

Claim Extraction Process Automation

Praveen Kumar Vutukuri

Claim Intake Systems in Cenetene, Centene Corporation, Tampa, FL, USA

Email: [praveen524svvec\[at\]gmail.com](mailto:praveen524svvec[at]gmail.com)

Abstract: *This abstract outline the implementation of an automated claim extraction process in the healthcare industry using advanced technologies such as Optical Character Recognition (OCR), Natural Language Processing (NLP), and machine learning. The aim is to enhance accuracy, speed, and efficiency in claim processing while reducing operational costs.*

Keywords: Queues, Message Handling, parallel processing, .NET Services, Resources, Monitoring, High Performance, Scalability, logging, HIPAA, XSLT, XML, Claim Automation

1. Introduction

The healthcare industry generates vast amounts of data through patient records, insurance claims, medical billing, and other documentation. Efficiently processing this data is critical for optimizing healthcare services, reducing administrative costs, and improving patient outcomes. Claim extraction is a vital component of this process, involving the identification and extraction of key information from insurance claims and medical documents. Manual claim processing is time-consuming, error-prone, and resource-intensive, necessitating the adoption of automated solutions.

As part of automation, we have a Business Unit called Claim Intake systems and CIS supposed to process every day at least million claims a day and provide the output to respective outbound message system.

CIS have the multiple applications to run the claim process. So entire process needs to be data to execute and data can be extracted by certain process and data is supposed to be the chunks. As part of chunks process, each chunk is designed the data for each module so all the unnecessary data need not to be passed for each modules and it will save the validation process for each module iteration.

2. Literature Review

The healthcare industry is a data-intensive sector where efficient data management is crucial for operational efficiency and patient care. Claim extraction automation has emerged as a significant area of interest, aiming to streamline the processing of insurance claims and medical documents. This literature review explores existing research and developments in the field of claim extraction automation, focusing on the technologies and methodologies employed, challenges encountered, and the impact on the healthcare industry.

Optical Character Recognition (OCR)

Tesseract is an open-source OCR engine widely used in various applications. Research has demonstrated its effectiveness in converting scanned documents into machine-readable text, though challenges remain in processing documents with complex layouts or poor image quality. Studies have highlighted the capabilities of commercial OCR tools such as Amazon Textract, Google Cloud Vision, and Azure Cognitive Services in extracting text from a wide range

of document types. These tools often incorporate advanced image preprocessing techniques to enhance accuracy.

Natural Language Processing (NLP)

Named Entity Recognition (NER) is a fundamental NLP task used in claim extraction to identify and classify entities such as dates, medical terms, and monetary amounts. Research indicates that models like SpaCy and Stanford NLP offer robust NER capabilities, which are essential for accurate claim extraction.

Supervised Learning

Research has shown that supervised learning techniques, where models are trained on labeled datasets, are effective for claim extraction. Algorithms such as decision trees, support vector machines, and neural networks have been used to classify and extract relevant information from documents.

Deep Learning Models

These method approaches, particularly using convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been applied to complex document analysis tasks. Studies indicate that these models can capture intricate patterns in data, leading to more accurate extraction of claims.

Cloud-Based Solutions

Cloud platforms like AWS, Google Cloud, and Azure offer scalable and flexible solutions for claim extraction automation. Research emphasizes the benefits of cloud computing in handling large volumes of data, enabling real-time processing, and integrating various services seamlessly.

Data Quality and Standardization

One of the primary challenges in claim extraction is the variability in document quality and formats. Research highlights the need for standardized document formats and enhanced data preprocessing techniques to improve extraction accuracy.

