

Registrant Classification

Huayi Jing/Sebastian Castro, NZRS
Registrar Conference 2017



What? Why? How?



Problem definition

- Each domain registration has a registrant name
- We don't get what type of registrant is
Is a person? Is an organization?
- Motivation
 - Trend detection
 - Future targeted campaigns
 - Increase understanding

How?

- Registrant names

Jeremy Ashford → Person

Jacqui → Person

Treecare Ltd → Organization

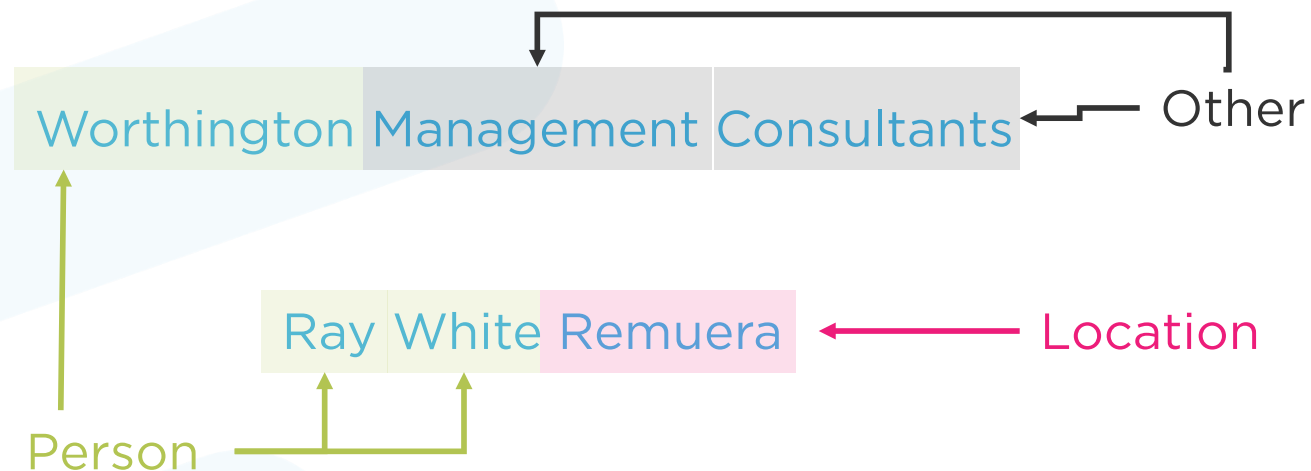
Techsoft → Organization

Tan → ???

wjja → ???

Attempt 1: NLP

- NLP: Natural Language Processing
Process Natural Language Corpora using computers
- NER: Named Entity Recognition
Extract named entities from text, such as names of persons, locations, expressions of times



Attempt 1

- 2000 hand classified registrant names
- Use a few rules
 - If contains Ltd, Limited, GmbH
 - Most of the words represent Person
- 96% accuracy
- Problems
 - It works well with the test list
 - It has a bias towards Organizations

Attempt 2

- Deep Learning
 - Character level aggregation
 - Word embeddings
- Not enough labelled data
 - Get a student to classify names, 30k unique registrant names
- CNN are complex artifacts
 - Slow to train
 - Multiple parameters



Attempt 3

- Traditional Machine Learning
 - As baseline for Deep Learning attempt
 - Random Forest
 - Naïve Bayes
 - SVM

Model Comparison

- Total training data : 29347 names
- Train / test : 90 / 10

Metric	CNN + word2vec	CNN + Glove	Character Level	Naïve Bayes	SVM	RF
Accuracy	96%	96%	94%	85%	86%	91%
Training Time	40 min	40 min	>7 hr	< 1 Sec	< 1 Sec	836 s



Future Work

- Fine tune
- Share aggregated counts with registrar, industry, general public
And publish regular updates on IDP
- Make code available

Contact: {huayi,sebastian}@nzrs.net.nz

www.nzrs.net.nz

NZRS 
an  InternetNZ company

