

# IDOL Education

Software Version 12.11.0

## Release Notes



Document Release Date: February 2022  
Software Release Date: February 2022

## Legal notices

© Copyright 2022 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:

- 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 .....	4
Resolved Issues .....	6
Notes .....	7
Requirements .....	10
Minimum System Requirements .....	10
Software Dependencies .....	10
Supported Operating System Platforms .....	11
Documentation .....	12

## New in this Release

The following new features were released in Education version 12.11.0.

- All of the Education APIs allow you to pass a license key to Education as a string. Micro Focus recommends embedding your license key in your application as a string, to avoid having the license key in a file on disk. The ability to specify a file path has been deprecated.
- You can use the new parameter `EntityMatchLimitN` to stop searching for an entity after a certain number of matches have been found. You might set this parameter if you want to see some matches for a given entity but would prefer to ignore further matches in favor of reducing the processing time.
- The output from `edktool list` (the list of entities, and each list of components returned by an entity), is now sorted alphabetically by name.

### Enhancements to Education Grammars

- The PII address entities for Bulgaria have been improved in order to improve recall, and the normalized form now better matches Bulgarian convention.
- The PII address entities for Cyprus, Ireland, Malta, and Romania have been improved in order to improve recall.
- The PII health ID entities for Canada (such as `pii/id/health/context/ca`) now match health care numbers for Canadian Forces, RCMP, Veteran Affairs, and NSOU. There are new entities for matching health care numbers for each province, for example `pii/health/id/context/alberta/ca`. Checksum validation has been added for the scheme numbers that support it.
- The PII national ID grammars have been improved.
  - The entities for France (such as `pii/id/context/fr`) support CNI (carte nationale d'identité) numbers.
  - The entities for Hungary (such as `pii/id/context/hu`) support the Hungarian TAJ (social security number). These numbers are validated by using a checksum.
  - Latvian national ID (Personas kods) numbers are now validated using a checksum.
  - The entities for Luxembourg (such as `pii/id/context/lu`) support non-natural-person ID numbers.
  - The entities for Poland (such as `pii/id/context/pl`) support ID card (Dowód osobisty) numbers. The card numbers are validated by using a checksum. Matches for the entity will contain a PESEL or DOWOD\_OSOBISTY component to indicate which identifier was detected.
  - The entities for Taiwan (such as `pii/id/context/tw`) match the numbers issued to resident non-nationals on their identity cards (ARC, Alien Resident Certificate or TARC, Taiwan Area Resident Certificate).
  - The patterns for Sweden and Portugal have been improved to enhance precision and recall.

- The PII tax identification number grammars have been improved.
  - The entities for Australia (`pii/tin/*/au`) support the Australian Company Number. These numbers are validated by using a checksum.
  - The British "Unique Taxpayer Reference" is now validated by checksum and the grammar includes landmarks such as "Unique Taxpayer Reference" and "UTR".
  - The entities for India (`pii/tin/vatin/*/in`) support the "Goods and Services Tax Identification Number" (GSTIN).
  - The entities for Poland (`pii/tin/*/pl`) match the REGON (Rejestr Gospodarki Narodowej – Register of the National Economy) system of ID numbers for businesses or individuals. Both national and local forms of the REGON are recognized, and both are validated using a checksum.
- The PII telephone grammar has new entities (for example `pii/telephone/context/c1`) to support telephone numbers for Chile.
- The Russian passport entities (`pii/passport/*/ru`) have been improved, to exclude some invalid internal passport numbers, and to add more landmarks relating specifically to the Russian internal passport.
- The PII, PCI, and PHI post-processing Lua scripts read environment variables so that you can customize checksum validation. To stop rejecting Education matches that fail checksum validation, set the environment variable `IDOL_PII_CHECKSUM_VALIDATION_STRICT` to `FALSE`. When you do this, a score penalty is applied and an `INVALID_CHECKSUM` component is added to the match. You can configure the score penalty by setting the environment variable `IDOL_PII_CHECKSUM_VALIDATION_FAIL_PENALTY`.

#### **Enhancements to Education Server**

- The path of the temporary directory, where the server stores files that were uploaded in multipart requests, can be configured using the new parameter `TempDir` in the `[Paths]` section of the configuration file.

## Resolved Issues

The following issues were resolved in Education version 12.11.0.

- When the Education SDK was used to process input in streaming mode, an application could terminate unexpectedly when a match crossed a streaming boundary and coincided with a match at the same position that didn't cross the same streaming boundary. The same issue could be encountered when `PreFilterMaxReturnedBytes` was set.
- Processing input containing bytes that should never appear in valid UTF-8 text could, with certain grammars, cause an application to terminate unexpectedly.
- When entities were created using the `edktool unify` command, Education could report matches using the unified entity name instead of the original entity name. This could occur when the original pattern used entity references (`?A^...`), or nested entity copies (`?A:...`), within a no-output group (`?A!...`).