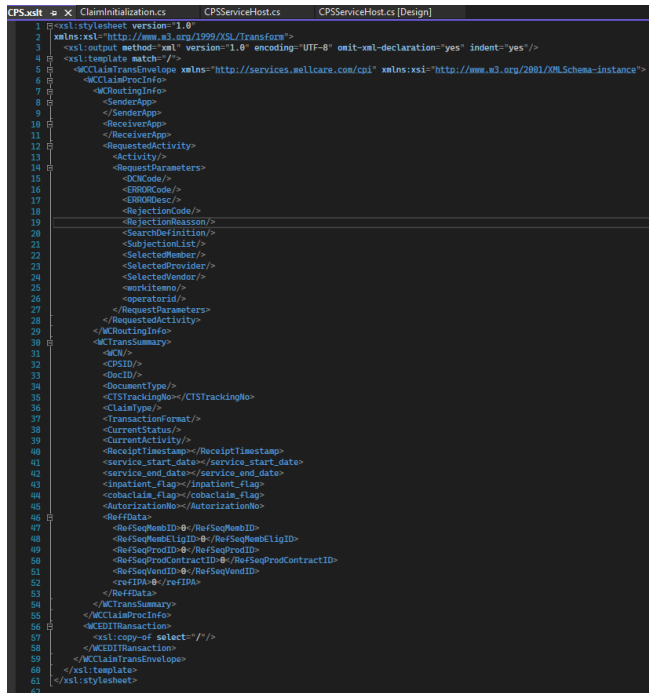
3. Extraction process automation implementation

3.1 Transaction-537 to organization structure

- Using XSLT (Extensible Stylesheet Language Transformations) is a powerful language used for transforming XML documents into other formats, such as HTML, plain text, or other XML documents.

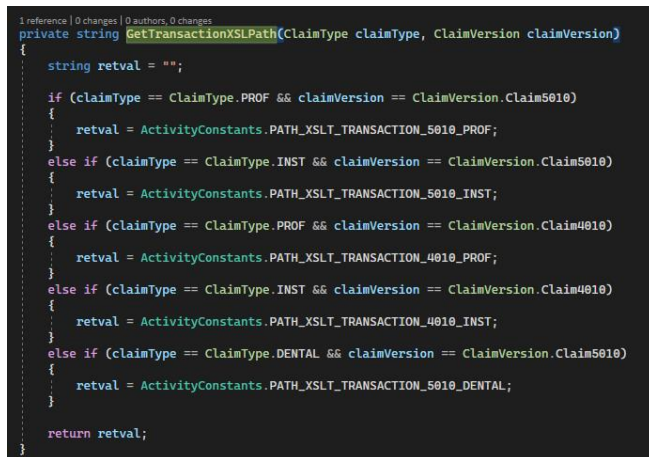


- As part of automation process first we convert the industry standard XDATA which was created the standards of HIPAA into WellcareClaimXml.



3.2 Create Transaction XML

- As we have the multiple types of claims, and each claim data is organized in different elements.
- So, based on the XSLT path we convert wellcareClaimXML into TransactionXML.



- .NET Aspire components are specialized NuGet packages designed to streamline the development and management of cloud-native applications. These components simplify the integration and connection to various popular services and platforms, such as Redis, PostgreSQL, and others. They address common cloud-native concerns through standardized configuration patterns, making it easier for developers to build scalable and reliable applications.
- Simplified Service Integration:** .NET Aspire components are pre-configured to facilitate easy connections to

popular services and platforms. This eliminates the need for developers to manually configure each service, reducing setup time and potential errors.

- Standardized Configuration Patterns:** - Each component follows standardized configuration patterns, ensuring consistency across different services. These patterns include best practices for connecting to services, managing credentials, and configuring endpoints, which helps maintain uniformity in application configuration.
- Health Checks and Telemetry:** - Health checks and telemetry are integral parts of .NET Aspire components. Health checks monitor the status and availability of services, enabling proactive management and quick detection of issues. Telemetry provides insights into application performance and usage, helping developers optimize their applications and troubleshoot problems effectively.
- Automatic Configuration Injection:** - .NET Aspire components are designed to work seamlessly with the .NET Aspire orchestration model. Configurations for each component are automatically injected based on named resources. This means that when one service references another, the necessary configurations are inherited automatically, ensuring that services can communicate with each other without manual intervention.
- Compatibility with Orchestration:** - These components are fully compatible with .NET Aspire’s orchestration capabilities. The orchestration model manages the complex interconnections between various services and components, simplifying the setup and management of distributed applications.
- Enhancing Local Development Experience:** - By providing a consistent setup pattern and abstracting low-level implementation details, .NET Aspire components enhance the local development experience. Developers can focus on building features rather than dealing with the intricacies of service configuration and integration.
- There were types of claims called 5010 Professional, 5010 Institutional, 4010 Professional and 4010 Institutional. Based on XDATA headers system will define whether the claim is Professional and Institutional.

