



## A Dummies Guide to Qualitative and Quantitative Text Analysis of Progress Notes

---

David Hunter

Information Analyst,  
*Townsville University Hospital*





## Pertinent patient information still hidden in progress notes



### Find Notes [Apache Solr](#)

- Author
- Keywords and phrases
- Patient demographics
- Diagnosis

### Parse Notes [Apache cTakes](#)

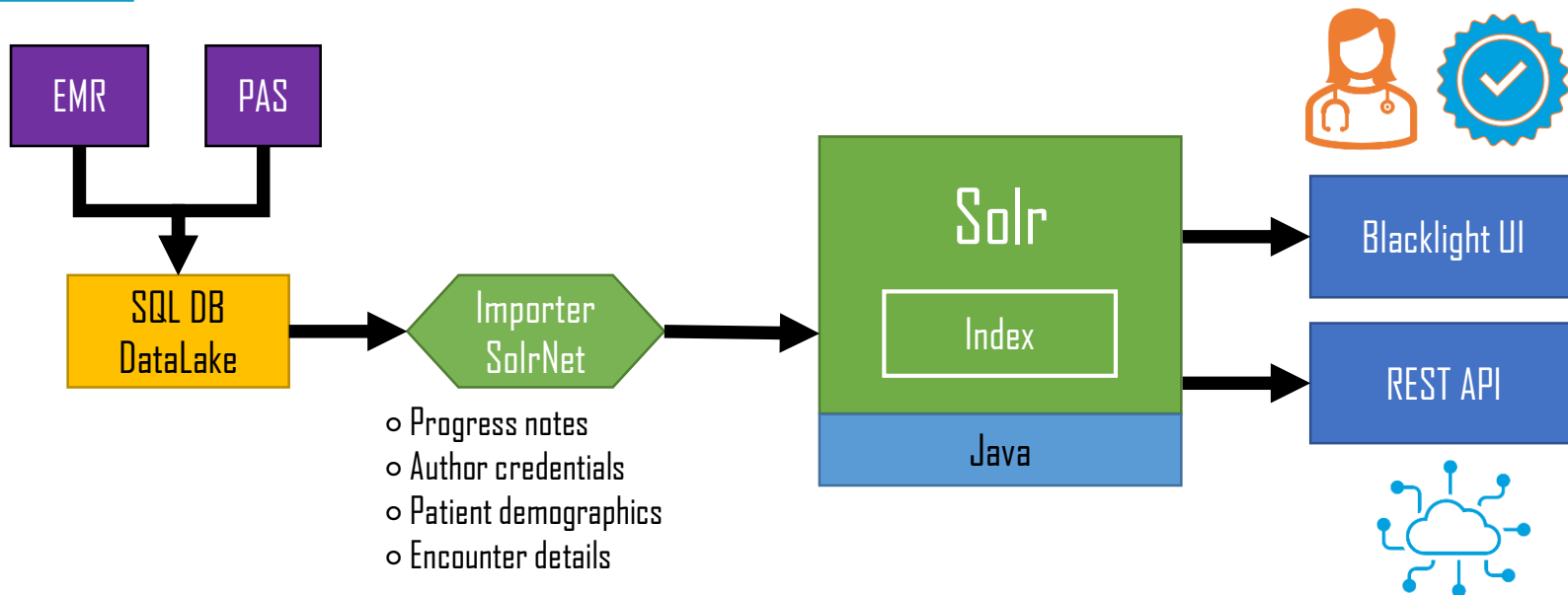
- Medications
- Symptoms
- Diseases

### Text Comparison [FuzzierSharp](#)

- Note subject matching
- GP address validation



## Apache Solr Data Fabric





## Apache Solr Installation

<https://solr.apache.org/>

<https://projectblacklight.org/>

<https://github.com/SolrNet/SolrNet>



- Logout solr
- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump
- disco\_data
- Overview
- Analysis
- Dataimport
- Documents
- Files
- Ping
- Plugins / Stats
- Query**
- Replication
- Schema
- Segments info

Request-Handler (qt)

/select

— common —

q

q

fq

sort

start, rows

0 10

fi

df

Raw Query Parameters

key1=val1&key2=val2

wt

-----

indent off

debugQuery

dismax

edismax

hl

facet

spatial

spellcheck

**Execute Query**



## Blacklight UI

Notes ▼ MATER Search

Limit your search

You searched for: MATER × Patient Age at Note > 21 to 30 years × FIN > 3136649 × Start Over

