

Processing of Unstructured Short Text in Maintenance Records

**NIST Standards Requirements Workshop for Natural Language Analysis
May 2018**

Yiyang Gao, Caitlin Woods, Melinda Hodkiewicz, Tim French, Wei Liu
University of Western Australia – BHP Fellowship funded Siri for Maintenance Project
ARC ITTC for Transforming Maintenance through Data Science

Motivation

- Thousands of maintenance tasks are performed a year in most asset intensive industries
- What was done to assets including when, where, why and how is captured in short unstructured text.
- The knowledge in this text is vital to improving maintenance efficiency and asset reliability.

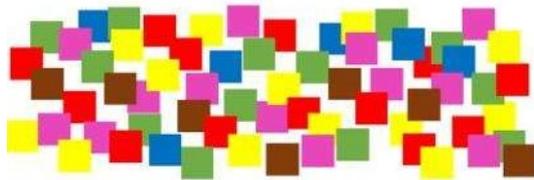


Image from: <https://www.oceanasolutions.org/the-importance-of-mechanical-maintenance/>

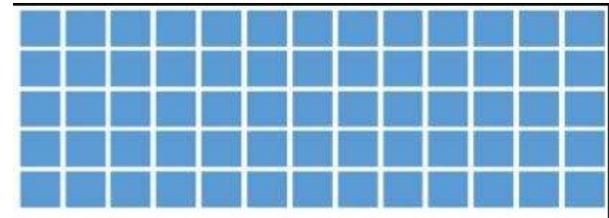
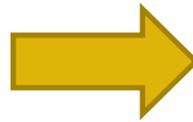
Problem

- Time Consuming
- Inconsistent
- Domain knowledge required

Can this process be semi-automated?

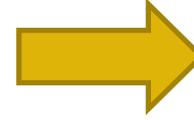


Unstructured maintenance records



Structured machine readable
maintenance records

Example of maintenance record text



Technicians Perform
Maintenance Task

Technicians Fill in
Comment

Reliability Engineer
Perform Analysis

Short Text

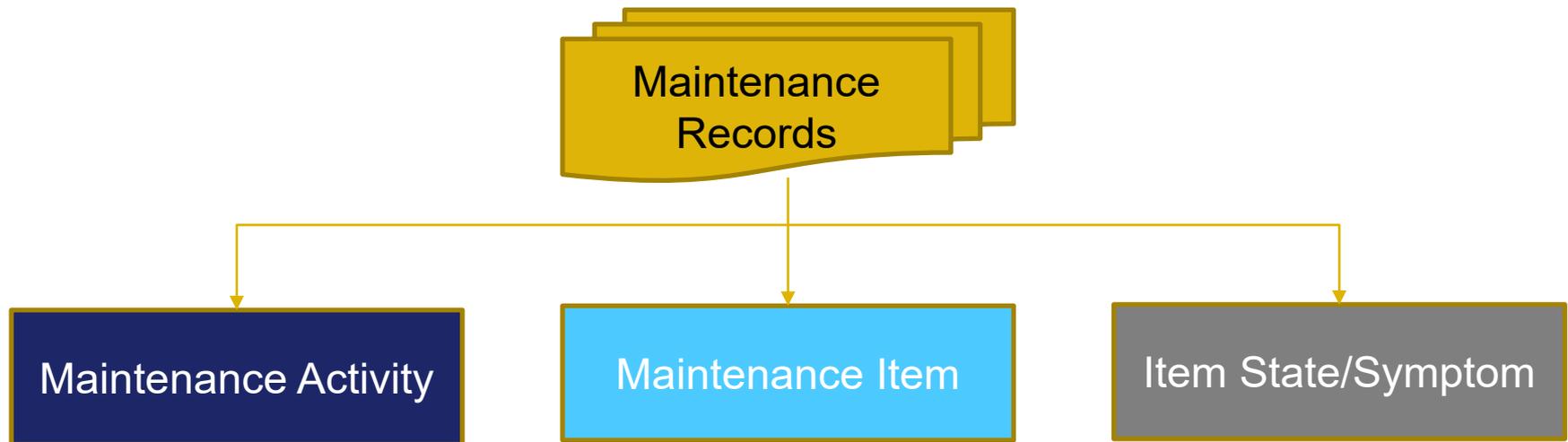
- replace t/con pump
- positioner arm out of alignment
- repair broken lift cylinders

