IDOL Eduction

Real-time discovery of sensitive data in flight and at rest

Product Highlights

Eduction Engine

The OpenText™ IDOL Eduction engine is a flexible, embeddable software component which finds and extracts sensitive and high-value entities from structured, semi-structured, and unstructured data using high performance natural language processing.

Peak performance is critical for our customers, and it's not just the speed of the engine itself that matters. IDOL Eduction supports 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.

Your product is unique. You can 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.

Eduction Grammars

Eduction grammars define which entities are of interest. You can create your own, or use the basic entity grammars which are included as standard. For full access to OpenText's established expertise in grammar design, use our highly targeted premium 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 add-on 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.

An up-to-date full list of available grammars is available in the IDOL documentation.

Where your customers need to detect entities which are truly unique to them, our optimized grammar language can be exposed to end users for the ultimate in value-added configurability.

Better Together

IDOL Eduction is designed to work well with third party components through the use of readily consumable APIs and architectural options. When paired with other IDOL SDKs—such as KeyView for text extraction—additional cross-component functionality is unlocked, such as table extraction for improved accuracy, or pipelining for better performance.

Key Benefits

Use IDOL Eduction and our curated add-on grammars to provide a risk reduction service to your end users who need compliance with regulations and corporate mandates.

- **Defensible compliance** with information provided on data sources and examples of match types.
- **Quick to market**—our grammars are ready to use today. There is no need to delay your product launch.
- **Added differentiation** with both our fully maintained grammar sets and end user customizable grammars.
- **Explainable decisions** through use of hand-curated grammars rather than corpus derived models.
- **Wide regulation support** covering many countries, languages, and entity types.
- **Lower hardware costs**—latency and throughput can be balanced against precision and recall to meet cost objectives at any scale.
- **Reduce ongoing engineering cost** as regulations and entity definitions evolve. We ensure that the grammars remain up-to-date and high-performing.
- **An experienced team**, we have the experience necessary to understand how to create the most efficient grammars. With a library of hundreds of grammars currently available, we have done this before.

Key Features

Eduction

- **Hybrid grammar support** including dictionary lists, regular expressions, componentized grammars and combinations.
- **Contextual processing** using 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** to give 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 KeyView text extraction component) for lower latency and lower memory usage.

- **Early termination** for lower CPU usage when risk is identified or ruled out before an entire file is processed.

- **Pre-processing optimizations** to accurately control scope for lower CPU usage, lower latency, and higher throughput.

- **Match validation** to ensure that error detection and correction codes such as checksums are valid, reducing false positives.

- **Normalization** of returned results such as dates, addresses, and names, for consistent downstream processing and lower risk.

- **Automatic governance** will redact, restrict and delete entities where they reside to ensure compliance.

Add-on Grammars

- **Personal Data / Personally Identifiable Information (PII)** entities covering 59 countries and 38 languages, designed for regulations such as GDPR (European Union), CCPA (California), KVKK (Turkey), POPI (South Africa), APP (Australia), NZPA (New Zealand), and CPPA (Canada).

- **Payment Card Industry (PCI)** entities covering 43 countries.

- **Protected Health Information (PHI)** entities covering 18 entity types, designed for regulations such as HIPAA (healthcare).

- **Government Classification**, including DoD markings, Controlled Unclassified Information (CUI), export control, and the Australian Protective Security Policy Framework.

- **Sentiment Analysis** covering 13 languages and providing clause-level analysis, meaning that each subject within a sentence is given its own sentiment score.

System Requirements by Form Factor

Embeddable library (Eduction SDK)
- Linux (x64), Windows (x64), macOS (x64, M1)
- C / C++, .NET, Java

Microservice (Eduction Server)
- Linux (x64), Windows (x64)
- Process-based service
- HTTP-based API
- Requires IDOL licensing service

NiFi processor
- Linux (x64), Windows (x64)
- NiFi API
- Requires IDOL licensing service

Learn more at
www.microfocus.com/IDOL

www.opentext.com