Performed Date > Within 7 days × Patient Gender > FEMALE ×

1 - 5 of 5 Share Export CSV Sort by Event Date (new to old) ▼ 10 per page ▼

1. Progress Notes Inpatient / Social Work Surg 3 referral received Bookmark

Date and time: 03-Mar-23 03:38PM  
Event Title: Social Work Surg 3 referral received  
Patient Name: ████████████████████  
Patient Gender: FEMALE  
Patient Weight: 110 kg  
FIN: 3136649  
URN: ██████████  
Encounter: Inpatient  
Treating Team: General Surgery  
Attending Clinician: ████████████████████  
Ward: TTH Ward Block Level Two Surgical 2 Ward  
Performed By: ████████████████████  
Notes: : 08/02/23 bariatric surgery at the Mater Since Friday has been feverish, nauseous, intermittent

2. Dietetic Initial Assessment Text / Dietetic Initial Assessment Bookmark

Date and time: 02-Mar-23 11:42AM  
Event Title: Dietetic Initial Assessment  
Patient Name: ████████████████████  
Patient Gender: FEMALE



## Apache Solr API - Python

```
1 import requests
2 import json
3
4 header = { "Authorization": "*****" }
5 url = 'https://*****.thsd.health.qld.gov.au/solr/disco_data/select'
6 params = dict(
7     fl='ENCNTR_ID, EVENT_ID, EVENT_CD, NOTE_DT_TM, AUTHOR_NAME, NOTE_HTML',
8     q='(NOTE_TEXT:\\"Mater\")== AND (EVENT_CD:22212489 or 9019048) AND NOTE_DT_TM:[NOW-366DAY TO NOW]',
9     rows='25000'
10 )
11
12 resp = requests.get(url, headers=header, params=params)
13 data = resp.json()
14 print (data['response']['numFound'], "documents found.")
15
16 for doc in data['response']['docs']:
17     print (doc['EVENT_ID'])
```



## Solr API - Dashboard example

Notes with "Mater" in text :

Author Credential	Note Author	Note Type	Encounter Type
# Notes	# Encounters	# Patients	
1	1	1	

DocID	Pt URN	Pt Age	Note Type	Note Subject	Au...	Note DtTm	Enc type	ED dep ...	Inpt LOS
1209696690			Assessments Nursing...	TCHS-Veterans' Lisalo...	RN		Emergency	AB L1 TSCDU	1
12093878072			Progress Notes Outpati...	ORTHO FRACTURE CL...	REG		Outpatient	-	-
12093750670			Progress Notes Inpatient	Ortho & OrthoGerI WR	RMO		Emergency	WB L2 TSG1	2
12093481152			Progress Notes Inpatient	AAU PT-AT WR	INT		Emergency	WB L1 TMW3	2
12093442710			Antenatal Booking Deta...	Antenatal Booking Det...	RM		Outpatient	-	-
12094959379			Progress Notes Inpatient	Ortho WR	INT		Emergency	AB L1 TSCDU	3
12092345724			ED Information for Pati...	ED Information for Pati...	INT		Emergency	-	-
12092299886			Progress Notes EmERGE...	ED Review - Intern	INT		Emergency	-	-
12092130301			Antenatal Booking Deta...	Antenatal Booking Det...	RNRM		Inpatient	-	0
12092061442			Progress Notes Inpatient	Surgical 1 - RN - Night ...	RN		Emergency	AB L1 TSCDU	3
12091988552			Progress Notes Inpatient	PCC RN ND	RN		Emergency	NB G TESS	2
12091469197			Progress Notes Inpatient	ENT Review	REG		Emergency	-	-
12091420971			ED Assessment Adult-T...	ED Assessment Adult	RN		Emergency	-	-
12090369348			Progress Notes EmERGE...	ED Review - RMO	RMO		Emergency	-	-
12089767447			Progress Notes Inpatient	Physiotherapy Acute R...	PHYSIO		Emergency	AB L1 TSCDU	3
12089496532			Progress Notes Inpatient	IM reg review	REG		Emergency	WB L1 TMW1	15
12095380247			Corro and Referrals Out...	Urgent - Renal Letter - ...	AO		Outpatient	-	-
12088748183			Progress Notes Inpatient	AAU PT Cons WR	INT		Emergency	WB L1 TMW3	3

Last Reload: [blurred]

Display Note Text - Select DocID

Ortho WR [blurred]

82F  
IHT from Mater with likely haemarthrosis of R knee  
MOI: fall 1/7 prior, on apixaban

