Eduction

Software Version 12.7.0

Release Notes

Document Release Date: October 2020
Software Release Date: October 2020

Legal notices

Copyright notice

The only warranties for products and services of Micro Focus and its affiliates and licensors (“Micro Focus”) are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Micro Focus shall not be liable for technical or editorial errors or omissions contained herein. The information contained herein is subject to change without notice.

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:

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

Many areas of the portal require you to sign in. If you need an account, you can create one when prompted to sign in. To learn about the different access levels the portal uses, see the Access Levels descriptions.

Contents

New in this Release

The following new features were released in Eduction version 12.7.0.

Enhancements to Eduction Server and Eduction SDK

- You can now configure Eduction in Table mode to extract entities from structured data. For example, you can use this mode to extract all driving licenses from a column titled 'driving license number' in a particular table.

  This functionality uses the new `HeaderEntity`*`N`* and `CellEntity`*`N`* configuration parameters. It can extract entities from tables in CSV and TSV format.

  For more information, refer to the *Eduction User and Programming Guide*.

  When you run Eduction through Connector Framework Server oo IDOL NiFi Ingest, you can also extract entities from structured output (for example, OCR output from Media Server), by using the `TableCellPath` and `ColumnSpanAttribute` parameters. For more information, refer to the *Connector Framework Server* documentation.
- The `psi/private_key/pem` entity now matches data containing extraneous newlines (for example, any newlines that are present because of formatting on source input).
- For the `psi/api_credentials/secret_key/aws` entity, the `psi_api_credentials_postprocessing.lua` script now applies a score to the matches based on randomness, to attempt to eliminate false positives.
- Eduction grammar compilation now produces compressed grammar files. Eduction can also now read compressed grammars. This change reduces the amount of disk space Eduction grammars take up.

  > **NOTE:** Eduction can still handle existing uncompressed grammars as before.

Enhancements to Eduction Server

- The OpenSSL library has been updated to version 1.1.1g.

Enhancements to Eduction SDK

- The Eduction SDKs now have functions to allow you to retrieve the row and column of a match:
  - C API: `EdkGetMatchTablePosition`
  - Java API: `public EDKMatch.TablePosition getTablePosition()`
  - .NET API: `public IExtractionMatchTablePosition TablePosition`

  There are also new sample programs available to show how to use these functions.
- The .NET component now targets .NET Standard 2.0 (netstandard 2.0).
- The .NET Eduction SDK is now available on Linux platforms. This version provides netstandard2.0 and netstandard1.1 assemblies.

- You can now add a compilation configuration file to the compile command for edktool to allow you to perform character expansions. These expansions allow you to specify that Eduction detects one or more characters as if they are a particular source character. For example, you can use this option to treat particular punctuation characters as equivalent to a standard form that you use in your grammars.
- You can now specify the path to a compilation configuration file when compiling a grammar with the Eduction SDKs:
  - C API: `EdkLoadResourceFileWithCompileConfig` and `EdkLoadResourceBufferWithCompileConfig`.
  - Java API: `loadResourceFile`, `loadResourceFiles`, and `loadResourceBuffer` methods in the `TextExtractionEngine` interface.
  - .NET API: `GetCompiler` method on the `EDKFactory` class.
- The EductionDotNet assembly is now strong named.
- New logging callback functionality has been added to the Eduction C and .NET SDKs:
  - ```
    typedef void(*EdkLoggingFn)(void* pUserData, int nLogLevel, const char*
    szMsg);
    EDKAPI EdkError EdkEngineSetLoggingFn(EdkEngineHandle pEngine, EdkLoggingFn
    loggingFn);
    EDKAPI EdkError EdkEngineSetLoggingUserData(EdkEngineHandle pEngine, void*
    pUserData);
    ```

    > **NOTE:** Micro Focus recommends that you use this new logger function, which is more flexible than the existing `EdkEngineSetLoggerFn` function.

  - ```
    public delegate void fnEngineLoggingFunction(IntPtr user_data, int log_
    level, string msg)

    ITextExtractionEngine:
    void SetLoggerFn(fnEngineLoggingFunction loggerFn);
    void SetLoggerUserData(IntPtr user_data);
    ```

    These methods set a logger function, which is called whenever Lua post-processing has information to log (such as an error). The pointer that is passed in to the user data function is the first argument in the logger function. The message string that the function receives is fully formatted and there are no variable arguments to consume.

Resolved Issues

The following issues were resolved in Eduction version 12.7.0.

- Very rarely, the normalized text of a match for a pattern using the exclusion operator (`?A!...`) could miss part of the text lying outside the exclusion.
- Eduction processing could be slower than expected when text normalization was not being conducted. The text processing speed is now comparable to that of the 12.2 release.
- When using Eduction in CFS or NiFi, Lua post-processing scripts that renamed the matching entity could drop matches, or return the match under an unexpected field name (such as "_"). For example, the post-processing scripts shipped in the IDOL PII Package rename entities from the combined grammars ending "/all" to correspond to the matched language.
- Running edktool benchmark mode on a Windows (CRLF) file that contained many lines and matches near the end of the file could return an incorrect set of matches unless the -b flag (binary mode) was specified.
- The `person_politician_engus` grammar has been updated with an up-to-date list of senators, governors, and members of the house of representatives. It also includes new entities for members of Donald Trump's cabinet.

Resolved Issues for Eduction Server

- Various `GetRequestLog` (`GRL`) HTML issues have been fixed:
  - Actions were not properly escaped in some places, which could allow malicious users to inject some alternative HTML in the `GetRequestLog` response.
  - The Open link in new window option did not work.
  - The auto refresh option would modify the selected open link option.
  - Links to the obsolete confighelp action have been removed.
  - The HTML page now uses a Content-Security-Policy to protect against malicious data injection by crafted actions sent to the server.

Resolved Issues for Eduction SDK

- When creating an Eduction engine from a configuration file in the Eduction SDK (C, Java, or .NET), Eduction did not respect the configured `PostProcessThreshold` value during matching.

Notes

These notes provide extra information about installing and using Eduction.

- The default values of the parameters `OutputSimpleMatchInfo` and `OutputScores` have been updated, to `False` and `True` respectively. These parameters apply only to `edktool` and were deprecated in IDOL 12.5. They might be removed in future.

Deprecated Features

The following features have been deprecated.

Category	Deprecated Feature	Deprecated Since
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.	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

The following notes apply to Eduction Server only:

Requirements

This section describes the system requirements, supported platforms, and software dependencies for Eduction 12.7.0.

Minimum System Requirements

The following are minimum system requirements for Eduction 12.7.0.

- 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 12.7.0 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 12.7.0 on UNIX platforms, 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 12.7.0.

- Windows x86 64
- Linux x86 64
- Solaris x86 64
- Solaris SPARC 64
- Mac OS X (Eduction SDK only)

Documentation

The following documentation was updated for Eduction version 12.7.0.

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