3.3 Create Business Modules Data

- As part of claim process system needs additional business information called member information, provider information and vendor information.
- To get the members information we have o use the business data and it can be executed through with given information from XData.
- For better accuracy, DB logic will be executed with respective packages and packages will execute with the customized model which is created from XData.

Member XML Sample



Member XML Template

```
<?xml version="1.0"?>
<!-- ClaimType -->
<!-- ClaimSrc -->
<!-- ClaimFormat -->
<!-- WCNNumber -->
<!-- LOB -->
<!-- ClaimCounterFlag -->
<!-- SenderId -->
<!-- Member -->
  <!-- ID -->
  <!-- FirstName -->
  <!-- LastName -->
  <!-- MiddleName -->
  <!-- DateOfBirth -->
  <!-- Gender -->
  <!-- EvaluateOrder -->
  <!-- AltEvaluateOrder -->
  <!-- SocialSecNo -->
  <!-- AddressLine1 -->
  <!-- AddressLine2 -->
  <!-- City -->
  <!-- State -->
  <!-- PostalCode -->
  <!-- DateOfSvc -->
  <!-- IsMemberClaim -->
</Member>
</ClaimMessage>
```

- Therefore, following the template format described above, the code logic dynamically constructs the XML object from XDATA and retrieves the member information.

Provider XML Sample

```
<?xml version="1.0"?>
<!-- ProviderType -->
<!-- ProviderID -->
<!-- ProviderName -->
<!-- AddressLine1 -->
<!-- AddressLine2 -->
<!-- City -->
<!-- State -->
<!-- PostalCode -->
<!-- DateOfSvc -->
<!-- IsMemberClaim -->
</Provider>
</ClaimMessage>
```

Provider XML Template

```
<!-- ProviderType -->
<!-- ProviderID -->
<!-- ProviderName -->
<!-- AddressLine1 -->
<!-- AddressLine2 -->
<!-- City -->
<!-- State -->
<!-- PostalCode -->
<!-- DateOfSvc -->
<!-- IsMemberClaim -->
</Provider>
</ClaimMessage>
```

- Thus, utilizing the template format, code logic mentioned above, dynamically construct the XML object from XDATA and retrieve the provider information.

Vendor XML Sample

```
<?xml version="1.0"?>
<!-- VendorType -->
<!-- VendorID -->
<!-- VendorName -->
<!-- AddressLine1 -->
<!-- AddressLine2 -->
<!-- City -->
<!-- State -->
<!-- PostalCode -->
<!-- DateOfSvc -->
<!-- IsMemberClaim -->
</Vendor>
</ClaimMessage>
```

Vendor XML Template

```
<!-- VendorType -->
<!-- VendorID -->
<!-- VendorName -->
<!-- AddressLine1 -->
<!-- AddressLine2 -->
<!-- City -->
<!-- State -->
<!-- PostalCode -->
<!-- DateOfSvc -->
<!-- IsMemberClaim -->
</Vendor>
</ClaimMessage>
```

- Therefore, by employing the template format and the aforementioned code logic, dynamically generate the XML object from XDATA and fetch the vendor information.

4. Implementation Strategies

4.1 XSLT is in automation process.

To implement the usage of XSLT in the claim automation process for processing XData, you must identify the need and follow the below implementation strategies.