Action, Item

Item, Item symptom

Action, Item, Item state

A pipeline for semi-automatic processing of unstructured short text into a structured format



Experimental Setup

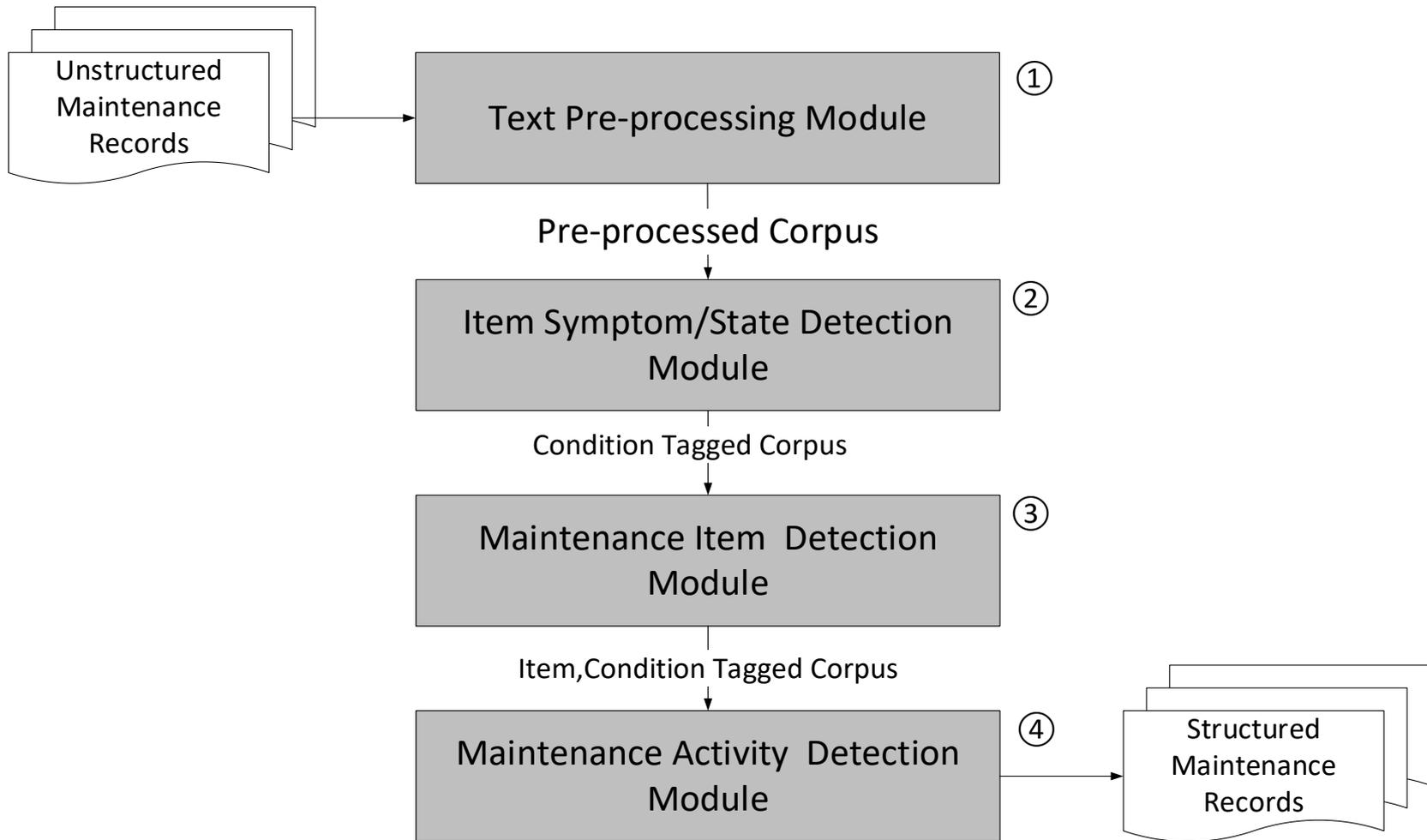
- Approx 700,000 historical maintenance records
- Domain → Maintenance of Heavy Mobile Equipment



