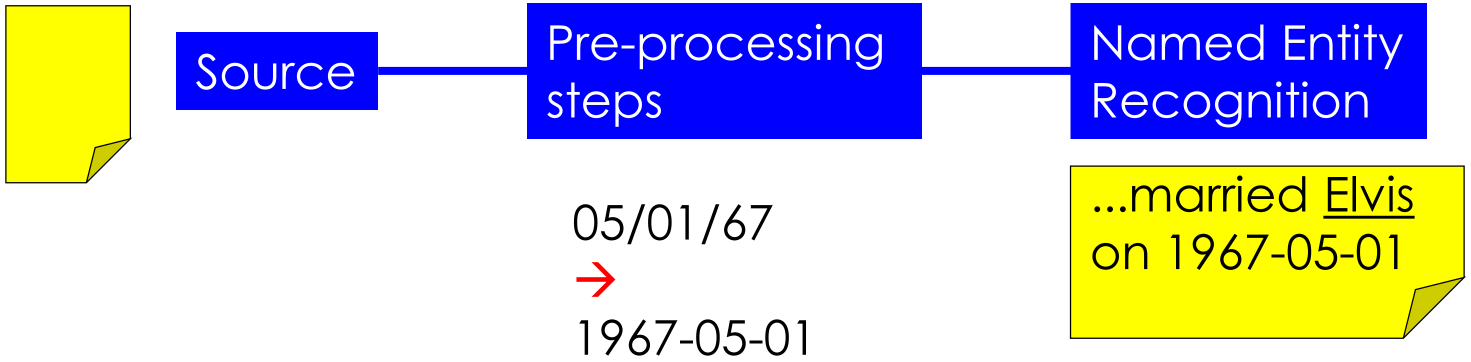
An Overview of Named Entity Recognition (NER)

Lim Cheng Yang

Introduction (Information Extraction)

Information Extraction (IE) is the process of extracting structured information from unstructured machine-readable documents



Other processes

Elvis Presley	PERSON
1967-05-01	TIME

Introduction

- What is NER ?
  - NER is a subtask of information extraction (IE) that aims to identifies which snippets in a text are mentions of entities in the real world
- Extract information from unstructured texts
  - With the purpose of classifying them into their respective class
- Common classes
  - Person
  - Organization
  - Location
  - Misc.
    - Money
    - Percent
    - Date
    - Time

NE and Question Answering

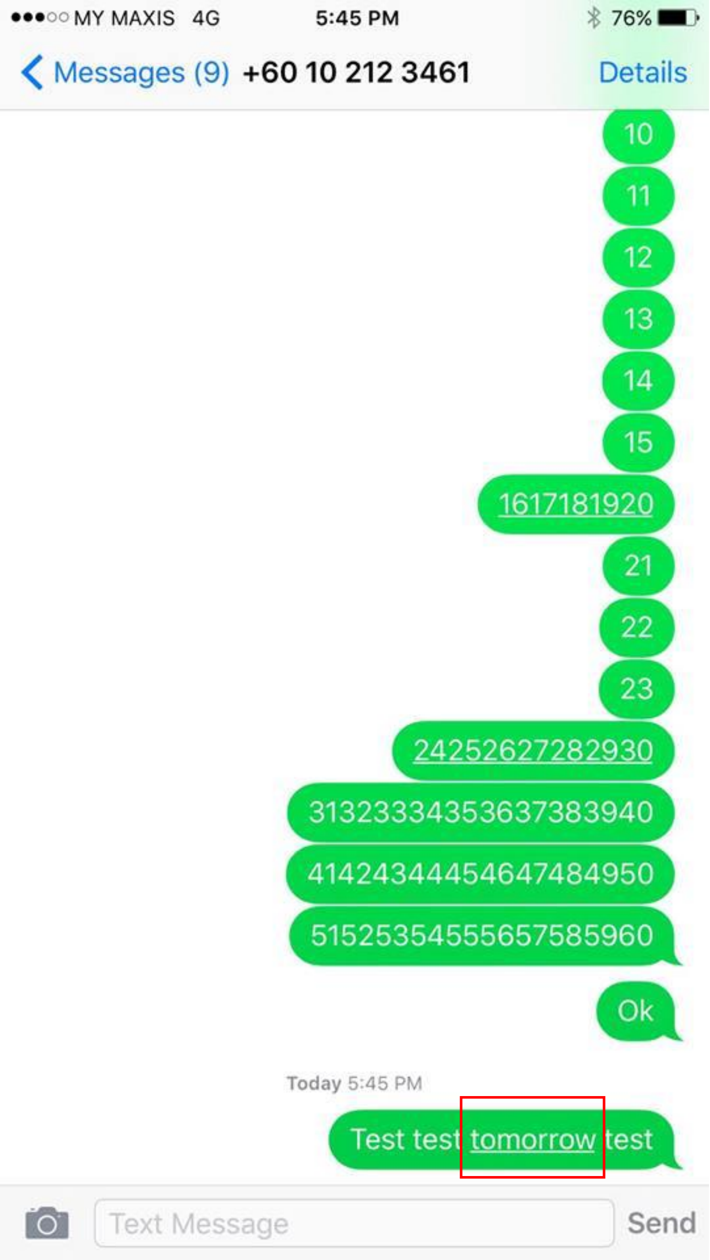
- NER is used to extract the who, where, when, and how in the sentence.
- Often, the expected answer type of a question is a NE
  - Where does the bombing incident happened ?
    - The answer type is a LOCATION
  - When will be the next talk ?
    - The answer type is a DATE
  - Who is the first Malaysian astronaut to do a spacewalk ?
    - The answer type is a PERSON
- NE are used to answer questions and usually are noun phrases

