Information Extraction Strategy :: LossRun Tabular Format

**Strategy**: follow an increasingly complex approach to find the matching data point: to-the-right, n-down, and n-intersection.

0.- have a complete/configurable dictionary of context words

1.- find context words in the document considering that these words can be split into several rows

2.- use distance to find data points

3.- validate using expected data types

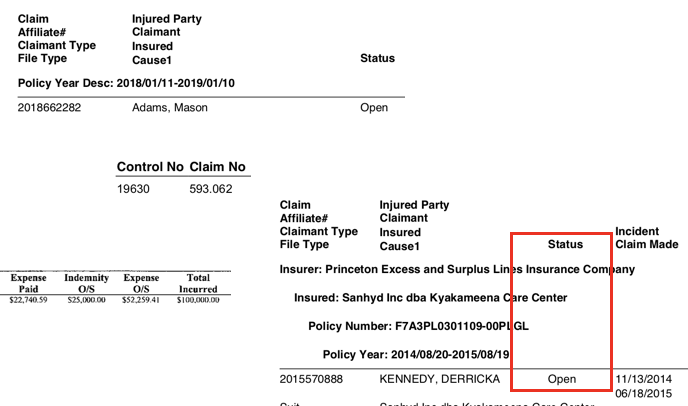
4.- Extend the NPDB strategy to 3 steps:

4.1 if (Find Right (NPDB general logic) == T) then Stop

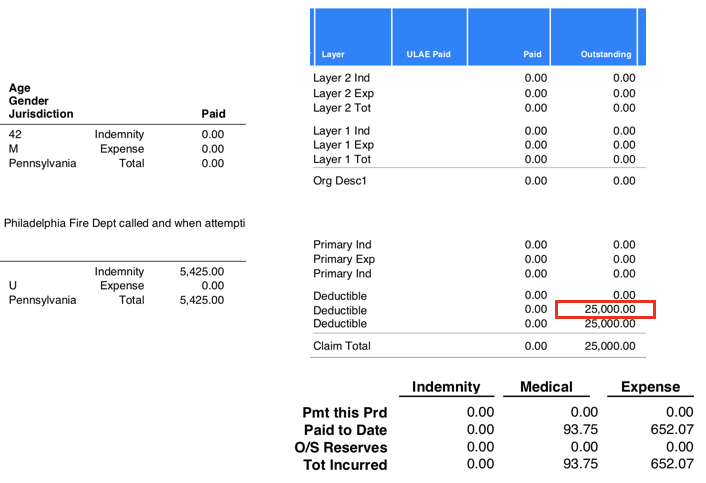




4.2 Else if (Find N-Down) <N = number of rows down> == T) then Stop



4.3 Else (Find N-Intersection) this is a new scenario where the spatial intersection of two groups of context words determine the position of the data point.



5.- For multiple entries in a report vectors of values per data point types are created and are assigned sequentially by claim number.

# TARGET FIELDS

loss-runs

1. Date when the report was run.

2. Dates of the incidents for the claim.

3. Dates when the incidents were reported.

4. Date when it was closed/paid.

5. Amounts paid for each claim.

1. Expenses paid. Amount paid to defend the doctor.

2. Expenses reserved. Amount of money that is needed to close the claim.

3. Indemnity paid. Amount paid in the form of damages to the plaintiff.

4. Indemnity reserved. Amount that the claim's manager thinks needs to be paid in order to close the claim.

5. Total incurred. The total of the above amounts.

6. Status. (open, closed, open pending final payment, reopened)

7. Category.

8. Comments/description.

Nice to have

1. Physician name. (link with opportunity data)

2. Entity name (who they work for).

3. Carrier/payer (who paid for it).

4. Description of the claim.

# PHASE 1

# UNIT TESTING

### 1.1 FIND RIGHT

1.1.1

File: LRT001.pdf

Field: Insured

Expect: D4 Inc

1.1.2

File: LRT001.pdf

Field: As of <Date Run>

Expect: 02/18/2016

1.1.3

File: LRT001.pdf

Field: Incident Desc <Description>

Expect: TREE BRANCH ON INSURED PROPERTY FELL ONTO CLMTS’ CARS IN ADJACENT PARKING LOT / PD CLAIM

1.1.4

File: LRT002.pdf

Field: As of <Date Run>

Expect: 06/03/2015

1.1.5

File: LRT003.pdf - Page 5

Field: Account <Insured>

Expect: 8005109 \*can we work with this number?