**VTE/Anticoagulation: TEDS**  
**ABx: trimethoprim (UTI)**  
**ARP: inactive**

**On Review**  
Going well  
Knee feels a lot better today  
Less pain and swelling  
Feels increased ROM  
Targin helped  
Discussion re previous knee operations, patient confused as to timing of bilateral knee replacements  
Surgical records from Mater obtained this morning indicated R) TKR was done in 2020 - printed and placed in patient's chart  
Encouraged patient to mobilise with PT as much as able

**O/E**  
Temperature 36.5 (07:34)  
Systolic Blood Pressure 167 (07:34)  
Diastolic Blood Pressure 69 (07:34)  
Pulse 78 (07:34)  
SpO2 95 (07:34)  
Respiratory Rate 18 (07:34)

Patient lying in bed, engaging in conversation  
Reduced swelling and bruising compared to yesterday  
Improved ROM 0-60 degrees



## Solr API - Estimated date of discharge

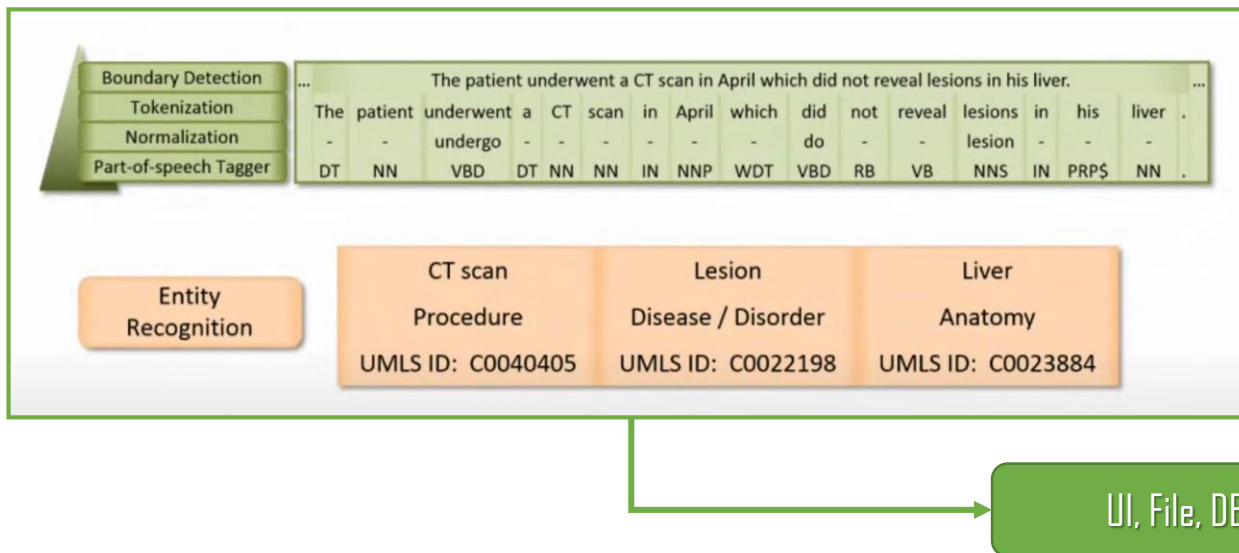
# Patients		Current Ward		Current Inpatients with Inpatient Progress Note authored by SMO/RMO/REG in last 3 days with Discharge, IHT, Transfer, Inter hospital, Home Today, D/C, DC in note text and date indication in same sentence	
54		WB L1 TMW1		Excludes TGSNH and HOMEPB - reloads every 30 mins	
		WB L1 TMW2			
		WB L1 TMW3			
URN	Pt Name	Ward	Bed	Note Date	Note Sentence
		WB L2 TSG2	79		Aim discharge tomorrow
		WB L2 TSG1	14		26/2/23 & dose on D/C was 36 units TWICE a day., pack(s), 0 Refill(s).
		AB G TTCW3	03		- saw a physio on monday 27/02/2023- as physio was percussing on chest had large amount of sputum come up, some difficulty breathing so physio called QAS, waited 4 hours but felt better so DC'ed herself
		AB G TTCW3	03		4. aim d/c monday - as pt not confidence in returning home
		WB L2 TSG1	06		Review tomorrow, then likely for discharge
		WB G TMW5	26		- D6 Mild COVID-19 infection - unable to do discharge as nowhere else to isolate besides hospital
		SAU G TMSAU	45		Can attend IVIG from outpatient DAY LEAVE? Yes with family DISCHARGE PLAN EDD 31/3/23- a/w nursing home availability REFERRALS TO BE COMPLETED Rehab
		SAU G TMSAU	04		Pt NDIS participant and on DSP. Pt may require plan review dependant upon current support needs for discharge. SW will liaise with support co-ordinator r.e. current supports. Support Co-ordinator is
		SAU G TMSAU	14		1 male who was admitted fall , he resides with his wife awaiting weight bearing status to commence discharge plan NEUROPSYCH PSYCHOLOGY WEEKLY REHAB GOAL SUMMARY Potential FM 7/3 - 9am
		WB L2 TSG1	87		Discharge likely for Mon/Tues
		WB L2 TSG1	87		Discharge likely for Mon/Tues
		SAU G TMSAU	25		- May need long-stay supports for discharge
		WB G TONC	15		Completed MASS application - need to give to D/c co-ordinator tomorrow
		AB L1 TMIDW	12		3. Transfer to med 5 when completed isolation
		SAU G TMSAU	10		x2 assist step transfer by 10/3/23 OT PLOF:
		SAU G TMSAU	12		1. Accepted by Rehab - for SACU transfer today at 1:00 PM, thank you
		WB G TMW5S	21		D/C Planning - aim home on Monday if remains well
		SAU G TMSAU	07		- RFDS DAY LEAVE? Wheelchair with family/friends, able to car transfer DISCHARGE PLAN EDD 3/3/23 REFERRALS TO BE COMPLETED
		WB L1 TMW3S	40		ED have informed me that transfer of patient.
		NB L1 TCCU	31		Reports after her initial discharge (10/2/23) had terrible diarrhoea for about 4 days
		WB L2 TSG1	82		Aim DC tomorrow
		WB L2 TSG1	82		- Note plan for DC tomorrow.





