

# TDSI

Temasek Defence Systems Institute

# Temasek Defence Systems Institute

Contact: tdsbox2@nus.edu.sg

## CONDITION-BASED MAINTENANCE (CBM) FOR RAIL VEHICLES: A DATA-DRIVEN APPROACH

Goh Yu Heng, ST Engineering  
Professor Emel Aktas, Cranfield University

Objective: Information Extraction of Event Data (Fault Description) for CBM Program.

### Problem:

- Event data recorded in unstructured free text.
- Event data consists of affected component (e.g. tamping tine), and actual defect observed (e.g. damage, missing, break).
- Information cannot be readily extracted as they are all combined in a body of text in a single column.
- The intention is to utilize text mining techniques to extract relevant information.

### Text Analytics Process

#### Information Extraction via Text Analytics

##### Data Preparation

Import data & clean corpus

##### Keyword Dictionary

N-gram modelling & dictionary building

##### Information Extraction

Classification logic

- Lower case transformation
- Remove of numbers
- Remove of punctuations
- Remove stopwords (from tm package in R)
- Remove user-defined words (e.g. LHS, L/H/S)
- Remove whitespaces

N-gram	Rows	Columns
1	2599	742
2	2599	665
3	2599	202
4	2599	67

```
#manual stemming + short forms handling
word <- c('leaking', 'replaced', 'damaged', 'hyd', 'tamping', 'banks',
'gear box', 'cyl', 'seals', 'snapped', 'doors', 'yoyo', 'washers',
'welding', 'welds', 'twisted', 'lines', 'tightened', 'tested', 'stripped',
'faulty', 'extinguishers', 'display', 'debur', 'cracked', 'chafe',
'chaffed', 'chaffing', 'buttons', 'bolts', 'bearings', 'auxiliary',
'alternators', 'adjusted', 'accumulator', 'fray', 'broken')

tran <- c('leak', 'replace', 'damage', 'hydraulic', 'tamp', 'bank',
'gearbox', 'cylinder', 'seal', 'snap', 'door', 'yoyo', 'washer',
'weld', 'welds', 'twist', 'lines', 'tighten', 'test', 'strip',
'fault', 'extinguisher', 'display', 'debur', 'crack', 'chafe',
'chafe', 'chafe', 'button', 'bolt', 'bearing', 'auxiliary',
'alternator', 'adjust', 'accumulator', 'fray', 'break')

library(stringr)
df_corpus <- tm_map(df_corpus, function(x) str_replace_all(fixed(x, word, tran,
vectorize_all = FALSE)))

df_corpus <- tm_map(df_corpus, function(x) str_replace_all(fixed(x, word2, tran2,
df_corpus <- tm_map(df_corpus, function(y) str_replace_all(fixed(y, word2, tran2,
vectorize_all = FALSE)))
```

Data Preparation - Cleaning

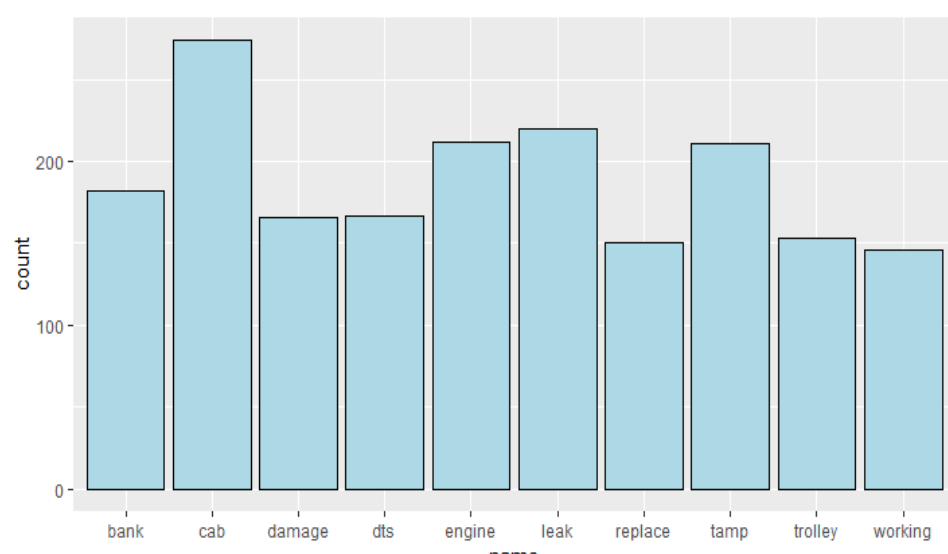
```
#Input dictionary
comp_dict <- c('wire', 'engine', 'tamp', 'tamp bank', 'cab b', 'dts', 'oil', 'machine',
'pump drive', 'pump drive gearbox', 'lining wire', 'hose', 'yoyo wire',
'auto greaser', 'cab b heater', 'seals', 'pick arm seal', 'engine radiator',
'drp trolley', 'dts rerailing arm', 'dts unit', 'dts engine', 'door fly',
'rail sensor', 'seat arm rest', 'filler cap', 'fire system', 'strip light',
'tail lamp', 'measuring trolley', 'arm rest', 'alternator belts', 'm', 'm',
'dts unit', 'cab heater', 'tamping tine', 'roller clamp', 'pressure hose',
'fly', 'hydraulic squeeze pressure hose')

defect_dict <- c('leak', 'fray', 'missing', 'damage', 'replace', 'worn', 'break', 'refuel',
'topped', 'snap', 'open', 'fault', 'twist', 'strip', 'chafe')
```

N-gram Modelling + Keyword Building



Word Cloud for 1-gram Modelling



Frequency Bar Plot for 1-gram Modelling

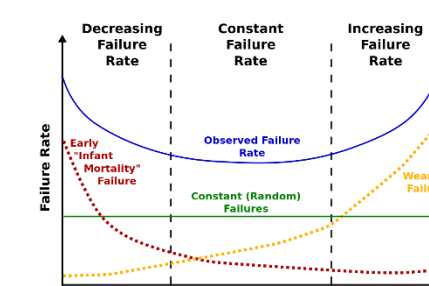
### Application of Research

- CBM programme consist of mapping condition-monitored data with event data.
- Cleaning and processing of text data is critical for a comprehensive CBM programme.

