**Introduction**

In Supply-Chain world, a large corporate will be procuring goods and raw materials from hundreds and thousands of vendors. Procurement is done by issuing purchase orders to impaneled vendors. Once the vendor supplies the goods as per the agreed terms and conditions, they send invoices asking for the payment. The buyer has to make a rigorous validation process of these invoices before making the payments. This manual invoice handling processes are time consuming and cumbersome. The industry is exploring the automation possibilities.

Each vendor has his own invoice format and extracting the required data from their invoice file is one of the key challenges to Invoice Processing Automation. You are required to develop a ML Based Invoice format matching module exposed as a REST Service. It takes an input and a target out fields list and tries to map the fields based on ***field name similarity***. The matching algorithms need to consider basic String similarity, Acronyms or short forms (like Date as dt, Invoice as Inv), domain specific synonyms (Vendor === Supplier) and so on.

The API will have two modes of operation

1. Supervised mode
   1. API takes a set of source and target fields and returns the mapping with matching confidence score
   2. The trainer reviews the response and corrects any in-appropriate mappings and resubmits the corrected mapping
   3. The service learns the mapping

.

We iterate this process with 3-4 source format fields for the same target format.

1. Unsupervised mode
   1. We give a new source format fields along with same target format fields
   2. The service should predict the mapping for source fields

**Reference article** : <https://bergvca.github.io/2017/10/14/super-fast-string-matching.html>

Please note that, this article gives a good glimpse of problem space. But solution provided there may not be sufficient to use as is and may need some ML tailoring for our problem statement. Basic String matching is not suffice here. The service should capture the patterns during the training mode (and memorize) and apply that for prediction.

**Sample Data for Test** : Given at the end of this document

**Matching prediction confidence score : It** is in percentage with value ranging between 0 and 100. overAllConfidence is a simple average of all field match confidence.

**API 1 - Matching in Supervised mode**

POST http://localhost:NNNN/train/format/match

{

"source" : {

"formatName": "supplier1Format",

"formatFields": [ "BATCH\_REFERENCE", "TRANSACTION\_REFERENCE",

"BUYER\_NAME", "SUPPLIER\_NAME" ,

"INVOICE\_REFERENCE", "PO\_REFERENCE",

"INVOICE\_DATE", "INVOICE\_AMOUNT", "INVOICE\_CURRENCY",

"MATURITY\_DATE", "PAYMENT\_DATE", "TENOR", "NET\_AMOUNT",

"GROSS\_AMOUNT", "DISCOUNT\_AMOUNT",

"ADJUSTMENT\_AMOUNT", "ADJUSTMENT\_REASON\_CODE",

"PAYMENT\_TERM", "STATUS", "NOTES"

]

},

"target" : {

"formatName": "buyer1Format",

"formatFields": [ "batchReference", "transactionReference", "buyerName",

"supplierName", "invoiceReference", "poReference",

"invoiceDate", "invoiceAmount", "invoiceCurrency", "maturityDate",

"paymentDate", "tenor", "netAmount", "grossAmount",

"discountAmount", "adjustmentAmount",

"adjustmentReasonCode", "paymentTerm", "status", "notes"

]

}

}

**API 1 Response**

{

"sourceformatName": "supplier1Format",

"targetformatName": "buyer1Format",

"overallConfidence": 91,

"mappings": [

{"sourceField" : "BATCH\_REFERENCE","targetField" : "batchReference","confidence" : 93},

{"sourceField" : "TRANSACTION\_REFERENCE","targetField" : "transactionReference", "confidence" : 85},

{"sourceField" : "BUYER\_NAME","targetField" : "buyerName","confidence" : 89},

{"sourceField" : "SUPPLIER\_NAME","targetField" : "supplierName","confidence" : 72},

{"sourceField" : "INVOICE\_REFERENCE","targetField" : "invoiceReference","confidence" : 97},

{"sourceField" : "PO\_REFERENCE","targetField" : "poReference","confidence" : 97},

{"sourceField" : "INVOICE\_DATE","targetField" : "invoiceDate","confidence" : 95},

{"sourceField" : "INVOICE\_AMOUNT","targetField" : "invoiceAmount","confidence" : 82},

{"sourceField" : "INVOICE\_CURRENCY","targetField" : "invoiceCurrency","confidence" : 82},

{"sourceField" : "MATURITY\_DATE","targetField" : "maturityDate","confidence" : 80},

{"sourceField" : "PAYMENT\_DATE","targetField" : "paymentDate","confidence" : 95},