NER in action

May 19 1995, Atlanta -- The Centers for Disease Control and Prevention, which is in the front line of the world's response to the deadly Ebola epidemic in Zaire , is finding itself hard pressed to cope with the crisis…

Information Extraction System

Date	*Disease Name*	*Location*
May 1995	Ebola	Zaire
July 1995	Mad Cow Disease	U.K.
Feb. 1995	Pneumonia	U.S.



NER in action

- NER is widely being used in various application
- Example: Iphone imessage
  - Detects time

Background

- NER was designed to be use on long texts
  - Well structured
  - Correct grammar
  - Achieved 85 – 90% accuracy
- However, its performance drops significantly when it comes to Internet Relay Chat (IRC) type of texts
  - 30 – 50%
- Why ?
  - NER is a sequential labelling tasks which utilizes contextual clues to label the entities
  - Traditional NER also relies heavily on capitalization

Background (cont.)

- IRC texts are short, less grammatical, contains unorthodox capitalization and contains other elements.
  - Hashtags
  - User mentions
  - Abbreviation
- These elements are noises which will cause the performance of the system to drop.
- Currently, researchers are focusing on solving this problem.

Challenges

- NER systems, although as good as it is claimed, it still have some limitations
- It is claimed that:
  - NER systems only perform well in the domain where it is trained in
  - NER systems accuracy will drop if it encounters entities that it haven't seen before

Existing tools

- There are a lot of existing tools that offers the functionality of NER

Feature	ANNIE	Stanford NER	Ritter et al.	Alchemy API	Lupedia
Approach	Gazetteers and Finite State Machines	CRF	CRF	Machine Learning	Gazetteers and rules
Languages	EN, FR, DE, RU, CN, RO, HI	EN	EN	EN, FR, DE, IT, PT, RU, ES, SV	EN, FR, IT
Domain	newswire	newswire	Twitter	Unspecified	Unspecified
# Classes	7	4, 3 or 7	3 or 10	324	319
Taxonomy	(adapted) MUC	CoNLL, ACE	CoNLL, ACE	Alchemy	DBpedia
Type	Java (GATE module)	Java	Python	Web Service	Web Service
License	GPLv3	GPLv2	GPLv3	Non-Commercial	Unknown
Adaptable	Yes	Yes	partially	No	No