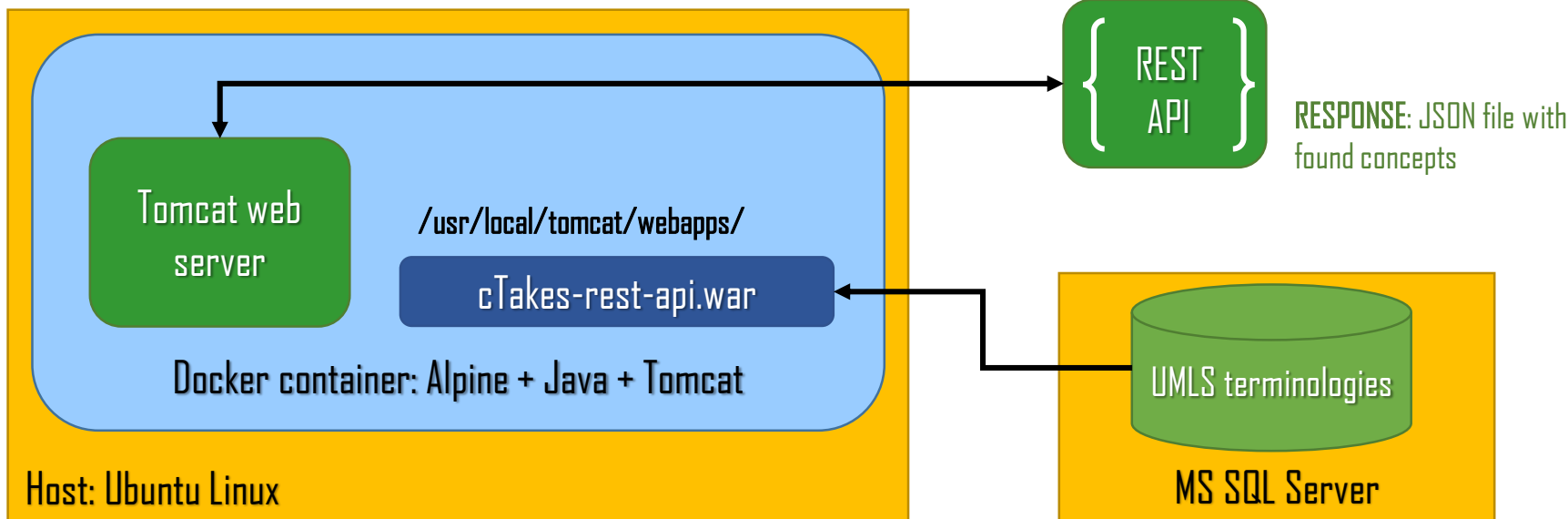
## Apache cTakes

- ✓ Natural language processing system (NLP)
- ✓ Extracts clinical information from EMR notes





