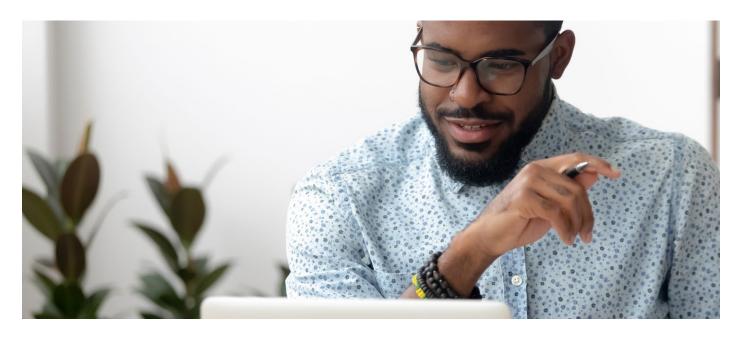


# **OpenText Named Entity Recognition**

Provides a risk reduction service to your end users who need compliance with regulations and corporate mandates, by using named entity extraction and our curated grammars.



#### **Benefits**

- Defensible compliance with information on data sources and match type examples
- Quick to market with readyto-use grammars
- Wide regulation support covering many countries, languages, and entity types

Peak performance is critical for finding and extracting sensitive and high-value entities from structured, semi-structured, and unstructured data—and it's not just the speed of the engine that matters.

The need to support pre-match and post-match operations to perform optimizations such as scope control, deferred validation, and selective early termination for immediate performance gains, and componentization, extraction, and normalization for downstream efficiency.

OpenText™ Named Entity Recognition engine is a flexible, embeddable software component that finds and extracts sensitive and high-value entities from structured, semi-structured, and unstructured data using high performance natural language processing.

It is designed to work well with third-party components using readily consumable APIs and architectural options. When paired with other SDKs—such as OpenText™ File Content Extraction for text extraction—additional cross-component functionality is unlocked, such as table extraction for improved accuracy or pipelining for better performance.

## Defensible compliance

OpenText Named Entity Recognition is designed with grammars that define which entities are of interest. You have full access to OpenText's established

# Additional key benefits

Use OpenText Named Entity Recognition with curated grammars to provide a risk reduction service to your end users who need compliance with regulations and corporate mandates. We provide:

- Added differentiation with both curated grammars and end-user customizable grammars.
- Explainable decisions through use of handcurated grammars rather than corpus derived models.
- Lower hardware costs latency and throughput can be balanced against precision and recall, to meet cost objectives at any scale.
- Reduced ongoing engineering costs as regulations and entity definitions evolve. We ensure that the grammars remain up to date and high performing.
- An experienced team, with the experience necessary to understand how to create the most efficient grammars. With a library of hundreds of grammars available, we have done this before.

expertise in grammar design, using our highly targeted grammar modules, which provide optimized, context-aware, maintained packages for specific geographic and industry-specific entity classes to meet common use cases such as:

- Personal data/Personally identifiable information—including fast and accurate address detection
- Protected health information
- · Payment card industry entities

These grammars allow your product to offer extensive compliance and governance capabilities as a primary use case, straight away and without any up-front grammar development or ongoing engineering cost as they are curated.

#### **Quick to market**

Your product is unique, and OpenText Named Entity Recognition allows you to choose from a variety of form factors, with support for several programming languages, to find the best fit for your environment. You can choose whether to optimize for speed or for accuracy and to what degree. You can choose whether to maximize entities discovered or minimize false positives, by using the returned confidence score for each discovered entity.

Combine these options to get the balance between precision, recall, performance, and convenience that is best for you and your product.

### Wide regulation support

With curated grammars covering personal data/personally identifiable information (PII) entities fully covering 46 countries and 42 languages, as well as many partially covered countries and languages, designed for regulations, such as GDPR (European Union), CCPA (California), KVKK (Turkey), POPI (South Africa), APP (Australia), NZPA (New Zealand), and CPPA (Canada). Protected health information (PHI) entities covering entity types, designed for regulations such as HIPAA (healthcare) and payment card industry (PCI) entities.

It also covers government classifications, including DoD markings, controlled unclassified information (CUI), export control, and the Australian Protective Security Policy Framework.



Figure 1 Global PII Discovery locations

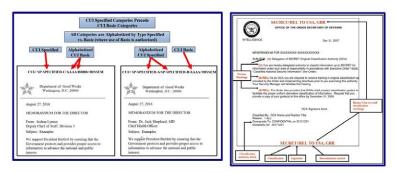


Figure 2 Examples of Entity Recognition

Quest Diagnostics is a leading healthcare provider with more than 2,000 patient service centers across the US. It uses OpenText Named Entity Recognition within OpenText Knowledge Discovery to enable fast and in-depth indexing, search, and retrieval.

Quest Diagnostics have unlocked more than 20 years' worth of data contained in 3.4 billion documents.

Read more >

Additional features	Description
Hybrid grammar support	Includes dictionary lists, regular expressions, componentized grammars and combinations.
Contextual processing	Uses structure and landmarks to improve recall and reduce false positives.
Custom grammar extensions	Which can be exposed to end users or become part of a value-add service.
Compiled grammars	For efficiency and IP protection—but still extendible.
Confidence scoring	Provides fine control over precision/recall curve.
Component tagging	Allows specificity as to exactly which bytes match with which components in a grammar, for use in downstream processing.
Fully streamed operation	For use in pipelined architectures (such as with OpenText File Content Extraction's text extraction component) for lower latency and lower memory usage.
Early termination	Lowers CPU usage when risk is identified or ruled out before an entire file is processed.
Pre-processing optimizations	Accurately controls scope for lower CPU usage, lower latency, and higher throughput.
Match validation	Ensures that error detection and correction codes, such as checksums, are valid, reducing false positives.
Normalization	Normalizes returned results, such as dates, addresses, and names, for consistent downstream processing and lower risk.
Automatic governance	Redacts, restricts and deletes entities where they reside to ensure compliance.
Sentiment analysis	Provides clause-level analysis, meaning that each subject within a sentence is given its own sentiment score.

## OpenText Named Entity Recognition deployment options:

#### **Extend your team**

 On-premises software, managed by your organization or OpenText

#### Resources

Request a demo >

Main web page >

What's New - Rolling blog >

## System requirements by form factor

#### **Embeddable library (Named Entity Recognition SDK)**

- Linux (x32&x64), Windows (x32&x64), macOS (x64, M1)
- C/C++, .NET, Java

#### Microservice (Named Entity Recognition Server)

- Linux (x64), Windows (x64)
- Process-based service
- HTTP-based API
- Requires Knowledge Discover licensing service

#### NiFi processor

- Linux (x64), Windows (x64)
- NiFI API
- Requires Knowledge Discover licensing service

### **About OpenText**

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on-premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit: opentext.com.