	DBpedia Spotlight	TextRazor	Zemanta	YODIE	NERD-ML
Approach	Gazetteers and Similarity Metrics	Machine Learning	Machine Learning	Similarity Metrics	SMO and K-NN and Naive Bayes
Languages	EN	EN, NL, FR, DE, IT, PL, PT, RU, ES, SV	EN	EN	EN
Domain	Unspecified	Unspecified	Unspecified	Twitter	Twitter
# Classes	320	1779	81	1779	4
Taxonomy	DBpedia, Freebase, Schema.org	DBpedia, Freebase	Freebase	DBpedia	NERD
Type	Web Service	Web Service	Web Service	Java (GATE Module)	Java, Python, Perl, bash
License	Apache License 2.0	Non-Commercial	Non-Commercial		GPLv3
Adaptable	Yes	No	No	Yes	Partially

NER approaches

- Most common ways to perform NER
  - Machine learning (ML)
  - Rule-based methods
  - Gazetteer lookup

- Lately, ML have been widely adopted
  - Especially SVM and CRF
- However, each methods have their own pros and cons.

Features

- Descriptors or characteristic attributes of words designed for algorithmic consumption
- The features can be classified into:
  - Word level features
  - List lookup features
  - Document and corpus features
- For Rule-based, NER problem is resolved by applying a rule system over the features
- For ML, the features will serve as input to detect the entities

Word	Lemma	PoS	case	Gaze
the	the	Art	low	
seminar	Seminar	Noun	low	
at	at	Prep	low	
4	4	Digit	low	
pm	pm	Other	low	timeid
will	will	Verb	low	

Pre-processing

- Tokenization
  - Split the sentence into individual words
- Part of speech (POS) tagging
  - Gives tags to each of the words
    - NN - noun
    - VB – verb
    - Etc.
- Shallow parsing (chunking)
  - Combining multiple words together into a phrase (commonly use clues from POS tags)
- Normalization
  - Process of transforming texts into a single canonical form.
- Stemming
  - Process of reducing inflected words into their root form

NER

- Sequential labelling tasks
  - Assign named entity tags to each word
  - Commonly used scheme: BIO
    - B – begin
    - I – Inside
    - O – outside
  - Or BILOU
    - L – Last
    - U - Unit
- Common tags for NER are:
  - Person (PER), Location (LOC), Organization (ORG), Miscellaneous (MISC)
  - B-PER, I-PER, B-LOC, etc.
- Example:

O		O	O	O	O	B-LOC	I-LOC	I - LOC	O	O	O
NPE	:	Flash	flood	near	the	Tmn	Maju	Jaya	U	-	Turn

O	B-LOC	I-LOC	O	O	O	O	O	O	O
towards	Jalan	Templer	,	pls	avoid	this	route	for	now

NER (Machine Learning)

- Utilizes the features to predict the entities
  - Based on the features of the words in the surroundings:
    - POS tags
    - Position of words
    - Capitalization
- After obtaining these information
  - The system will calculate the probability of each words respective to each class
    - The word will be classified as the class which have the highest number
      - Example “Barack” from “In the recent years, Barack Obama has …”
        - B-Person 0.875
        - I-Person 0.645
        - B-ORG 0.223
        - I-ORG 0.345

Sample

DAP secretary - general Lim Guan Eng will brief top PKR leaders Tuesday night on his party's case for snap elections in Penang.

-	N-2	N-1	N	N+1	N+2
Word	-	-	DAP	secretary	-
POS tag	-	-	NN	NN	-
Capital	-	-	True	False	False
Position	-	-	1	2	3

Tag	Probability
B-PER	0.085
I-PER	0.035
B-ORG	**0.258**
I-ORG	0.156
B-LOC	0.225
I-LOC	0.103

Sample

DAP secretary - general Lim Guan Eng will brief top PKR leaders Tuesday night on his party's case for snap elections in Penang.

-	N-2	N-1	N	N+1	N+2
Word	-	general	Lim	Guan	Eng
POS tag	-	JJ	NN	NN	NN
Capital	False	False	True	True	True
Position	3	4	5	6	7

Tag	Probability
B-PER	0.585
I-PER	0.235
B-ORG	0.258
I-ORG	0.156
B-LOC	0.225
I-LOC	0.103

NER (Rule based)

- Why rules ?
- Many real-life extraction tasks can be conveniently handled through a collection of rules
- In general, rule-based systems consists of:
  - A collection of rules
  - A set of policies to control the firings of multiple rules