1.1.6

File: LRT003.pdf - Page 5

Field: Total Incurred

Expect: 1,089.46

1.1.7

File: LRT004.pdf

Field: Date Reported as of

Expect: January 1, 2020 \* Need to transform this to our standard MM/DD/YYYY

1.1.8

File: LRT004.pdf

Field: Insured

Expect: Mel’s Transportation II, LCC

1.1.9

File: LRT004.pdf

Field: Policy Number

Expect: 001000022377118

1.1.10

File: LRT004.pdf

Field: Status

Expect: Closed

### 1.2 FIND N-DOWN

1.2.1

File: LRT001.pdf

Field: Status

Expect: Closed, Closed, Closed

1.2.2

File: LRT001.pdf

Field: Incident <Incident Date>

Expect: 09/14/2012, 09/14/2012, 05/26/2011

1.2.3 The data is in the same column, but there is a suitable data point that has to be ignored

File: LRT001.pdf

Field: Claim Made <Claim Date>

Expect: 09/14/2012, 09/14/2012, 05/26/2011

1.2.3 The data is in the same column, but there is a suitable data point that has to be ignored

File: LRT001.pdf

Field: Closed <Close Date>

Expect: 02/25/2013, 05/21/2013, 05/21/2012

1.2.4

File: LRT002.pdf

Field: Claim <Claim Number>

Expect: 2013495999

1.2.5

File: LRT003.pdf - Page 5

Field: Claim Number

Expect: PHNH14040799831, PHNH14040799996, PHNH14040799996

1.2.6 Need a mapping table CL - Close to store a standard name in the database

File: LRT003.pdf - Page 5

Field: Status

Expect: CL, CL, CL

1.2.7

File: LRT003.pdf - Page 5

Field: Loss Date

Expect: 04/08/2014, 04/08/2014,04/08/2014

1.2.8

File: LRT003.pdf - Page 5

Field: Loss/Exps Paid

Expect: 500.89, 0.00, 588.57

### 1.3 FIND N-INTERSECTION Notation - Row/Column

1.3.1

File: LRT001.pdf

Field: Indemnity/Paid

Expect: 0.00, 0.00, 0.00

1.3.2

File: LRT001.pdf

Field: Indemnity/Outstanding <Reserved>

Expect: 0.00, 0.00, 0.00

1.3.3

File: LRT001.pdf

Field: Expense/Paid

Expect: 0.00, 0.00, 0.00

1.3.4

File: LRT001.pdf

Field: Expense/Oustanding <Reserved>

Expect: 0.00, 0.00, 0.00

1.3.5

File: LRT001.pdf

Field: Total/Incurred

Expect: 0.00, 0.00, 0.00

1.3.6

File: LRT002.pdf

Field: Indemnity/Paid

Expect: 0.00

1.3.7

File: LRT002.pdf

Field: Indemnity/Outstanding <Reserved>

Expect: 20,000.00

1.3.8

File: LRT002.pdf

Field: Expense/Paid

Expect: 0.00

1.3.9

File: LRT002.pdf

Field: Expense/Oustanding <Reserved>

Expect: 0.00

1.3.10

File: LRT002.pdf

Field: Total/Incurred

Expect: 20,000.00

1.3.10 the row label is different

File: LRT002.pdf

Field: TOTAL/Incurred <Total Sum of Expenses>

Expect: 20,000.00

1.3.11 The order of intersection fields can be reversed

File: LRT004.pdf

Field: Paid to Date/Indemnity

Expect: 0.00

1.3.12

File: LRT004.pdf

Field: O/SReserves/Indemnity

Expect: 0.00

1.3.13

File: LRT004.pdf

Field: Paid to Date/Expense

Expect: 652.07

1.3.14

File: LRT004.pdf

Field: O/SReserves/Expense

Expect: 0.00

1.3.15

File: LRT004.pdf

Field: Tot Incurred/Total

Expect: 745.82