### Resolved Issues for Education Grammars

- Education did not match Japanese names (`pii/name/jp`) where there was a space between the surname and given name.
- Capitalized words containing a title prefix could be incorrectly matched as a name. For example, "DRESS SMART" could be matched as "DR ESS SMART".
- When matching Spanish tax identification numbers (`pii/tin/*/es`), the scores for NIE numbers were incorrectly penalized if the match did not include a hyphen.
- In the Government Education package, the entity `gov/regulation_references/usc/us` could produce duplicate matches differing only by whitespace.

### Resolved Issues for Education Server

- When the server was configured to require SSL and GSS authentication for incoming connections, POST requests over a certain size could fail with the error "httpClient error - could not receive request (HTTP message-header is too long)".
- The server was unable to renew its license and would therefore stop, if SSL was required for the connection to the IDOL License Server.

## Notes

- The behavior of character expansions has changed. (You can configure character expansions in a compilation configuration file, which you use when compiling grammar files with `edktool`. For an introduction to character expansions, refer to the *Eduction User and Programming Guide*.)

Eduction now considers the source and destination characters as a single list where any character in the list is expanded to any other. The character chosen as the "src" character is significant only because it is used in normalized matches in place of any "dest" character. For example, consider the following configuration:

```
{  
  "expansions": [  
    { "src": "a", "dest": ["b", "c"] }  
  ]  
}
```

The following table shows how Eduction 12.10 and Eduction 12.11 expand some example patterns with this configuration:

Example pattern	Expanded pattern
<code>&lt;pattern&gt;ade&lt;/pattern&gt;</code>	Both Eduction 12.10 and 12.11 expand this pattern to: <code>&lt;pattern&gt;[abc]de&lt;/pattern&gt;</code> Matches are normalized to <code>ade</code> .
<code>&lt;pattern&gt;bde&lt;/pattern&gt;</code>	Eduction 12.10 does not expand the pattern. Eduction 12.11 expands the pattern as follows, and normalizes matches to <code>ade</code> : <code>&lt;pattern&gt;[abc]de&lt;/pattern&gt;</code>
<code>&lt;pattern&gt;bbc&lt;/pattern&gt;</code>	Eduction 12.10 does not expand the pattern. Eduction 12.11 expands the pattern as follows, and normalizes matches to <code>aaa</code> : <code>&lt;pattern&gt;[abc][abc][abc]&lt;/pattern&gt;</code>

These notes provide extra information about installing and using Eduction.

### Deprecated Features

The following features have been deprecated.

Category	Deprecated Feature	Deprecated Since
----------	--------------------	------------------

Education SDK	<p>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.</p> <p>In the C API the following functions have been deprecated:</p> <ul style="list-style-type: none"> <li>• EdkEngineCreate()</li> <li>• EdkEngineCreateFromConfigFile()</li> <li>• EdkSetLicenseKey()</li> <li>• EdkSetLicenseKeyFromFile()</li> </ul> <p>In the Java API the following have been deprecated:</p> <ul style="list-style-type: none"> <li>• Both of the constructors for EDKEngine.</li> <li>• The setLicenseKey method of EDKEngine.</li> </ul> <p>In the .NET API, the following has been deprecated:</p> <ul style="list-style-type: none"> <li>• The EDKFactory constructor: <code>public EDKFactory(string license_key_path)</code></li> </ul> <p>Micro Focus recommends supplying the license key as a string, and creating an Education engine by using an engine factory. For more information, refer to the API reference documentation.</p>	12.11
.NET Education SDK	<p>Education 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.</p>	12.11
C Education SDK	<p>EdkSessionSetMatchTimeout</p> <p>EdkSessionSetRequestTimeout</p> <p>EdkGetNextMatchTimed</p>	12.8
.NET Education SDK	<p>ITextExtractionSession::SetMatchTimeout</p> <p>ITextExtractionSession::SetRequestTimeout</p> <p>ITextExtractionSession::SetStartTime</p>	12.8
Java Education SDK	<p>TextExtractionSession&lt;T&gt;::setMatchTimeout</p> <p>TextExtractionSession&lt;T&gt;::setRequestTimeout</p>	12.8
Java Education SDK	<p>EDKMatch::getTimedOut function</p>	12.6
edktool configuration	<p>SearchFields parameter.</p> <p>This parameter applies only to IDX document input,</p>	12.6

	which was deprecated in 12.5.0.	
Configuration	[PostProcessingTasks] configuration section. Use the PostProcessingTaskN and PostProcessThreshold parameters in the [Education] section.	12.5
edktool	The OutputScores and OutputSimpleMatchInfo configuration parameters.  <b>NOTE:</b> 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 also deprecated as part of this change: <ul style="list-style-type: none"><li>• AllowDuplicates</li><li>• CantHaveFieldCSVs</li><li>• Databases</li><li>• DocumentDelimiterCSVs</li></ul>	12.5

# Requirements

This section describes the system requirements, supported platforms, and software dependencies for Education 12.11.0.

## Minimum System Requirements

The following are minimum system requirements for Education 12.11.0.

- 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 12.11.0 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 12.11.0 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 12.11.0.

- 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 12.11.0.

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