{"sourceField" : "TENOR","targetField" : "tenor","confidence" : 96},

{"sourceField" : "NET\_AMOUNT","targetField" : "netAmount","confidence" : 90},

{"sourceField" : "GROSS\_AMOUNT","targetField" : "grossAmount","confidence" : 78},

{"sourceField" : "DISCOUNT\_AMOUNT","targetField" : "discountAmount","confidence" : 99},

{"sourceField" : "ADJUSTMENT\_AMOUNT","targetField" : "adjustmentAmount","confidence" : 71},

{"sourceField" : "ADJUSTMENT\_REASON\_CODE","targetField" : "adjustmentReasonCode","confidence" : 72},

{"sourceField" : "PAYMENT\_TERM","targetField" : "paymentTerm","confidence" : 82},

{"sourceField" : "STATUS","targetField" : "status","confidence" : 92},

{"sourceField" : "NOTES","targetField" : "notes","confidence" : 99}

]

}

**API 2 - Training the service**

POST http://localhost:NNNN/train/format/learn

{

"source" : {

"formatName": "supplier1Format",

"formatFields": [ "BATCH\_REFERENCE", "TRANSACTION\_REFERENCE",

"BUYER\_NAME", "SUPPLIER\_NAME" ,

"INVOICE\_REFERENCE", "PO\_REFERENCE",

"INVOICE\_DATE", "INVOICE\_AMOUNT", "INVOICE\_CURRENCY",

"MATURITY\_DATE", "PAYMENT\_DATE", "TENOR", "NET\_AMOUNT",

"GROSS\_AMOUNT", "DISCOUNT\_AMOUNT",

"ADJUSTMENT\_AMOUNT", "ADJUSTMENT\_REASON\_CODE",

"PAYMENT\_TERM", "STATUS", "NOTES"

]

},

"target" : {

"formatName": "buyer1Format",

"formatFields": [ "batchReference", "transactionReference", "buyerName",

"supplierName", "invoiceReference", "poReference",

"invoiceDate", "invoiceAmount", "invoiceCurrency", "maturityDate",

"paymentDate", "tenor", "netAmount", "grossAmount",

"discountAmount", "adjustmentAmount",

"adjustmentReasonCode", "paymentTerm", "status", "notes"

]

},

"mappings": [

{"sourceField" : "BATCH\_REFERENCE","targetField" : "batchReference"},

{"sourceField" : "TRANSACTION\_REFERENCE","targetField" : "transactionReference"},

{"sourceField" : "BUYER\_NAME","targetField" : "buyerName"},

{"sourceField" : "SUPPLIER\_NAME","targetField" : "supplierName"},

{"sourceField" : "INVOICE\_REFERENCE","targetField" : "invoiceReference"},

{"sourceField" : "PO\_REFERENCE","targetField" : "poReference"},

{"sourceField" : "INVOICE\_DATE","targetField" : "invoiceDate"},

{"sourceField" : "INVOICE\_AMOUNT","targetField" : "invoiceAmount"},

{"sourceField" : "INVOICE\_CURRENCY","targetField" : "invoiceCurrency"},

{"sourceField" : "MATURITY\_DATE","targetField" : "maturityDate"},

{"sourceField" : "PAYMENT\_DATE","targetField" : "paymentDate"},

{"sourceField" : "TENOR","targetField" : "tenor"},

{"sourceField" : "NET\_AMOUNT","targetField" : "netAmount"},

{"sourceField" : "GROSS\_AMOUNT","targetField" : "grossAmount"},

{"sourceField" : "DISCOUNT\_AMOUNT","targetField" : "discountAmount"},

{"sourceField" : "ADJUSTMENT\_AMOUNT","targetField" : "adjustmentAmount"},

{"sourceField" : "ADJUSTMENT\_REASON\_CODE","targetField" : "adjustmentReasonCode"},

{"sourceField" : "PAYMENT\_TERM","targetField" : "paymentTerm"},

{"sourceField" : "STATUS","targetField" : "status"},

{"sourceField" : "NOTES","targetField" : "notes"}

]

}

**API 2 Response**

{

"sourceformatName": "supplier1Format",

"targetformatName": "buyer1Format",

“Message” : “Learned the mappings”

}

**API 3 - Format matching Automation**

POST http://localhost:NNNN/format/match

{

"source" : {

"formatName": "supplierNFormat",

"formatFields": [ "Batch #", "Transaction #", "Buyer Name", "Supplier Name",

"Inv #", "PO #", "Invoice Date", "Invoice Amount",

"Invoice Currency", "Maturity Date", "Payment Date", "Tenor",

"Net Amount", "Gross Amount", "Discount Amount",

"Adjustment Amount", "Adjustment Reason Code",

"Payment Term", "Status", "Notes"