Asset	Description	Reported Date
DR73116	R/H rear full tamping tine damaged	22/04/2019
DR73114	Washer missing from tamping tine middle inner LHS	16/04/2019
DR73114	RHS middle inner tamping tine damaged	09/04/2019

Description	Component	Defect
R/H rear full tamping <u>tine damaged</u>	tine	damage
<u>Washer missing</u> from tamping tine middle inner LHS	washer	missing
RHS middle inner tamping <u>tine damaged</u>	tine	damage

### Findings



### Time Series Fault Distribution

Group	Asset Number	Type	Year Introduced
1	DR 73111	09-3X Dynamic [WA 4605]	2005
	DR 73113	09-3X Dynamic [WA 4607]	2005
	DR 73114	09-3X Dynamic [WA 4857]	2006
	DR 73115	09-3X Dynamic [WA 5008]	2007
2	DR 73116	09-3X Dynamic [WA 5525]	2010
	DR 73117	09-3X Dynamic [WA 5526]	2010
	DR 73118	09-3X Dynamic [WA 5527]	2010
3	DR 73120	09-3X Dynamic [WA 6413]	2016
	DR 73121	09-2X Dynamic [WA 6414]	2016
	DR 73122	09-2X Dynamic [WA 6415]	2016

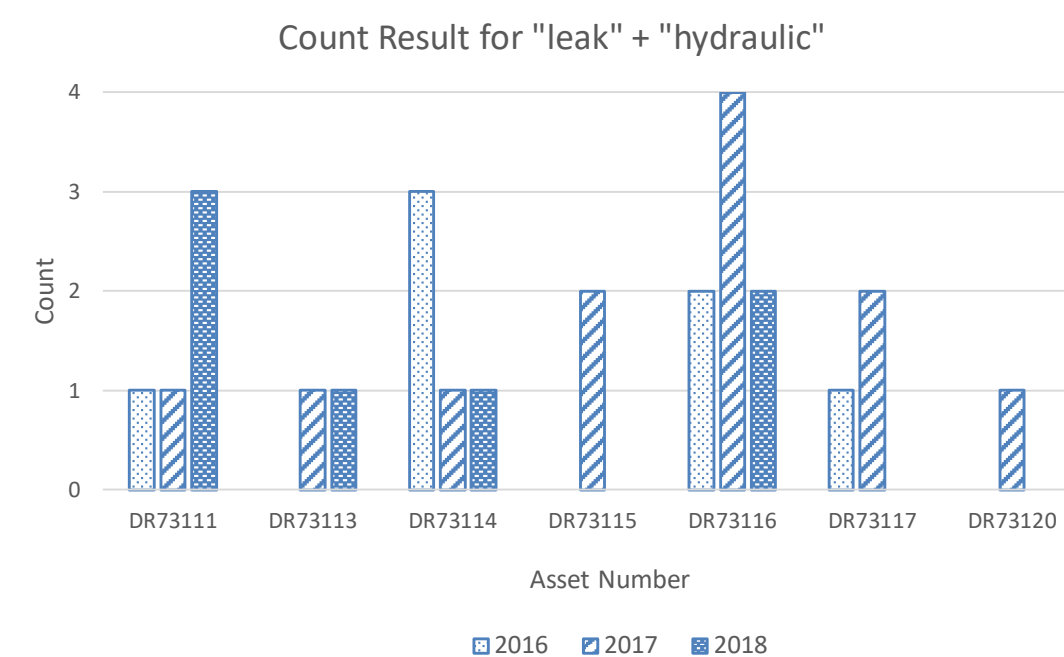
Group	Asset Number	Average Age, n	Average Fault per Quarter	Mean, $\mu$	Standard Deviation, $\sigma$
1	DR 73111	13	12.6	11.1	1.1
	DR 73113		11.5		
	DR 73114		10.2		
	DR 73115		10.0		
2	DR 73116	9	11.4	9.1	2.8
	DR 73117		10.7		
	DR 73118		5.1		
3	DR 73120	3	9.3	5.7	2.6
	DR 73121		3.7		
	DR 73122		4.0		

### Deep-dive into Event Data: “leak” + “hydraulic”

Top three defects

- leak” (186 observations)
- “damage” (118 observations)
- “missing” (93 observations)

Hydraulic System account for 26% of all “leak” entries.



### Recommendation to Industry

Existing Situations	Recommendations
1. Multiple WO entries of same event.	1. Review the process for creating a new Work Order. Ideally, there should be a one-to-one mapping from the event to a work order.
Event Data	
1. Use of short forms. Standardisation of entries. E.g. Ratification in fault description. Multiple names of same component. E.g. Radiator & heat exchanger.	1. Use of drop-down list for data entries. One category per fault. One category per component. If the component has two faults, repeat entry.

NOVEMBER 2019

