IDOL
Eduction

Software Version 12.12.0

Release Notes

Micro Focus®

Legal notices

Documentation updates

The title page of this document contains the following identifying information:

- Software Version number, which indicates the software version.
- Document Release Date, which changes each time the document is updated.
- Software Release Date, which indicates the release date of this version of the software.

To check for updated documentation, visit https://www.microfocus.com/support-and-services/documentation/.

Support

Visit the MySupport portal to access contact information and details about the products, services, and support that Micro Focus offers.

This portal also provides customer self-solve capabilities. It gives you a fast and efficient way to access interactive technical support tools needed to manage your business. As a valued support customer, you can benefit by using the MySupport portal to:

- View information about all services that Support offers
- Submit and track service requests
- Contact customer support
- Search for knowledge documents of interest
- View software vulnerability alerts
- Enter into discussions with other software customers
- Download software patches
- Manage software licenses, downloads, and support contracts

Many areas of the portal require you to sign in. If you need an account, you can create one when prompted to sign in.

Contents

New in this Release

The following new features were released in Eduction version 12.12.0.

- Logging callbacks registered with an Eduction engine (through the C or .NET APIs) receive messages about all eduction events, such as timeouts. In earlier versions of Eduction, these callbacks only received information about events related to post-processing.

- Your Lua post-processing scripts can now define a function `resetprocessor`, which Eduction calls whenever the session resets (for example when it receives a new input stream, or more text after it has already processed a final input). Eduction passes the current user parameters in to the reset hook for use, if required. You can use this function to reset any global state that the script maintains with details of previous matches when you reuse a session to process another document or input buffer.

- Matching efficiency has been improved when all entities have an `EntityMatchLimit` configured; Eduction does no extra work after all individual entity limits have been met.

Enhancements to Eduction Grammars

- The advanced date/time grammar (`datetime_advanced_eng.ecr` has new entities to match contextual dates:

  - `datetime/advanced_date_context/eng` A context-aware date, for example *two days after*.

  - `datetime/advanced_date_and_time_context/eng` A contextual date with a time, for example *7pm the Tuesday before that*.

  - `datetime/advanced_context/eng` A combination of the other two entities; a contextual date with optional time.

  To use this entity, you must configure it explicitly; it is not included in the `datetime/advanced/eng` combined entity. You must also configure at least one date or datetime entity for the context entities to use for reference.

  You can optionally use proximity to filter out candidate relative dates that are too far from the contextual date, by using the following user parameters in the `datetime.lua` post-processing script:

  - `contextmaxdistance` Discard contextual matches more than this number of characters from the last date match.

  - `contextpenaltydistance` Reduce score for matches lying between this distance from the last date match and the maximum distance. This option applies a linear reduction, scaling to zero at the maximum distance.

  For information about how to use user parameters, refer to the *Eduction User and Programming Guide*.

- The PII name grammar has been improved when matching names where the surname occurs first, for example "Smith, Peter John".

- The `pii/banking/swiftcode` entities have been extended to support all SWIFT registered countries.

- The `pii/name/title_surname/`*cc*, `pci/name/title_surname/`*cc*, and `phi/name/title_surname/`*cc* entities have been improved. For example, spurious matches such as "King Street" are no longer returned.

- The `pii/banking/iban/context/`*cc* and `pii/banking/iban/nocontext/`*cc* entities have been updated with the following additional countries: Andorra (AD), Azerbaijan (AZ), Georgia (GE), Gibraltar (GI), Holy See (VA), Iraq (IQ), Jordan (JO), Kosovo (XK), Lebanon (LB), Libya (LY), Mauritania (MR), Mauritius (MU), Palestine (PS), Qatar (QA), Saint Lucia (LC), Sao Tome and Principe (ST), Seychelles (SC), Sudan (SD), Timor-Leste (TL), Tunisia (TN), Virgin Islands British (VG)

  In addition, entities have been added for the Faroe Islands (FO). These patterns were previously incorrectly matched by the Denmark (DK) entities.

- The `pii/banking/context/ie` entity has been adjusted for sort codes, so that anything that does not match a known sort code has a score penalty, and anything that does not start with a 9 digit has a larger score penalty.