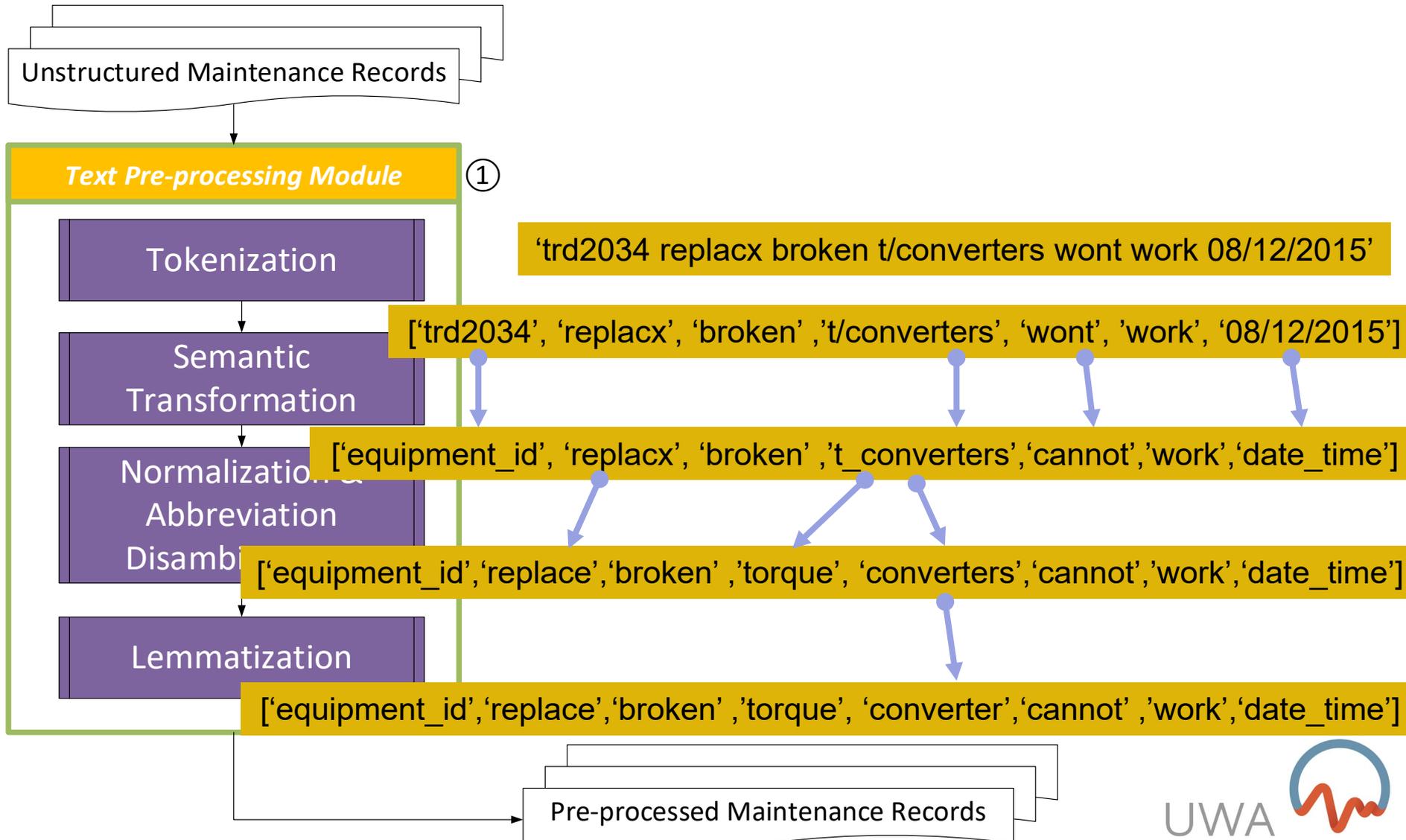
WOType	PONumber	BscStartDate	ActStartDate	ShortText	WorkCentre	FuncLoc	FuncLocDesc	UserStatu	SystemSta	Actual
PM01	1	10/07/2005	10/07/2005	replace lube tank fill hose.	WC1	FL1	grease lube system	Removed	Removed	651
PM01	2	10/07/2005	10/07/2005	checked rotary head alignment	WC1	FL2	mast	Removed	Removed	434
PM01	3	10/07/2005	10/07/2005	reweld shock sub to rotary head	WC1	FL3	subs/adaptors	Removed	Removed	361
PM01	4	27/07/2005	28/07/2005	centre tank not filling with water	WC1	FL4	water injection system	Removed	Removed	144
PM01	5	2/07/2001	2/07/2001	hydraulic pump failure	WC2	FL5	implement pump	Removed	Removed	NA
PM01	6	1/07/2001	2/07/2001	pump drive box coupling and shaft u/s.	WC2	FL6	#NAME?	Removed	Removed	172
PM01	7	11/07/2005	0/1/00	repair engine excessive blowby	WC1	FL7	engine	Removed	Removed	0
PM01	8	16/07/2005	0/1/00	two broken growser bolts r/h side	WC1	FL8	track - right	Removed	Removed	19
PM01	9	27/07/2005	0/1/00	adjust all head slides	WC1	FL9	rotary guide frame	Removed	Removed	0
PM01	10	27/07/2005	28/07/2005	change out main gearbox oil to 85/140	WC1	FL10	oil - pump drive gearbox	Removed	Removed	72
PM01	11	27/07/2005	28/07/2005	change out both finaldrive oil to 85/140	WC1	FL11	final drive assemblies	Removed	Removed	144
PM01	12	27/07/2005	28/07/2005	change out rotary head gearbox oil	WC1	FL12	oil - rotary gearcase	Removed	Removed	72