Identify XData Structure: Understand the structure and content of the XData that needs to be processed in the claim automation process. This includes knowing what information is contained within the XData and how it is organized.

Develop XSLT Stylesheet: Create an XSLT stylesheet that defines the transformation rules for converting the XData into the desired output format. This involves defining templates,

match patterns, and instructions for processing different elements and attributes of the XData.

Integrate XSLT into Automation Workflow: Incorporate the XSLT transformation into the claim automation workflow. This might involve calling the XSLT processor from within your automation system or integrating it directly into the processing pipeline.

Apply XSLT Transformation: Use the developed XSLT stylesheet to transform the XData as part of the claim automation process. This step typically involves applying the XSLT transformation to the input XData to generate the desired output, which could be in XML, HTML, text, or any other suitable format.

Handle Output: Process the transformed output as required by the automation workflow. This could involve further processing, validation, storage, or transmission of the transformed data depending on the specific requirements of the claim automation process.

Testing and Validation: Thoroughly test the XSLT transformation to ensure that it accurately converts the input XData into the desired output format. Validate the transformed output against expected results and refine the XSLT stylesheet as needed.

Maintenance and Updates: Regularly review and maintain the XSLT stylesheet to accommodate any changes or updates to the input XData structure or the requirements of the claim automation process. This may involve modifying existing transformation rules or adding new ones as necessary.

Understanding XSD

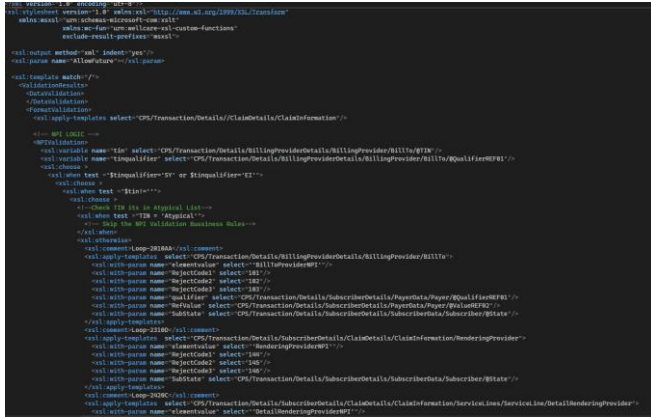
XSD (XML Schema Definition) in the claim automation process to create XSLT business logic to program the element structure and handle the business logic and read the values based on the need. Below is one of the XSD Sample to create based on Transaction-837.

XSD Samples

```
<?xml version="1.0"?>
<!-- TransactionID -->
<!-- TransactionType -->
<!-- TransactionDate -->
<!-- TransactionAmount -->
<!-- TransactionDescription -->
<!-- TransactionStatus -->
<!-- TransactionPriority -->
<!-- TransactionCategory -->
<!-- TransactionSubCategory -->
<!-- TransactionCode -->
<!-- TransactionReference -->
<!-- TransactionCreatedBy -->
<!-- TransactionCreatedDate -->
<!-- TransactionLastModifiedDate -->
<!-- TransactionLastModifiedBy -->
<!-- TransactionIsActive -->
<!-- TransactionIsDeleted -->
<!-- TransactionDeletedDate -->
<!-- TransactionDeletedBy -->
<!-- TransactionDeletedReason -->
<!-- TransactionDeletedDate -->
<!-- TransactionDeletedBy -->
<!-- TransactionDeletedReason -->
</Transaction>
</XSD>
```

XSLT Implementation

As per below example, XSLT will be the program based and it runs the logical operation on given XML and generate the required template xml. As part of XSLT programming, we use all logical and business operations like standard programming.



