

Kofax Capture 9.0 ABBYY Engine

Application Note

Date	February 01, 2011
Applies To	Kofax Capture 9.0
Summary	This application note provides information about the new ABBYY FineReader 9.0 that is used within Kofax Capture (KC) 9.0 as a default recognition engine, such as: ABBYY new features, features translation from OCE (the previous recognition engine) to ABBYY, character type scope of ABBYY, and some known issues that ABBYY may encounter. It also addresses the use of OCE engine in KC 9.0 as optional installation and licensing.
Revision	1.3

Kofax Capture 9.0 ABBYY Engine Overview

Beginning with Kofax Capture 9.0, the OCE Recognition Engine, which was used in Ascent Capture 7.x and Kofax Capture 8.0, is no longer used and is replaced with ABBYY FineReader Engine. This change will affect the recognition result for Index Zones, Registration Zones, Form ID Zones, Bar Code and Patch Code recognition during Scan and Recognition time. This change is NOT related to / will NOT affect the recognition result during PDF Generation time.

The new ABBYY FineReader Engine has different characteristics compare to the OCE Engine. This application note will provide some information of the new recognition engine features to help Kofax Capture users to better understand its features and what to expect from the new engine.

Although the new ABBYY FineReader Engine has several new features, Kofax Capture 9.0 will not expose all the new features. The following will be discussed in this document:

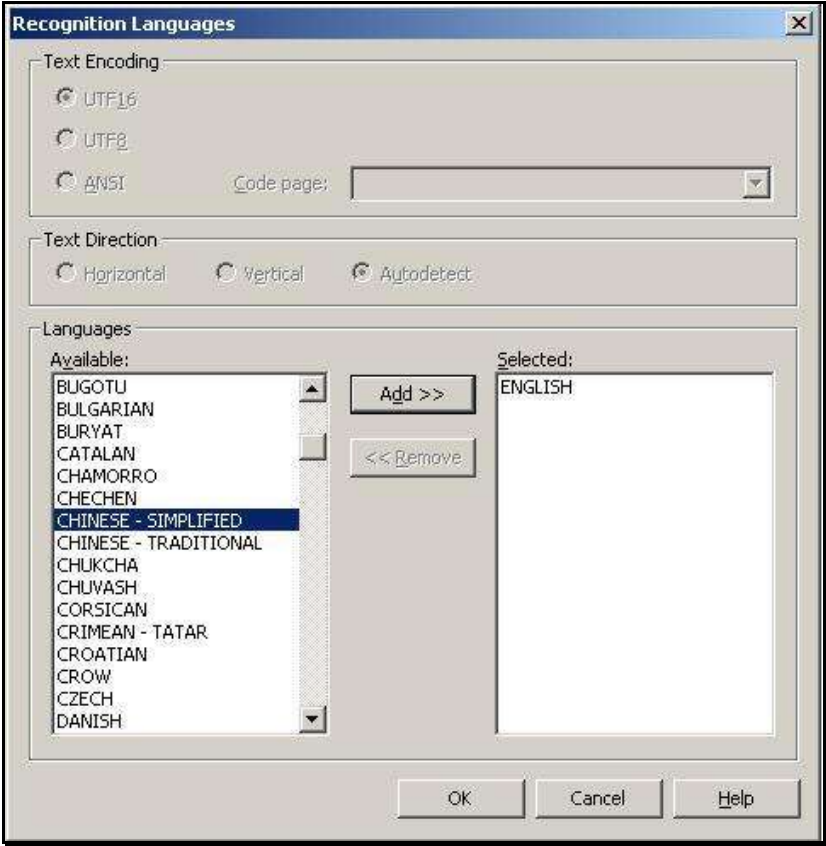
- New Languages
- PDF Versioning
- MICR barcode font CMC-7
- MICR text type E13B
- OCE vs. ABBYY
- Optional OCE's RecoStar Installation

Kofax Capture 9.0 ABBYY Engine

New Languages

ABBYY FineReader 9.0 supports several new languages:

- Chinese Simplified
- Chinese Traditional
- Japanese
- Korean
- Thai
- Hebrew
- Yiddish