Data of interest: short text in maintenance records

Proposed Pipeline



Text Pre-processing Module



Text Pre-processing Result

Task	Size of Vocabulary
Tokenization	54288
Semantic Transformation	20200
Normalization & Abbreviation Disambiguation	12743
Lemmatization	11618

Symptom/State Detection Module

What is bigram for symptom/
state detection?

Pre-processed Maintenance Records

Symptom/State Detection Module ②

Bigram Detection

Keyword based
bigram filtering

Symptom/State
Dictionary Building

Tagging

`['equipment_id', 'replace', 'broken', 'torque', 'converter', 'cannot', 'work', 'date_time']`

`['equipment_id', 'replace']`

`['converter', 'cannot']`

`['replace', 'broken']`

`['cannot', 'work']` ✓

`['broken', 'torque']`

`['work', 'date_time']`

`['torque', 'converter']`

`['pump', 'is', 'broken']`

`['cylinders', 'not', 'lifting']`

`['pump', 'is']`

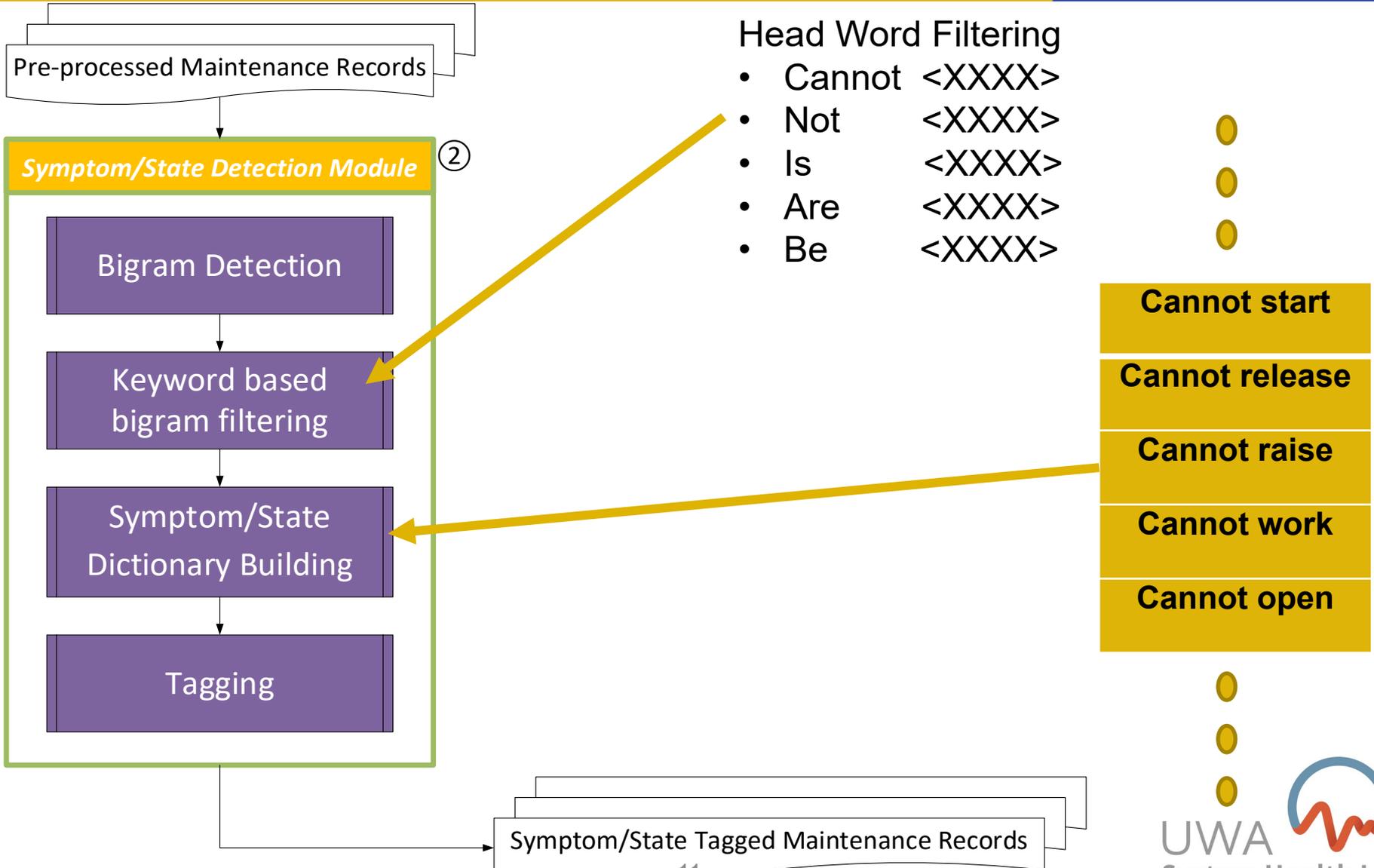
`['cylinders', 'not']`

`['is', 'broken']` ✓

`['not', 'lifting']` ✓

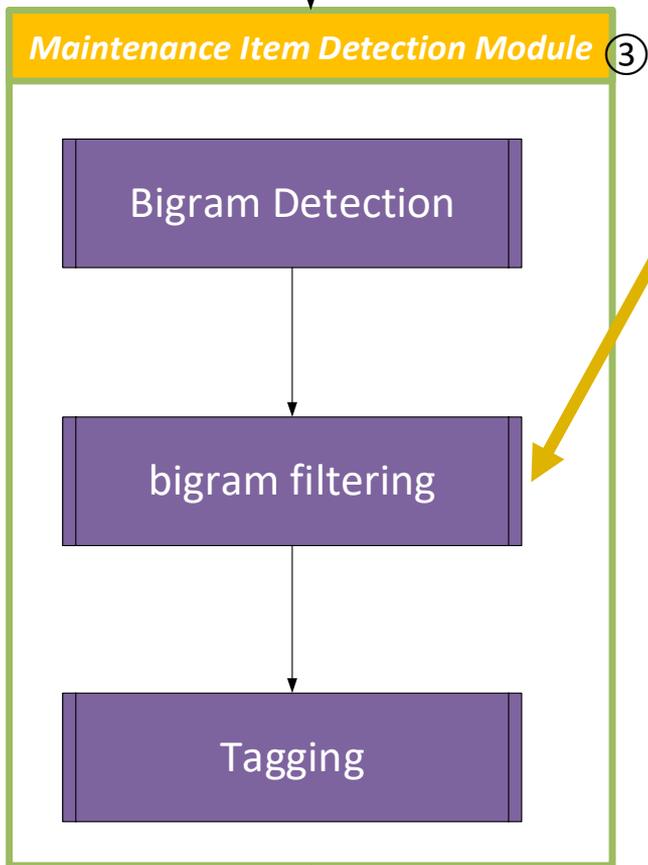
Symptom/State Tagged Maintenance Records