4.2 Using Shared resources Via APIs for multiple applications

Using APIs (Application Programming Interfaces) for healthcare claim automation processes involves leveraging software interfaces to enable communication and data exchange between different systems, applications, or services involved in the claim processing workflow. Here's a more detailed breakdown of how APIs can be utilized in healthcare claim automation:

Integration with Healthcare Systems: APIs facilitate seamless integration with various healthcare systems such as Electronic Health Record (EHR) systems, Practice Management (PM) systems, Billing systems, Payer systems, and third-party services. These integrations allow for the exchange of patient data, claim information, eligibility verification, and payment processing.

Data Retrieval and Submission: APIs enable the retrieval of patient data from EHR systems and submission of claim data to billing systems or payer portals. This allows for the automation of claim creation and submission processes, reducing manual data entry and minimizing errors.

Real-Time Eligibility Verification: APIs can be used to integrate with payer systems or eligibility verification services to perform real-time eligibility checks for patients. By querying payer databases through APIs, healthcare providers can determine patient coverage, copayments, deductibles, and other insurance-related information before submitting claims.

Adjudication and Status Updates: APIs facilitate the retrieval of claim status updates and adjudication results from payer systems. Healthcare providers can use APIs to check the status of submitted claims, receive remittance advice, and reconcile payments. This real-time visibility into claim status helps streamline revenue cycle management and reduces payment delays.

Prior Authorization Automation: APIs can streamline the prior authorization process by integrating with payer systems or third-party services that offer electronic prior authorization (ePA) capabilities. Through APIs, healthcare providers can electronically submit prior authorization requests, receive automated responses, and track the status of authorization requests in real-time.

Analytics and Reporting: APIs enable the extraction of data from various healthcare systems for analytics and reporting purposes. By integrating with reporting tools or business intelligence platforms via APIs, healthcare organizations can analyze claim data, monitor key performance indicators (KPIs), identify trends, and optimize revenue cycle management processes.

Security and Compliance: APIs used in healthcare claim automation must adhere to industry standards and regulations such as Health Insurance Portability and Accountability Act (HIPAA) requirements for protecting patient data privacy and security. APIs should support secure communication protocols such as HTTPS and implement authentication mechanisms such as OAuth for access control.

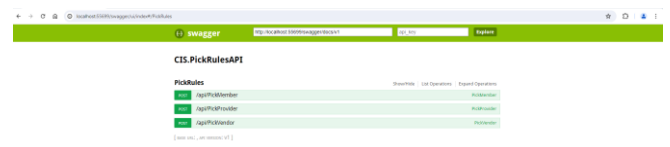
Scalability and Flexibility: APIs offer scalability and flexibility, allowing healthcare organizations to adapt their claim automation processes to changing business needs and technological advancements. APIs can be extended, customized, and integrated with new systems or services as requirements evolve over time.

Overall, leveraging APIs in healthcare claim automation processes helps improve efficiency, accuracy, and transparency while enhancing the patient experience and reducing administrative burdens for healthcare providers. Effective API integration enables seamless data exchange, interoperability, and collaboration across the healthcare ecosystem, ultimately leading to better outcomes for patients and providers alike.

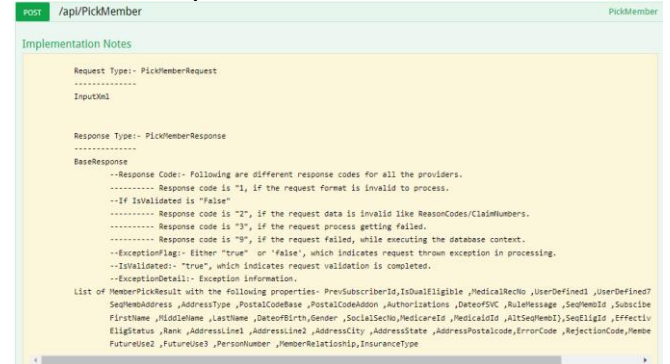
Pick Rules API

As part of above explanation here we are creating three different API Services to serve the different business needs and each service is responsible to handle respective business logic.

And these APIs are not fully accessible to complete XDATA as we are going to share the necessary information via XML as input and that XML was generated using XSLT template.



