

IDOL Education

Software Version 12.10.0

Release Notes



Document Release Date: October 2021
Software Release Date: October 2021

Legal notices

Copyright notice

© Copyright 2021 Micro Focus or one of its affiliates.

The only warranties for products and services of Micro Focus and its affiliates and licensors (“Micro Focus”) are as may be 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	4
Resolved Issues	6
Notes	7
Requirements	9
Minimum System Requirements	9
Software Dependencies	9
Supported Operating System Platforms	10
Documentation	11

New in this Release

The following new features were released in Eduction version 12.10.0.

- The Eduction APIs (C, Java, and .NET) provide a way to iterate through repeated matches before returning to the next unique match. A *repeated match* occurs when Eduction finds another match (exactly the same text), at a different location in the input (a different offset). This feature provides a convenient way to stop processing repeated matches and move on to the next unique match or maybe even the next document.
- The performance of case-insensitive matching has improved. If you configure `MatchCase=False`, entities are optimized for case-insensitive matching when they are loaded. This can increase the loading time, so if you regularly use a grammar file with `MatchCase=False` you can optimize the entities for case-insensitive matching at compile time instead. To do this, specify `{"alternativeCaseArcs": true}` in a compilation configuration file.
- `edktool` has a new command, `unify`, that you can use to create combined entities. A *combined entity* is a single entity that combines the patterns for a predetermined set of entities in order to improve the speed of matching at run-time.
- The Eduction parameter `PreFilterMaxReturnedBytes` applies to pre-filter output after a match is found (the counter is reset and another match must occur before the limit is reached, otherwise Eduction stops processing). In previous versions of Eduction, `PreFilterMaxReturnedBytes` was disabled after the first match.
- Entity patterns support Unicode character classes. For example, you can use the regular expression `\p{Sc}` to match any currency symbol.
- Eduction no longer creates a temporary file when loading a compressed `.ecr` grammar file.

Enhancements to Eduction Server

- The third-party libraries `libxml2` and `libxslt` have been updated, to versions 2.9.11 and 1.1.34 respectively.

Enhancements to Eduction Grammars

- The PII grammars include an HTML file that lists the entities that are available. This file makes it easier to see which countries or languages are supported for particular entities.
- The PII grammars include new grammar files, `banking.ecr` and `banking_cjkvt.ecr` (and equivalent `.ejr` files). These provide entities to locate bank account details (such as account numbers, routing numbers, and SWIFT codes) for the countries AU, CA, CN, DE, FR, GB, IE, IL, JP, NZ, and US. There are also entities to locate International Bank Account Numbers (IBANs).
- The PII grammars include a new grammar file, `medical_terms_cjkvt.ecr`, which contains entities to locate medical conditions or procedures in Chinese, Japanese, and Korean. There is also a combined grammar named `combined_medical_terms_cjkvt.ecr` that contains the entity `pii/medical_terms/all_cjkvt`. This offers better performance if you want to use all of the available languages.

- The PII address entities for Austria, Brazil, Bulgaria, and Croatia have been improved in order to improve recall.
- The PII entities for Japan have been improved:
 - New entities have been added for pension numbers (for example `pii/id/context/pension/jp`) and unemployment insurance numbers (for example `pii/id/employment_insurance/context/jp`).
 - The entities `pii/id/context/jp`, `pii/id/landmark/jp`, and `pii/id/nocontext/jp` now cover Residence Card number and Resident Registration number. Earlier versions of Education included only Individual Number (also known as "My Number").
 - The driving license number landmarks have been extended.
 - The entities `pii/health/id/context/jp`, `pii/health/id/landmark/jp`, and `pii/health/id/nocontext/jp` have been updated to include Japanese patterns and landmarks for Long-term care insurance numbers and Worker's accident compensation numbers.
- The PII entities for matching driving license numbers (`pii/driving/...`) for Germany, Ireland, and Italy have been improved, to eliminate some false positives.
- The IDOL Government Education Grammars include a new grammar file, `us_regulation_references.ecr`, which contains entities for matching citations of U.S. regulations.

Resolved Issues

The following issues were resolved in Education version 12.10.0.

- When pre-filtering generated many overlapping windows, Education could erroneously ignore the value of the `PreFilterMaxReturnedBytes` parameter.

Notes

These notes provide extra information about installing and using Education.

Deprecated Features

The following features have been deprecated.

Category	Deprecated Feature	Deprecated Since
C Education SDK	EdkSessionSetMatchTimeout EdkSessionSetRequestTimeout EdkGetNextMatchTimed	12.8
.NET Education SDK	ITextExtractionSession::SetMatchTimeout ITextExtractionSession::SetRequestTimeout ITextExtractionSession::SetStartTime	12.8
Java Education SDK	TextExtractionSession<T>::setMatchTimeout TextExtractionSession<T>::setRequestTimeout	12.8
Java Education 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 [Education] 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	Education on IDX documents. The following configuration parameters that apply only to IDX are	12.5

	<p>also deprecated as part of this change:</p> <ul style="list-style-type: none">• AllowDuplicates• CantHaveFieldCSVs• Databases• DocumentDelimiterCSVs	
--	--	--

Requirements

This section describes the system requirements, supported platforms, and software dependencies for Education 12.10.0.

Minimum System Requirements

The following are minimum system requirements for Education 12.10.0.

- 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 Education Java SDK requires the Java 8.0 JDK and JRE.

Microsoft Visual C++ Redistributable Package

To run Education version 12.10.0 on the Microsoft Windows operating system, you might need to install Microsoft Visual C++ Redistributable packages. The Education SDK and Education 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 Education version 12.10.0 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 Education SDK and Education 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 Education 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 Education 12.10.0.

- Windows x86-64
- Linux x86-64
- Linux ARM 64 (centos AArch 64)
- MacOS x86-64 and M1 (Education 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 Education version 12.10.0.

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