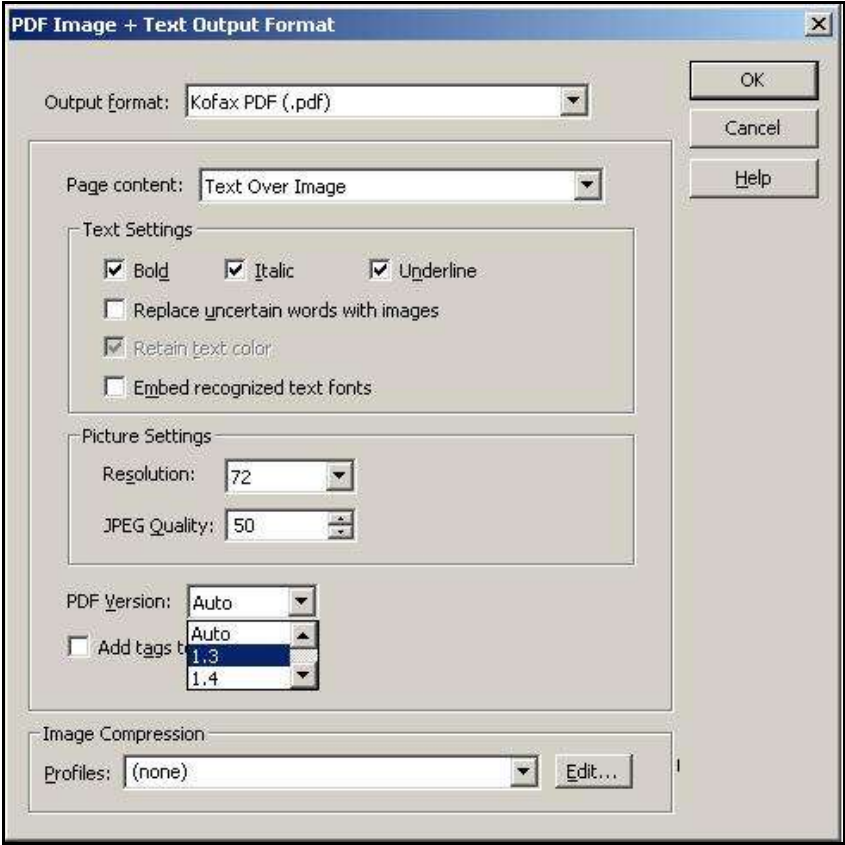
Kofax Capture 9.0 ABBYY Engine

Application Note

PDF Versioning

In Kofax Capture 9.0, the exported PDF version can be configured to one of the following version numbers:

- Auto
- 1.3
- 1.4
- 1.5
- 1.6



When Auto is selected, the PDF version number is determined by the engine and set to the actual PDF version number.

The PDF version number can never be less than the actual PDF version number; for example, if the actual PDF Version number supposes to be 1.6 and the Administrator configures the PDF Version number to be 1.5, the exported PDF image version number will be 1.6 and not 1.5.

Kofax Capture 9.0 ABBYY Engine

Application Note

MICR Barcode Font CMC-7

CMC-7 font is a special MICR barcode font that is used in Mexico, France, Spain, Brazil and most Spanish speaking countries. It is normally used to print characters for magnetic recognition and optical character recognition systems.

CMC-7 font contains 10 specially designed numeric characters 0 through 9, and **five** special symbols.



MICR Text Type E13B

The MICR E13B font is a special font that is used on bank checks and drafts in the United States, Canada, Puerto Rico, Panama, UK, and a few other countries. It is normally used to print magnetic recognition and optical character recognition systems.

E13B font contains 10 specially designed numeric characters 0 through 9, and **four** special symbols:

- Transit
- Amount
- On-us
- Dash



OCE vs. ABBYY

Engine Features Translation

The table below is the engine features translation when running the High Performance recognition profile using ABBYY Engine.

Feature	OCE	ABBYY
Languages	Central European	Russian Czech Croatian Hungarian Polish

Kofax Capture 9.0 ABBYY Engine

Application Note

Feature	OCE	ABBYY
	English – Canadian English – Irish English – United Kingdom English – United States	English
	French French – Belgian French – Canadian French – Switzerland	French
	German German – Austrian German – Switzerland	German
	Italian Italian – Switzerland	Italian
	South American	Spanish Portuguese
	Scandinavian	Swedish Norwegian
	Western European	English French German Italian Spanish

NOTE: The languages that are listed in the table are only the languages that need special conversion. If no special conversion is needed, the language is not listed in the table.

Feature	OCE	ABBYY
Print Type	Unknown	Automatic
	Machine print	Normal
	Fixed	Normal
	Handprint	New
	OCR-A	New
	OCR-B	New
	Farrington 7B	N/A
	CMC7	CMC7
	E13B	E13B

Kofax Capture 9.0 ABBYY Engine

Application Note

Feature	OCE	ABBYY
Content	Alphanumeric	Any
	Numeric	Numbers only
	Amount	Numbers only
Character set – Regular Expression	Character set	New
Machine Print Pitch Handprint Pitch	Unknown	N/A
	Fixed	
	Variable	
	Manual	
Mark Settings	Mark Flag	Mark Flag
	Mark Level	Mark Level
Dictionary Settings	Word type Maximum length difference	N/A
Votting	RecoStar AEG Recognitoin	N/A
Text Rotation	0, 90, 180, 270	0, 90, 180, 270
Light text on dark background	Yes	N/A
Include characters on zone boundary	Yes	N/A

NOTES:

- “New” means that the feature has never been implemented for ABBYY before. This was added to ABBYY after the OCE removal but is transparent to the user.
- “Normal” means that ABBYY will try to recognize the character as regular machine printed characters. This is the default value for OCR recognition profiles.
- “Automatic” means that ABBYY will try to recognize the character by searching the best-matched print type among the available print types.
- “N/A” means that selecting one of these options will not affect the recognition results.