Symptom/State Detection Module



Maintenance Item Detection Module

Symptom/State Tagged Maintenance Records



Bigram that does not contain

1. Symptom/State,
2. Maintenance activity,
3. Common stop words

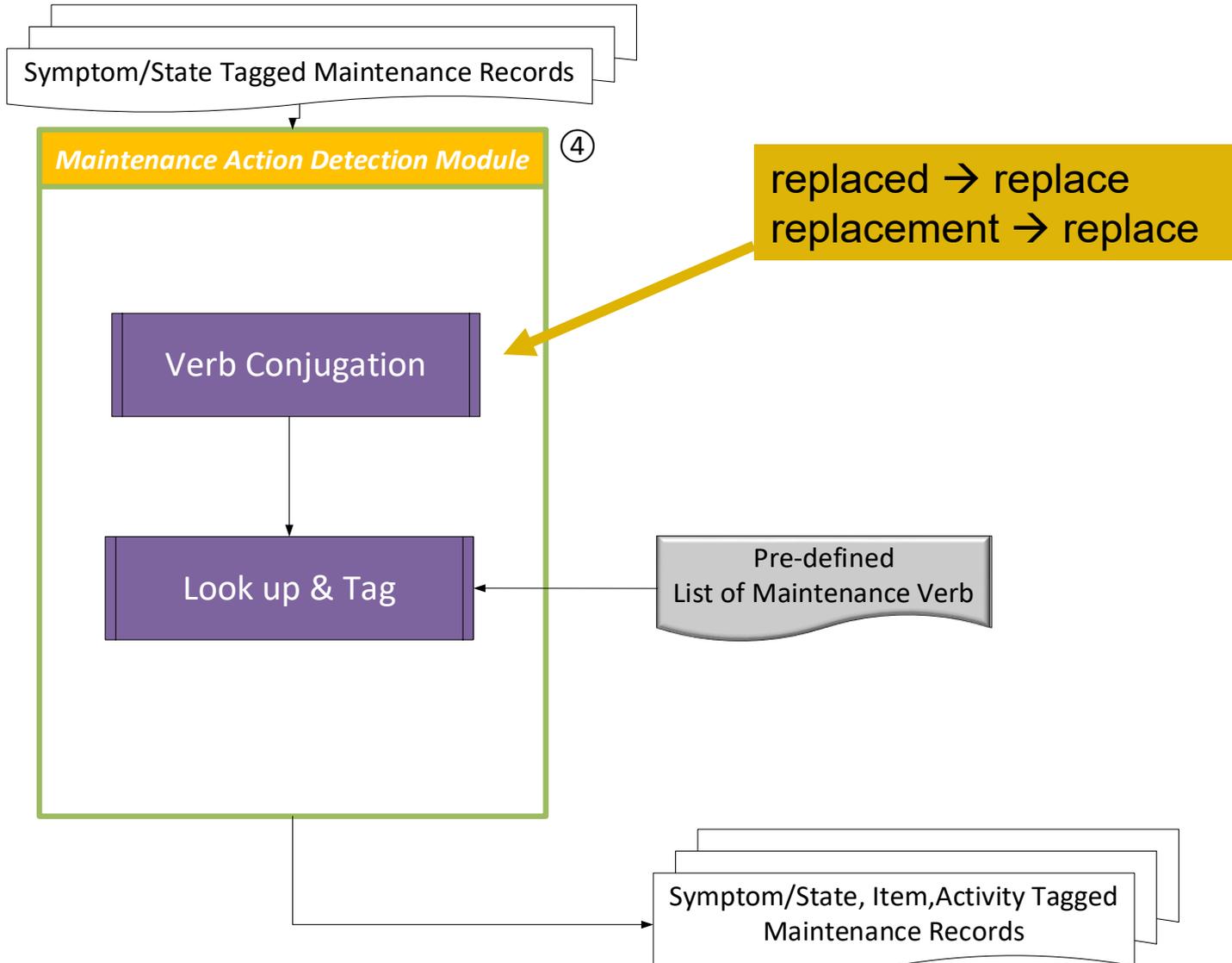
Is a bigram for maintenance item

['equipment_id','replace','broken','torque','converter','cannot','work','date_time']