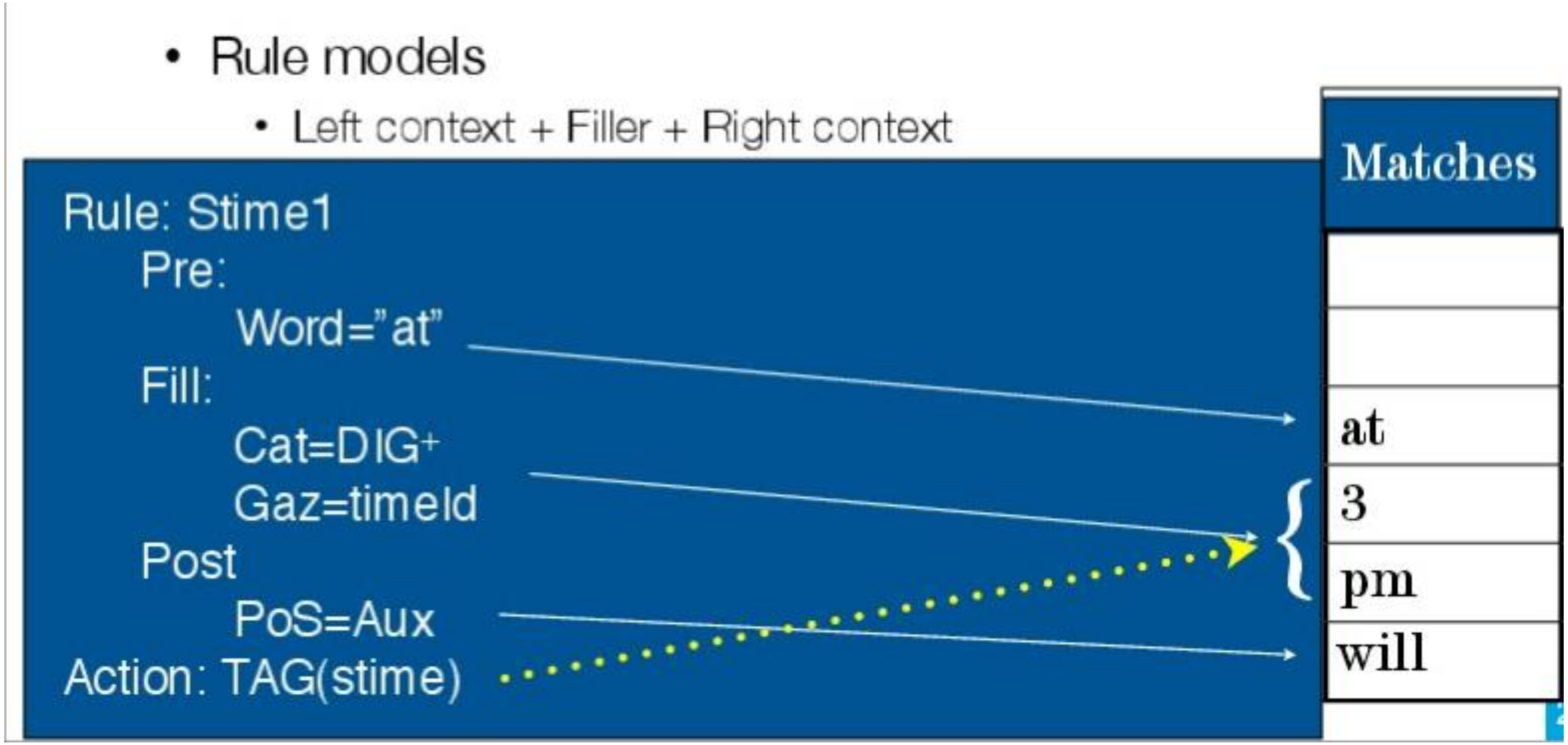
Rule based Example (1)

- Sample rules:
  - If the text before “:” is not number, it is a location
  - The text after “at”, and is a noun, it is a location
  - The text after “a”, “an” is a state
    - A stalled lorry

from gate.ac.uk

```
Rule: Company1
  ( ( {Token.orthography == upperInitial} )+
      {Lookup.kind == companyDesignator}
  ):match
 -->
   :match.NamedEntity = { kind=company, rule="Company1" }
```

Rule based Example (2)



NER (Gazetteer)