- The `gov/document_markings/us_dod/banner_line/nocontext/us` entity has been updated with an additional banner line.

- The recall of the `finance/news` entity in the `financial_strength.ecr` grammar has been improved.

- Match quality has been improved for driving licenses, by automatically stoplisting short numerical matches and matches that consist entirely of alphabetic characters.

Resolved Issues

The following issues were resolved in Eduction version 12.12.0.

- `EdkGetNextMatch` could incorrectly return `EdkErrInvalidUTF8`.

- Eduction could match a substring of an entity composed of full-width characters, even if `MatchWholeWord=TRUE`.

- In certain situations, using multiple ECR files with Eduction could cause internal entities to become public. This could result in unexpected matches.

- The C example programs could fail to build using the Visual Studio compiler if the supplied license key file exceeded a certain size.

- When `EdkGetNextRepeatedMatch` and pre-filter tasks were used, some input could cause an interruption of service.

- Eduction versions 12.8 to 12.11 could take longer than necessary to create an `EdkSession`.

- When attempting to use the `address_cjkvt` prefilter dictionary file with CJKVT input, Eduction could exit unexpectedly.

Resolved Issues for Eduction Grammars

- The advanced date/time grammar (`datetime_advanced_eng.ecr` with the script `datetime.lua`) would give a post-processing error when matching timezones in the format "UTC±hh" or "GMT±hh".

- The US national ID entities (such as `pii/id/context/us`) would match some invalid social security numbers with a single repeated digit (`111-11-1111`) or the sequential digits 1-9 (`123-45-6789`).

- Some national ID and tax identification numbers, without context, would fail to match at `MinScore=0.4`. The scores for distinctive `nocontext` national ID and tax identification numbers, including those without checksum validation, have been increased so that they score >=0.5.

- The `pii/banking/iban/context/dk` and `pii/banking/iban/nocontext/dk` entities matched patterns for Greenland and the Faroe Islands.

  > **NOTE:** New `pii/banking/iban/context/fo` and `pii/banking/iban/nocontext/fo` entities have been added to match Faroese patterns.

- The `pii/banking/account_number/nocontext` and `pii/banking/roll_number/nocontext` entities scored matches as 0.2, which meant that they were not returned as nocontext matches. These entities now score 0.4.

- When ambiguous entity handling was enabled, Switzerland was sometimes dropped from the list of candidate countries from driving license matches. For example, the text *führerausweis 006551496001* was returned as a match for `pii/driving/context/li`, rather than the ambiguous match `pii/driving/context/ambiguous/ch_li`.

- The **pii/health/id/context/jp** entity sometimes incorrectly treated Latin-alphabet words as landmarks if they began with a valid landmark string, generating spurious matches (for example *Codeine* was mistakenly treated as an occurrence of the landmark *Code*).

- For the **pii/health/id/context/jp** entity, the score has been reduced for context matches for this entity when a generic landmark (such as *Code* or *Sign*) appears in conjunction with a very short entity match (1 to 4 digits).

- For PII and PCI grammars, customization of the name lists in the CJKVT name grammar did not work as described in the documentation.

- The `medical_terms` grammar could return erroneous matches for certain English words.

Resolved Issues for Eduction Server

- **(Security update)** The third-party OpenSSL library has been upgraded to version 1.1.1n to resolve a security vulnerability, CVE-2022-0778.

Notes

These notes provide extra information about installing and using Eduction.

Deprecated Features

The following features have been deprecated.