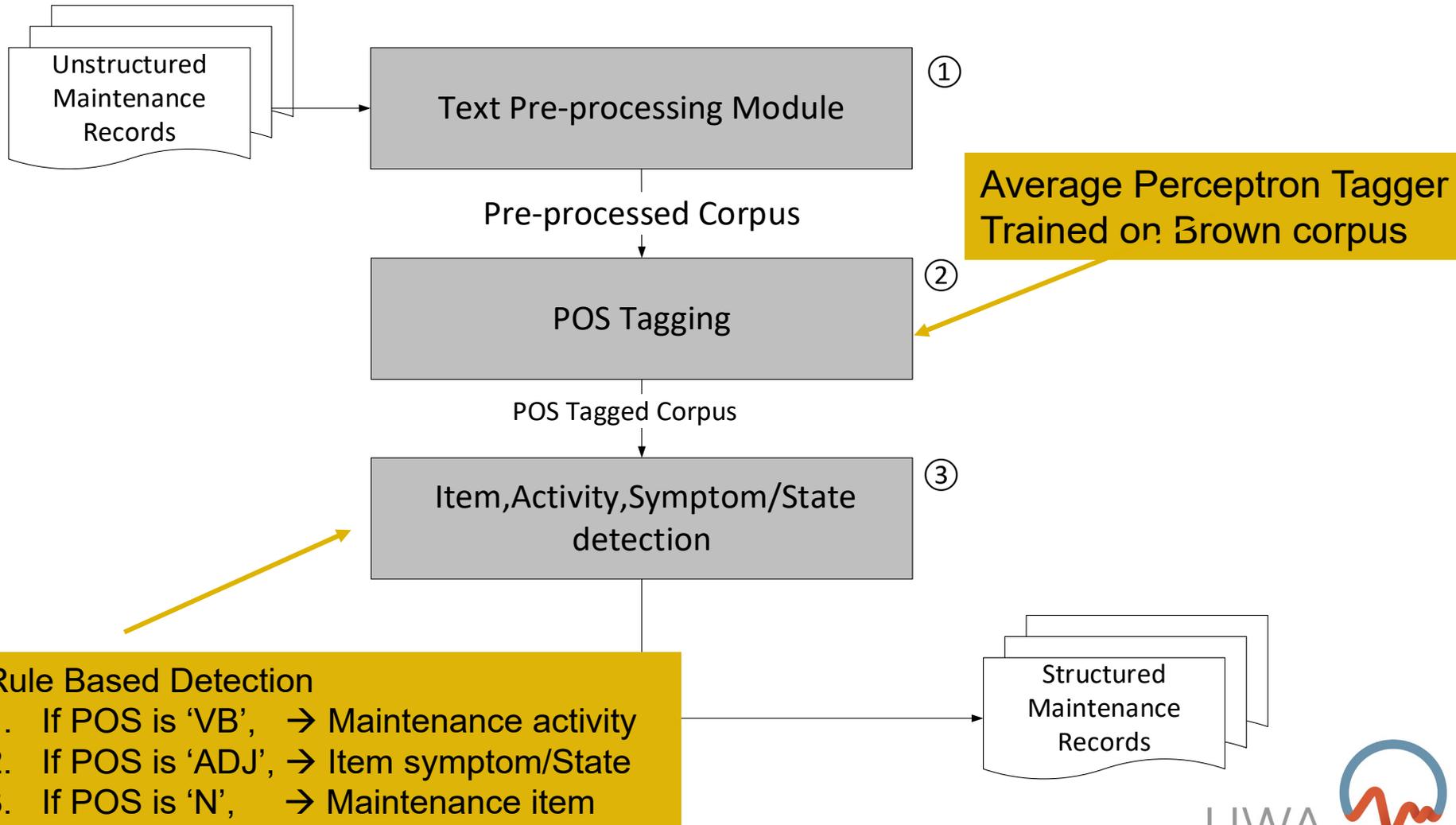
- | | |
|-------------------------------|---------------------------|
| ['equipment_id', 'replace'] ❌ | ['converter', 'cannot'] ❌ |
| ['replace', 'broken'] ❌ | ['cannot', 'work'] ❌ |
| ['broken', 'torque'] ❌ | ['work', 'date_time'] ❌ |
| ['torque', 'converter'] ✅ | |

Symptom/State, Item Tagged Maintenance Records

Maintenance Activity Detection Module



Base Pipeline



Performance Evaluation

Validation Corpus

- Size : 360
- Human Processed
 - Normalization
 - Abbreviation Disambiguation
 - Item symptom/state tagging(#)
 - Maintenance Item tagging(~)
 - Maintenance activity tagging(=)

Jaccard Index for similarity measurement

A: Expert tagged
B: Pipeline tagged

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Raw: 'replacx broken arm rest'

A: ['replace=' , 'broken#' , 'arm~rest']

B: ['replace=' , 'arm~rest']

$$\frac{\text{Number of Common Tokens}}{\text{Total Number of Unique Tokens}}$$

$$\text{Jaccard Index} = \frac{2}{3} = 66\%$$

Result

Task	Proposed Pipeline Result	Base Pipeline Result
Item Symptom/State Detection	63.2%	8.9%
Maintenance Item Detection	54.9%	14.5%
Maintenance Activity Detection	88.9%	24.4%
Overall	56.7%	12.8%

Proposed Pipeline > Base Pipeline

Next steps

Our next steps

- Working on improving the code
- Explore LSTM and other methods
- Work on other corpus
- Building on this work through our new \$8.8M industry and government funded Centre for Transforming Maintenance through Data Science
<https://www.maintenance.org.au/>

Proposed next steps for this community

- Develop shared test sets
- Develop shared dictionaries specifically for maintenance including lists of synonyms, misspellings, etc.
- Share code, papers etc. (e.g. on GitHub)
- Assist each other in peer reviewing code