## Apache cTakes - Architecture at TUH





## Apache cTakes 4.0 Installation

<https://ctakes.apache.org/>

### Download cTAKES™

Current Version: 4.0.0.1    Released: January 20, 2021    ( [Release Notes](#) / [SVN Tag](#) )

#### System Requirements:

- All Users:
  - 64 bit hardware
  - Any major 64 bit operating system
  - 1 GB Disk Space
  - Oracle® Java™ 1.8 or higher
- Dictionary Users:
  - UMLS® license
  - [Specification of license](#) with UMLS API Key parameter [Piper specification](#) or cli parameter.
  - Network Connection for license verification or custom dictionary.
- Developers:
  - [Apache Subversion™ \(SVN\)](#) Client or equivalent [IDE](#) plugin
  - [Apache Maven™](#) or equivalent [IDE](#) plugin
  - Oracle® Java™ 1.8 or higher
  - 6 GB disk space



## Apache cTakes 4.0 Installation at TUH

---

- Based on step-by-step instructions here:

<https://github.com/GoTeamEpsilon/ctakes-rest-service>

- Optional edit files:

`\dictionary\lookup\fast\customDictionary.xml`

`\resources\pipers\Default.piper`

`\resources\org\apache\ctakes\ytex\uima\annotators\negex_triggers.txt`

- Compile using Maven:

`ctakes-web-rest\target\ctakes-web-rest.war`



## cTakes REST API

POST Progress Note plain text in Body

Response as JSON

The screenshot shows a Postman interface for a REST client. The request is a POST to `http://localhost:8080/ctakes-web-rest/service/analyze?pipeline=Default`. The body is plain text containing a medical progress note. The response is a JSON object with a `"DiseaseDisorderMention"` property.

```
1 Dr. Nutritious Medical Nutrition Therapy for Hyperlipidemia Referral from: Julie Tester, RD, LD, CNSD Phone contact: (555) 555-1212 Height: 144 cm
2 Current Weight: 45 kg Date of current weight: 02-29-2001 Admit Weight: 53 kg BMI: 18 kg/m2 Diet: General Daily Calorie needs (kcal): 1500 calories,
3 assessed as HB + 20% for activity.
4 Daily Protein needs: 40 grams, assessed as 1.0 g/kg. Pt has been on a 3-day calorie count and has had an average intake of 1100 calories.
5 Nil benadryl cough mixture
6 No alcohol
7 She was instructed to drink 2-3 cans of liquid supplement to help promote weight gain. She agrees with the plan and has my number for further
```

```
481 "DiseaseDisorderMention": {
482   "begin": 45,
483   "end": 59,
484   "text": "Hyperlipidemia",
485   "polarity": 0,
486   "conceptAttributes": [
487     {
488       "code": "55022004",
489       "cul": "C0020473",
490       "codingScheme": "SNOMEDCT_US",
491       "tui": "T047"
492     }
493   ],
494   "type": "DiseaseDisorderMention"
495 }
496 ]
497 }
```



## cTakes API

Browser address bar: <https://.health.qld.gov.au/api/cTakes?EventID=0x02000000F0227F45FD3398D9BE5CF...>

**Cardiology Ward Round**

Treating consultant: [REDACTED]

**High Risk Medications**

Antiplatelets: DAPT w/ Ticagrelor 90mg BD

Anticoagulants:

Antiglycemics:-

Antibiotics:

**On Review:**

Lying in bed

Feels well

Nil chest pain

Nil SOB

Nil palpitations

Partner by bedside with no further concerns

**Examination:**

GCS 15, lying in bed, nil distress

**Impression:**

S2M

Inferior STEMI requiring ongoing monitoring

Nil concerns post PCI

**Plan:**

D/C tomorrow if no concerns

Scripts completed - DAPT for 12 months then aspirin indefinitely

Charted Ramipiril 1.25mg PO mane, Bisoprolol 1.25mg mane

Text	Negated	Concept	Terminology	Code
Antibiotics		MEDICATION	SNOMED CT (US)	255631004
aspirin		MEDICATION	RxNorm	1191
Bisoprolol		MEDICATION	SNOMED CT (US)	108547003
Ticagrelor		MEDICATION	RxNorm	1116632
chest pain	X	SYMPTOM	SNOMED CT (US)	29857009
distress	X	SYMPTOM	SNOMED CT (US)	69328002
High Risk		SYMPTOM	SNOMED CT (US)	723509005
Lying		SYMPTOM	SNOMED CT (US)	255576009
pain	X	SYMPTOM	SNOMED CT (US)	22253000
palpitations	X	SYMPTOM	ICD10-CM	R00.2
Present		SYMPTOM	SNOMED CT (US)	52101004
Risk		SYMPTOM	SNOMED CT (US)	30207005
Anticoagulants		DISEASE	SNOMED CT (US)	67438002
chest	X	ANATOMY	SNOMED CT (US)	261179002



