Machine Learning Hackathon '17

November 24th – 25th

Milan Sismanovic, SUB IT

November 24th, 2017

Logistics

Schedule Friday 24th

08:00 - 08:30	Get together and Breakfast
08:30 - 10:00	Topic Intorduction
10:00 - 10:45	Q&A
10:45 - 12:30	Hackathon Part I
12:30 - 13:00	Lunch
13:00 - 18:00	Hackathon Part II
18:00 - 19:00	Dinner
19:00 - 22:00	Hackathon Part III
22:00	CS Tower closes

Schedule Saturday 25th

08:00 - 08:30   Breakfast

08:30 - 12:30   Hackathon Part I

12:30 - 13:30   Lunch

13:30 - 15:00   Solution submission

15:00 - 17:00   Presentations

17:00 - 18:00   Jury decision meeting

18:00 - 19:00   Award Ceremony
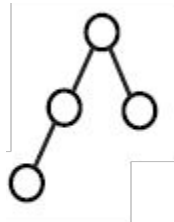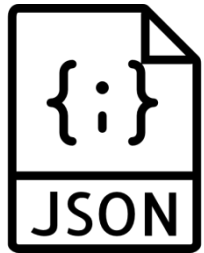
from 19:00   Apero

The Challenge

Taxonomy Definition

- Wikipedia Definition: **Taxonomy** is the practice and science of classification. The word finds its roots in the [Greek language](#) τάξις, *taxis* (meaning 'order', 'arrangement') and νόμος, *nomos* ('law' or 'science').

- Examples:
  - Amazon Product Classification
  - News Article Classification
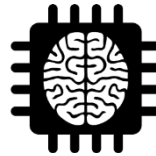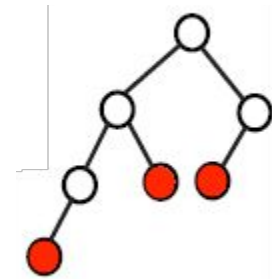 - Wikipedia category taxonomy

Hackathon ML Challenge

Reuters News
Articles

JSON

Taxonomy
0

Machine
Learning

Taxonomy
1

Add nodes only to
leafs

Reuters News Articles Format

seqNo: 5001

storyId: urn:newsml:reuters.com:20160509:nL2N18320D

date: 2016-05-09T11:00:00Z

headLine: INSIGHT-Before massive Bangladesh heist, New York Fed feared such cyber attacks

topics: "PXP", "HACK", "LAW", "CLJ", "CRIM", "MTGFX", "BANK", "RTRS", "BD", "FRAUD1", "FED", "MTPIX", "US", "EREP", "FINS", "ASIA", "EMRG", "POL", "BISV", "LEN", "NEWS1", "PH", "GEN", "INSGHT", "CEN", "MCE", "ASEAN", "BSVC", "AMERS"

topicsDescription: "Picture Available", "Computer Crime / Hacking / Cybercrime", "Lawmaking", "Crime / Law / Justice", "Crime", "Graphics", "Banks", "Reuters News", "Bangladesh", "White Collar Crime / Fraud / Corporate Crime", "US Federal Reserve", "Pictures", "United States", "Enterprise Reporting", "Financials", "Asia / Pacific", "Emerging Market Countries", "Government / Politics", "Banking & Investment Services", "English", "Major News", "Philippines", "General News / Events", "Central Banks / Central Bank Events", "Economic Events", "ASEAN Countries", "Banking Services", "Americas", "News"

story: (Repeating for additional clients with no changes to text) By Jonathan Spicer and Jim Finkle New York, May 9 (Reuters) - In the years before hackers stole $81 million from a Bangladesh central bank account at the Federal Reserve Bank of New York, senior Fed security officials examined the risk of such an attack - but judged the prospect unlikely, bank sources told Reuters. The Fed managers worried that […]

seqNo: 5002 #### storyId: urn:newsml:reuters.com:20160509:nL2N186013 #### date: 2016-05-09T11:00:00Z #### headLine: Movie theater chain goes on trial over 2012 Colorado massacre #### topics: "LAW", "US", "SOCI", "CYCS", "CMPNY", "CLJ", "CRIM", "ENT", "RELX", "POL", "ENTS", "LEN", "MURD", "FLM", "JUDIC", "RTRS", "GEN", "USACO", "GUNCTR", "CCOS", "AMERS" #### topicsDescription: "Lawmaking", "United States", "Society / Social Issues", "Cyclical Consumer Goods & Services", "Company News", "Crime / Law / Justice", "Crime", "Arts / Culture / Entertainment", "Leisure & Recreation", "Government / Politics", "Hotels & Entertainment Services", "English", "Murder / Manslaughter", "Film", "Judicial Process / Court Cases / Court Decisions", "Reuters News", "General News / Events", "Cyclical Consumer Services", "Americas", "News" #### story: By Keith Coffman DENVER, May 9 (Reuters) - Jury selection was due to begin on Monday in Colorado for the first civil trial of wrongful death and personal injury claims stemming from a 2012 mass shooting in which 12 people were killed and dozens wounded in a suburban Denver movie theater. A group of more than two dozen plaintiffs, including surviving victims and relatives of the dead, have sued the […]

Taxonomy File Format

☐ Indentation represents the depth of the node

Reuters News Topics;;;;;;;

;News Topics;;;;;;

;;General News Topics;;;;;

;;;Arts Culture and Entertainment;;;;

;;;;Dance;;

;;;;Exhibitions;;

;;;;Festivals;;

Results in

```
Reuters News Topics
|-News Topics
||-General News Topics
|||-Arts Culture and Entertainment
||||-Dance
||||-Exhibitions
||||-Festivals
```

Useful Information

Performance Measurement/Test Data

- Algorithm Performance Measurement Page
  - https://hackathon17.mope.ml/HackathonSite/
- Algorithm Performance Measurement Web Service
  - https://hackathon17.mope.ml/HackathonSite/upload
  - Use POST and the params
    - *teamname* (UTF-8 string) for your team name
    - *file* (UTF-8) for your taxonomy file
  - Response is a JSON: {d :0.3145674356, teamname:spartans}
- Test Data
  - Initial taxonomy 0: https://hackathon17.mope.ml/HackathonSite/T0.csv

  - 300 news items: https://hackathon17.mope.ml/HackathonSite/News300.txt
  - Taxonomy 1 after above items: https://hackathon17.mope.ml/HackathonSite/T1.csv
  - More news items for context: https://hackathon17.mope.ml/HackathonSite/NewsAll.txt