- Gazetteer approach is done by keyword searching
  - Based on a predefined dictionary
- Have the highest precision
- It will work well for formal texts that doesn’t have spelling problems
- However, the if the entities are not recorded in the dictionary, it couldn’t detect the entity
- Disadvantage:
  - Couldn’t cope with the expansion of words
  - Potential set of NE is too large to include in the dictionary
  - Names changing constantly and appear in many variant forms

Summary of approaches

	Machine Learning	Rule based	Gazetteer
Characteristic	• Accuracies varies with different input of features	• Works well in specific domain	• Very high precision • Works very well on formal documents/texts
Advantages	• Trainable • Adaptable • Reduces manual effort	• Easy to comprehend • Easy to maintain • Easy to incorporate domain knowledge • Easy to trace and fix cause error	• Sure fire method
Disadvantages	• Requires large training set • Requires manual annotation of data • Opaque	• Requires manual hand-crafted rules	• Unable to keep up with the expansion of vocabulary • Vulnerable to spelling error, abbreviations

ML vs Rule-Based

- In the recent years, most recent academic research starts from the assumption that ML approach is the best approach for solving IE problem.
- Only 6 papers relied solely on rule-based
- However, in commercial world, things are reversed

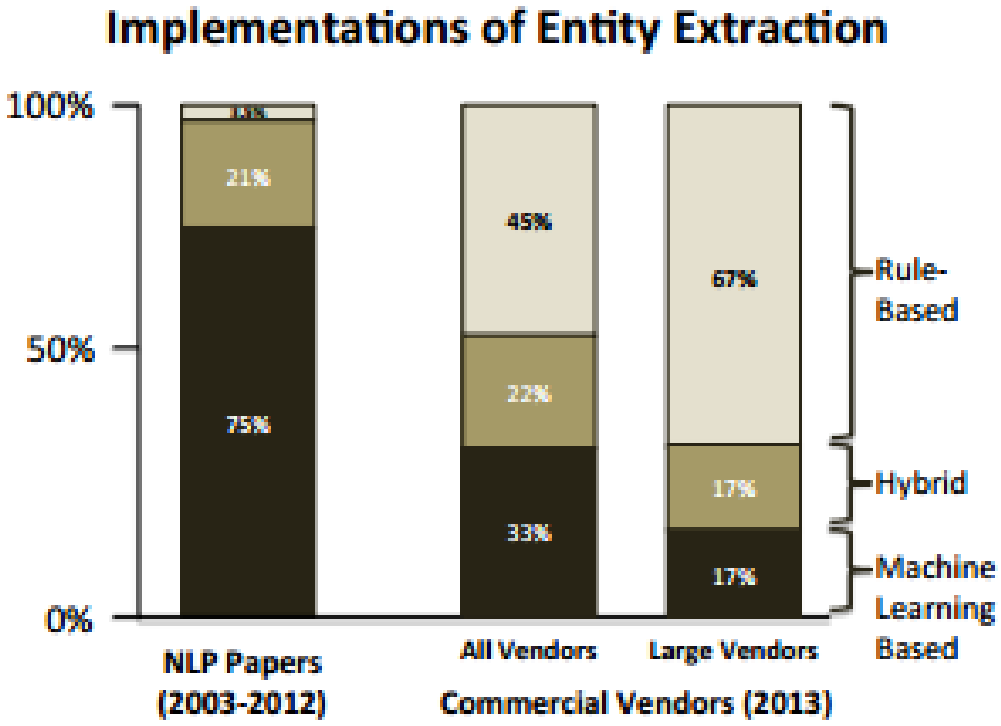


Image taken from Chiticariu, L et al. 2013

Source of disconnection

- Academic papers evaluates IE performance in terms of precision and recall over standard labelled data sets.
- Reality of business world is much more fluid and less well defined.
- In business context, IE must function well with metrics that are ill-defined and subject to change.
  - ML-based models requires a careful upfront definition of IE task, fit poorly in these metrics
  - Rule-based model is preferred for its interpretability, which makes it easier to adopt, understand and maintain

Key design decision in NER system

- How to represent text chunks in NER system?
- What inference algorithm to use?
- How to model non-local dependencies
- How to use external knowledge resources in NER ?

End