## Fuzzy String Matching

---

- Python: TheFuzz
- C#: FuzzierSharp

Find if strings match, or get closest match, or account for spelling mistakes

Not all algorithms may convert to lower case or remove punctuation first

<https://www.analyticsvidhya.com/blog/2021/07/fuzzy-string-matching-a-hands-on-guide/>



## GP Address Example

### ED Information for Patients GP Letters

STS SENT MESSAGE LOGS

STS ADDRESS BOOK



Select Date:

2/3/2023 21:42	DR TAIHS TAIHS, PO BOX 7534, GARBUTT 4814
2/3/2023 21:40	DR [REDACTED] AVENUES FAMILY M/C, 98/185 KERN BROS DRIVE, KIRWAN 4817
2/3/2023 21:36	DR [REDACTED] STRIVE HEALTH & PHYSIOTHERAPY, 703-705 ROSS RIVER ROAD, KIRWAN 4817
2/3/2023 21:35	DR TAIHS TAIHS, PO BOX 7534, GARBUTT 4814
2/3/2023 21:33	DR NORTHERN BEACHES GP SCLINIC NORTHERN BEACHES GP S/CLINIC, 12 DEERAGUN ROAD, DEERAGUN 4818
2/3/2023 21:33	DR NIL
	DR [REDACTED] UPPER ROSS M/C, 1199 RIVERWAY DRIVE, RASKILUSSEN 4815
	DR [REDACTED] CRANBROOK MEDICAL, 403 ROSS RIVER ROAD, CRANBROOK 4814
	DR [REDACTED] SMARTCLINICS ANNANDALE, 152 MARABOU DRIVE, ANNANDALE 4814
	DR [REDACTED] TAIHS, PO BOX 7534, TOWNSVILLE 4810
	DR TAIHS TAIHS, PO BOX 7534, GARBUTT 4814
	DR TAIHS TAIHS, PO BOX 7534, GARBUTT 4814
	DR [REDACTED] BLUEWATER MEDICAL PRACTICE BLUEWATER MEDICAL PRACTICE, 2 PURONO PARKWAY, BLUEWATER 4818
2/3/2023 20:47	DR FNQ FAMILY PRACTICE (STOCKLAND FAMILY PRACTICE), PO BOX 131, AITKENVALE 4814
2/3/2023 20:43	DR [REDACTED] JCU HEALTH GENERAL PRACTICE, PO BOX 217, JAMES COOK UNIVERSIT 4811

Suggested address from STS:

SEND LETTER TO CLINIC

Northern Beaches GP Superclinic  
Deeragun 4818 [Health Link](#)

85% match

TOWNSVILLE UNIVERSITY HOSPITAL  
EMERGENCY DEPARTMENT  
100 Angus Smith Drive, Douglas, QLD, 4814  
PH: [REDACTED] FAX: [REDACTED]  
TTH SWITCH: (07) 4433 1111  
Statement of Attendance  
Confidential

URN: [REDACTED]  
DOB: [REDACTED]  
[REDACTED] MOUNT LOW QLD 4818

Dear ,

[REDACTED] attended the Emergency Department at TOWNSVILLE UNIVERSITY HOSPITAL on the [REDACTED]. The presenting problem was Limb/joint pain.

The diagnosis was Closed fracture of fifth metacarpal

Emergency Department Summary:

[REDACTED] and sustained a closed non displaced 5th metacarpal fracture.  
A wrist splint was applied and she will be followed up in the fracture

```
int match_pct = Fuzz.PartialTokenSortRatio(pas_addr, sts_addr);  
if(match_pct >= 80)  
{  
    ...  
}
```





This functionality is amazing. I have been able to screen 97 charts (and still have time for a coffee!) this morning. I would usually quote a week to get through that volume. This has already had real impact for our ability to safety net patients

*Nurse Manager - Clinical Audit Unit*