]

},

"target" : {

"formatName": "buyer1Format",

"formatFields": [ "batchReference", "transactionReference", "buyerName",

"supplierName", "invoiceReference", "poReference",

"invoiceDate", "invoiceAmount", "invoiceCurrency", "maturityDate",

"paymentDate", "tenor", "netAmount", "grossAmount",

"discountAmount", "adjustmentAmount",

"adjustmentReasonCode", "paymentTerm", "status", "notes"

]

}

}

**API 3 Response**

{

"sourceformatName": "supplier1Format",

"targetformatName": "buyer1Format",

"overallConfidence": 100,

"mappings": [

{

"sourceField": "Batch #",

"targetField": "batchReference",

"confidence": 100

},

{

"sourceField": "Transaction #",

"targetField": "transactionReference",

"confidence": 100

},

{

"sourceField": "Buyer Name",

"targetField": "buyerName",

"confidence": 100

},

{

"sourceField": "Supplier Name",

"targetField": "supplierName",

"confidence": 100

},

{

"sourceField": "Inv #",

"targetField": "invoiceReference",

"confidence": 100

},

{

"sourceField": "PO #",

"targetField": "poReference",

"confidence": 100

},

{

"sourceField": "Invoice Date",

"targetField": "invoiceDate",

"confidence": 100

},

{

"sourceField": "Invoice Amount",

"targetField": "invoiceAmount",

"confidence": 100

},

{

"sourceField": "Invoice Currency",

"targetField": "invoiceCurrency",

"confidence": 100

},

{

"sourceField": "Maturity Date",

"targetField": "maturityDate",

"confidence": 100

},

{

"sourceField": "Payment Date",

"targetField": "paymentDate",

"confidence": 100

},

{

"sourceField": "Tenor",

"targetField": "tenor",

"confidence": 100

},

{

"sourceField": "Net Amount",

"targetField": "netAmount",

"confidence": 100

},

{

"sourceField": "Gross Amount",

"targetField": "grossAmount",

"confidence": 100

},

{

"sourceField": "Discount Amount",

"targetField": "discountAmount",

"confidence": 100

},

{

"sourceField": "Adjustment Amount",

"targetField": "adjustmentAmount",

"confidence": 100

},

{

"sourceField": "Adjustment Reason Code",

"targetField": "adjustmentReasonCode",

"confidence": 100

},

{

"sourceField": "Payment Term",

"targetField": "paymentTerm",

"confidence": 100

},

{

"sourceField": "Status",

"targetField": "status",

"confidence": 100

},

{

"sourceField": "Notes",

"targetField": "notes",

"confidence": 100

}

]

}

**Sample Source Format fields**

1. ["BATCH\_REF", "TXN\_REF", "BUYER", "SUPPLIER", "INV\_REF", "PO\_REF", "INV\_DATE", "INV\_AMT", "INV\_CCY", "MAT\_DATE", "PAY\_DATE", "TENOR", "NET\_AMT", "GROSS\_AMT", "DISC\_AMT", "ADJUST\_AMT", "ADJUST\_REASON\_CODE", "PAY\_TERM", "STATUS", "NOTE" ]
2. ["batchNum", "txnNum", "buyer", "supplier", "invNum", "poNum", "invDt", "invAmt", "invCcy", "matDt", "payDueDt", "tenor", "netAmt", "grossAmt", "discAmt", "adjustAmt", "adjustReason", "payTerm", "status", "notes"]
3. ["BatchNumber", "TransactionNumber", "Buyer", "Supplier", "InvoiceNumber", "PONumber", "InvoiceDate", "InvoiceAmount", "InvoiceCurrency", "MaturityDate", "PaymentDate", "Tenor", "NetAmt", "GrossAmt", "DiscountAmt", "AdjustmAmt", "AdjustReasonCode", "PayTerm", "Status", "Notes"]
4. [ "Batch #", "Transaction #", "Buyer Name", "Supplier Name", "Inv #", "PO #",

"Invoice Date", "Invoice Amount", "Invoice Currency", "Maturity Date", "Payment Date", "Tenor", "Net Amount", "Gross Amount", "Discount Amount", "Adjustment Amount", "Adjustment Reason Code", "Payment Term", "Status", "Notes"]

**Target Format fields**

[ "batchReference", "transactionReference", "buyerName", "supplierName",

"invoiceReference", "poReference", "invoiceDate", "invoiceAmount", "invoiceCurrency",

"maturityDate", "paymentDate", "tenor", "netAmount", "grossAmount", "discountAmount",

"adjustmentAmount", "adjustmentReasonCode", "paymentTerm", "status", "notes"]