Category	Deprecated Feature	Deprecated Since
Eduction SDK	The ability to specify a license key by supplying a file path has been deprecated. Micro Focus recommends embedding your license key in your application as a string, to avoid having the license key in a file on disk. In the C API the following functions have been deprecated: • `EdkEngineCreate()` • `EdkEngineCreateFromConfigFile()` • `EdkSetLicenseKey()` • `EdkSetLicenseKeyFromFile()` In the Java API the following have been deprecated: • Both of the constructors for `EDKEngine`. • The `setLicenseKey` method of `EDKEngine`. In the .NET API, the following has been deprecated: • The `EDKFactory` constructor: `public EDKFactory(string license_key_path)` Micro Focus recommends supplying the license key as a string, and creating an Eduction engine by using an engine factory. For more information, refer to the API reference documentation.	12.11
.NET Eduction SDK	Eduction SDK support for .NET Standard 1.1 has been deprecated and might be removed in future. Micro Focus recommends using a .NET implementation that supports .NET Standard 2.0.	12.11

C Eduction SDK	`EdkSessionSetMatchTimeout`
	
	`EdkSessionSetRequestTimeout`
	
	`EdkGetNextMatchTimed`
		12.8

.NET Eduction SDK	`ITextExtractionSession::SetMatchTimeout` `ITextExtractionSession::SetRequestTimeout` `ITextExtractionSession::SetStartTime`	12.8
Java Eduction SDK	`TextExtractionSession<T>::setMatchTimeout` `TextExtractionSession<T>::setRequestTimeout`	12.8
Java Eduction SDK	`EDKMatch::getTimedOut` function	12.6
edktool configuration	`SearchFields` parameter. This parameter applies only to IDX document input, which was deprecated in 12.5.0.	12.6
Configuration	`[PostProcessingTasks]` configuration section. Use the `PostProcessingTaskN` and `PostProcessThreshold` parameters in the `[Eduction]` section.	12.5
edktool	The `OutputScores` and `OutputSimpleMatchInfo` configuration parameters. **NOTE:** The default values of the parameters `OutputSimpleMatchInfo` and `OutputScores` have been updated, to `False` and `True` respectively.	12.5
edktool	XML configuration files are now deprecated. Use CFG format configuration files instead.	12.5
edktool	Eduction on IDX documents. The following configuration parameters that apply only to IDX are also deprecated as part of this change: • `AllowDuplicates` • `CantHaveFieldCSVs` • `Databases` • `DocumentDelimiterCSVs`	12.5

Requirements

This section describes the system requirements, supported platforms, and software dependencies for Eduction 12.12.0.

Minimum System Requirements

The following are minimum system requirements for Eduction 12.12.0.

- 512 MB RAM

- 1 GB disk space

- 800 MHz or higher processor

- 400 MB of space designated for `/tmp` or `/var/tmp` (UNIX), or user-defined temporary directory (Windows)

Software Dependencies

Java

Development with the Eduction Java SDK requires the Java 8.0 JDK and JRE.

Microsoft Visual C++ Redistributable Package

To run Eduction version 12.12.0 on the Microsoft Windows operating system, you might need to install Microsoft Visual C++ Redistributable packages. The Eduction SDK and Eduction Server stand-alone zip packages include the required redistributable files for Microsoft Visual C++ 2017. You can also update your packages by using the latest version at:
http://support.microsoft.com/kb/2019667.

UNIX Libraries

To run Eduction version 12.12.0 on UNIX platforms (except MacOS on the M1 platform), the server must have the following minimum versions of libraries:

- GLIBC_2.3.2

- GLIBCXX_3.4.21

- GCC_4.8.0

> **NOTE:** The Eduction SDK and Eduction Server stand-alone zip packages provide these libraries in the `libgcc_s` and `libstdc++` shared libraries.
>
> You might need to set the `LD_LIBRARY_PATH` to include the *InstallDir*`/bin` directory, to ensure that Eduction can access the installed shared libraries.
>
> You can also copy the shared libraries to the component working directory.

Supported Operating System Platforms

The following operating system platforms are supported by Eduction 12.12.0.

- Windows x86-64

- Linux x86-64

- Linux ARM 64 (centos AArch 64)

- MacOS x86-64 and M1 (Eduction SDK only)

- Solaris x86-64 and SPARC 64

> **DEPRECATED:** Support for Solaris x86-64 and SPARC 64 was deprecated in IDOL version 12.10, for all IDOL components except KeyView. IDOL components are currently still available for Solaris, but might not be available in future.

Documentation

The following documentation was updated for Eduction version 12.12.0.

- *Eduction Server Reference*
- *Eduction User and Programming Guide*