close January 15, 2020 at 12:59pm MST. Winner will be co
  prettyauthor: "Lekker Creations",
  topicssw7: [
    - {
      topic: "contest",
      prob: 0.78,
      weight: 16118
    },
    - {
      topic: "realestate",
      prob: 0.07,
      weight: 1541
    },
  ],
}
```

GlobeNow! available at <https://ahmado.com/GlobeNow>



Technologies

- Services running on an AWS machine
- Data collector, NLP, and ML models written in Python
 - NLP core written from scratch, NLTK was too general
 - TensorFlow used for inference
- Web interface written in HTML/CSS/PHP
- Android App written in Android



Current State

- Presented the idea at Dragons' Den:
 - Didn't make it to final round (project not profitable)
- Up and running for Vancouver and Victoria at:
 - ahmado.com/GlobeNow
- Listing about 200 events per day for Vancouver!



Thank you.

