

# An Introduction to Processing Unstructured Medical Data Using Amazon Comprehend Medical

Unstructured clinical documents such as hospital admission notes and patient medical histories contain critical health information necessary for efficient, high-quality care delivery. However, this data can prove difficult to extract in an efficient and meaningful way.

Healthcare organizations generate about 1.2 billion unstructured clinical documents each year. In 2018, Amazon Web Services (AWS) released Amazon Comprehend Medical, a natural language processing (NLP) service makes it easy to use machine learning to extract relevant medical information from unstructured text.

In a recent [HealthITAnalytics.com webcast](#), AWS explained how healthcare providers and health IT companies can use Amazon Comprehend Medical to retrieve unstructured patient data and leverage this information to satisfy a variety of use cases.

Currently, healthcare organizations often process unstructured data through coding.

"If you look at healthcare in general, a lot of retrospectives are done through coding whether it's an ICD-10 code, RX, SNOMED," said Arun Ravi, Senior Product Manager for Amazon Comprehend Medical. "There's a loss of meaning as you go from the unstructured text to actual codes."

Amazon Comprehend Medical is a HIPAA-eligible service that allows users to easily index patient data at the source to create more powerful applications based on accurate information. The service does not store any consumer data that runs through its service.

Providers and developers can use the tool to enhance existing health IT products or assist in building new products more quickly. The tool's capacity to facilitate innovation is key as the healthcare industry transitions to a value-based care system. The service extracts and contextualizes unstructured medical data through a three-step process, Ravi explained.

First, Amazon Comprehend Medical analyzes unstructured clinical documents and extracts data in

five categories: medications, medical conditions, test treatment procedures, anatomical terms, and protected health information (PHI.)

When Amazon Comprehend Medical pulls data about medications, it may extract data about dosages, route modes, strength, and frequency.

"Now, if I give you all these entities at once, it's valuable, but 15 milligrams by itself really doesn't tell you much," explained Ravi. "So we realized we had to take it one step further. The next step that we process through the API is called relationship extraction."

Relationship extraction attaches information within subtypes, such as dosage or frequency information, to its parent category.

"The dosage, router mode, strength, and frequency is actually tied to the medication," explained Ravi. "If there is a dosage and route mode associated with the medication, you'll see it associated to the right parent, and you'll see a relationship score that allows you to see if it's the right relationship for you to take in your downstream applications."

The third step of the process provides contextual information to fill out a complete picture of the extracted data. Contextual data are referred to as traits. One example of a trait is negation.

"Negation could be that a patient denies taking a medication," said Ravi. "That's really important when you're trying to understand what data elements you need to take for your downstream applications," he noted.

Another trait differentiates data that can be classified as a diagnostic sign or symptom. A sign is patient-reported, while a symptom is physician-reported.

"Being able to take different features from existing NLP platforms and put it into a simple API call is unique, and we see this as very powerful and as an enabler for our customers," said Ravi.

In addition to the API that extracts and processes data in the above three steps, Amazon Comprehend Medical also includes an API that extracts PHI. “We don’t only extract the PHI, we actually tell you what the identifier is,” said Ravi.

Amazon Comprehend Medical also includes search capabilities designed to improve usability and efficiency. Users can look up the name of a medication or a specific dosage and view all patients with the specified medication or value in their chart.

“We’re providing enough granularity on the data where you can make those searches very easily,” said Ravi. “The advantage with the deep learning model is that our power and competent medical is we look at context. And because we look at context, we’re able to increase our accuracy to two levels that allow you to do a lot more and essentially create much more accurate downstream applications.”

Established health companies and startups currently use Comprehend Medical during research and health IT development. Fred Hutchinson Cancer Research Center used the service to streamline clinical trial recruitment. Prior to implementing Comprehend Medical, researchers spent a total of 9,600 hours looking for patient participants to fill spots for clinical trials.

With Comprehend Medical, researchers at Fred Hutch can complete the recruitment process in under an hour. Running patient information through Comprehend Medical also enables researchers to apply processed data to different use cases in the future.

“You’ve created an indexed data lake that can power multiple applications,” Ravi said. “Right now it may be clinical trial recruitment for Fred Hutch, but tomorrow it could be something else. Now that the information is structured, they can do a lot more.”

Use cases for Amazon Comprehend Medical vary widely.

“There’s a lot of power with post-market and monitoring and leveraging those type of capabilities for, say, pharmaco-vigilance,” said Dr. Aaron Friedman, Healthcare & Life Sciences Partner Technical Lead at AWS.

PwC is currently applying the solution in its research to monitor and assess pharmaceutical drugs. “They’re working with large pharmaceutical companies to use Comprehend Medical to extract relevant information to look at adverse events and other types of workloads,” Friedman explained.

Ujjwal Ratan, Principal AI and Machine Learning Solutions Architect for Healthcare and Sciences at AWS, demonstrated of how to use Comprehend Medical and the data the technology outputs.

“Having these capabilities minimizes a lot of human intervention--the entire workflow,” said Ratan. “It also gives you the capability to scale this out to hundreds of thousands or even millions of notes.”

“But it’s just a building block,” he noted. “You’ll have to tie it up with upstream and downstream services.”

---

Produced by

**xtelligent**  
HEALTHCARE MEDIA

About Amazon Web Services



For 13 years, Amazon Web Services has been the world’s most comprehensive and broadly adopted cloud platform. AWS offers over 165 fully featured services for compute, storage, databases, networking, analytics, robotics, machine learning and artificial intelligence (AI), Internet of Things (IoT), mobile, security, hybrid, virtual and augmented reality (VR and AR), media, and application development, deployment, and management from 60 Availability Zones (AZs) within 20 geographic regions, spanning the US, Australia, Brazil, Canada, China, France, Germany, India, Ireland, Japan, Korea, Singapore, Sweden, and the UK. Millions of customers—including healthcare providers, healthcare payers, biotechnology and pharmaceutical companies—AWS to power their infrastructure, become more agile, and lower costs. To learn more about AWS, visit [aws.amazon.com/health](https://aws.amazon.com/health).