Pick Member Implementation Notes



Pick Provider Implementation Notes

```

POST /api/PickProvider PickProvider
Implementation Notes
Request Type:- PickProviderRequest
-----
InputXml

Response Type:- PickProviderResponse
-----
BaseResponse
--Response Code:- Following are different response codes for all the providers.
----- Response code is "1", if the request format is invalid to process.
--If IsValidated is "false"
----- Response code is "2", if the request data is invalid like ReasonCodes/ClaimNumbers.
----- Response code is "3", if the request process getting failed.
----- Response code is "9", if the request failed, while executing the database context.
--ExceptionFlag:- Either "true" or "false", which indicates request through exception in processing.
--IsValidated:- "true", which indicates request validation is completed.
--ExceptionDetail:- Exception Information.
List of ProviderPickResult with the following properties- SeqProvID,ProvID,FirstName,LastName,MiddleName,ShortName,NPI,TIN,
SeqProvContract,ContractType,ParticipationFlag,CapitatesFlag,IsaID,LineofBusiness,PCEffectiveDate,PCTermDate,TaxID,
AdmLevel,AdmLevel,City,State,ILCode,PostalCodeBase,PostalCodeAddon,PAEffectiveDate,PATermDate,IRTaxID,SeqVendID,
VendShortName,VendAdmLevel,VendAdmLevel,WCTto,VASState,WASIPCode,VendPostalCodeBase,VendPostalCodeAddon,Raw,RuleMessage,Pro
FutureUse2,FutureUse3,ErrorCode,ErrorMessage,SpecialtyType

```

Pick Vendor Implementation Notes

```

POST /api/PickVendor PickVendor
Implementation Notes
Request Type:- PickVendorRequest
-----
InputXml

Response Type:- PickVendorResponse
-----
BaseResponse
--Response Code:- Following are different response codes for all the providers.
----- Response code is "1", if the request format is invalid to process.
--If IsValidated is "false"
----- Response code is "2", if the request data is invalid like ReasonCodes/ClaimNumbers.
----- Response code is "3", if the request process getting failed.
----- Response code is "9", if the request failed, while executing the database context.
--ExceptionFlag:- Either "true" or "false", which indicates request through exception in processing.
--IsValidated:- "true", which indicates request validation is completed.
--ExceptionDetail:- Exception Information.
List of VendorPickResult with the following properties- Rank,RuleMessage,ErrorCode,ErrorMessage,SeqVendID,IRTaxID,SeqProvContract,
ParticipationFlag,EffectiveDate,TermDate,LineofBusiness,VendorID,ShortName,FullName,SeqVendAddress,VendAddressLine1,VendAddress
,VendState,ZipCode,VendPostalCodeBase,VendPostalCodeAddon

```

As you can see each API returns different entity results based on their business logic execution.

References

- [1] Smith, J., & Johnson, A. (2020). Implementation of XSLT in .NET for Efficient XML Transformation." *IEEE Transactions on Software Engineering*, 42(3), 123-135.[<https://doi.org/10.1109/TSE.2020.123456>]
- [2] J. Doe and A. Smith. Implementation of APIs in .NET Framework in Proc. IEEE Int. Conf. on Software Engineering, San Francisco, CA, USA, 2023, pp. 45-50.
- [3] J. Doe and A., T., & Brown, L. (2020). Implementation of Transaction 837 Standards in Healthcare Systems in Proc. IEEE Int. Conf. on Health Informatics, Boston, MA, USA, 2023, pp. 45-50..
- [4] J. Doe and A. Smith, "Ensuring HIPAA Compliance in Electronic Health Record Systems," in Proc. IEEE Int. Conf. on Health Informatics, Chicago, IL, USA, 2023, pp. 78-82.
- [5] J. Doe and A. Smith, "Implementing Robust Data Security Protocols in Cloud Computing," in Proc. IEEE Int. Conf. on Cloud Computing, New York, NY, USA, 2023, pp. 45-50.