Due to these features translation, there may be different results in the recognition process between OCE and ABBYY. If this occurs, it is recommended to do the followings in order to get the precise results:

- Adjusting the Print Type selection in Recognition profile settings.
- Adjusting the Image Cleanup tuning in Recognition profile settings.

Kofax Capture 9.0 ABBYY Engine

Application Note

Recognition Profiles Mapping

In the previous Capture Software releases (version 7.x and 8.0), the following are applied:

- All Kofax High Performance Recognition Profiles use OCE RecoStar
- All Kofax Advanced Recognition Profiles use ABBYY FineReader

However, in Kofax Capture 9.0, all Recognition Profiles use the ABBYY engine unless the OCE engine is installed and licensed (optional). For customers that upgrade their system to Kofax Capture 9.0:

- Kofax High Performance Recognition Profiles that are set within batch classes will be mapped to use the Kofax Advanced Recognition Profile.
- This mapping occurs only to the engine that runs behind the Recognition Profile. The Recognition Profile names are kept the same.

Character Type Scope

The following are the expected results when running High Performance recognition profile using ABBYY engine:

Source	OCE Numeric	OCE Amount	ABBYY Numbers
0123456789# \$ *+,./<>	0123456789#\$*+,./<>	0123456789#\$*+,./<>	0123456789 ^{^^} - [^] +,-./ 0
\$1,987.65	\$1,987.65	\$1,987.65	\$1,987.65
1+2 = 3	1+2 = 3	1+2 = 3	1+2 = 3
4*5 = 20	4*5 = 20	4*5 = 20	4 [^] 5 = 20
6-7= X	6-7=X	6-7=X	6-7=)(
8/4= x	8/4=X	8/4=X	8/4=)(
AB 123 CD \$Xx	8123\$60XX	8123\$60XX	/ [^] 6123 [^] 00)()(
3>2<4	3>2<4	3>2<4	3 [^] 2 [^] 4
#890,012.45	#890,012.45	#890,012.45	#890,012.45

Note: The characters in red show the difference between ABBYY and OCE.

In processing numbers, ABBYY does not support the following:

- <> sign.
- * sign.
- # sign and \$ sign if not the first character. However, if it is in the first character, then ABBYY will show it.
- X or x character.

Kofax Capture 9.0 ABBYY Engine

Application Note

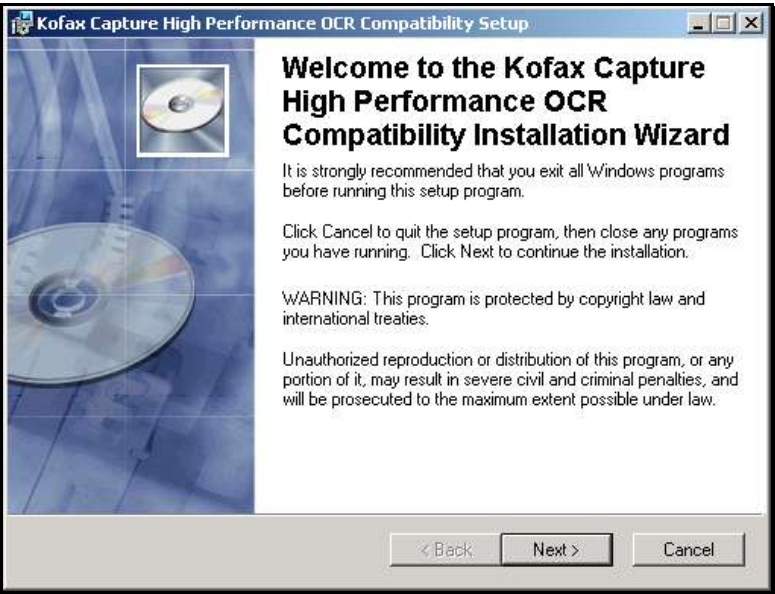
Known Issues

The following are known issues with ABBYY during the recognition process that will be addressed in the future:

- SPR 52585: ABBYY recognizes the text "is" as "IS" (upper case).
- SPR 52943: Period character is not being recognized when zone testing a particular sentence. For example: "This is some sample text for testing." is recognized as "This is some sample text for testing".

Optional OCE RecoStar Installation

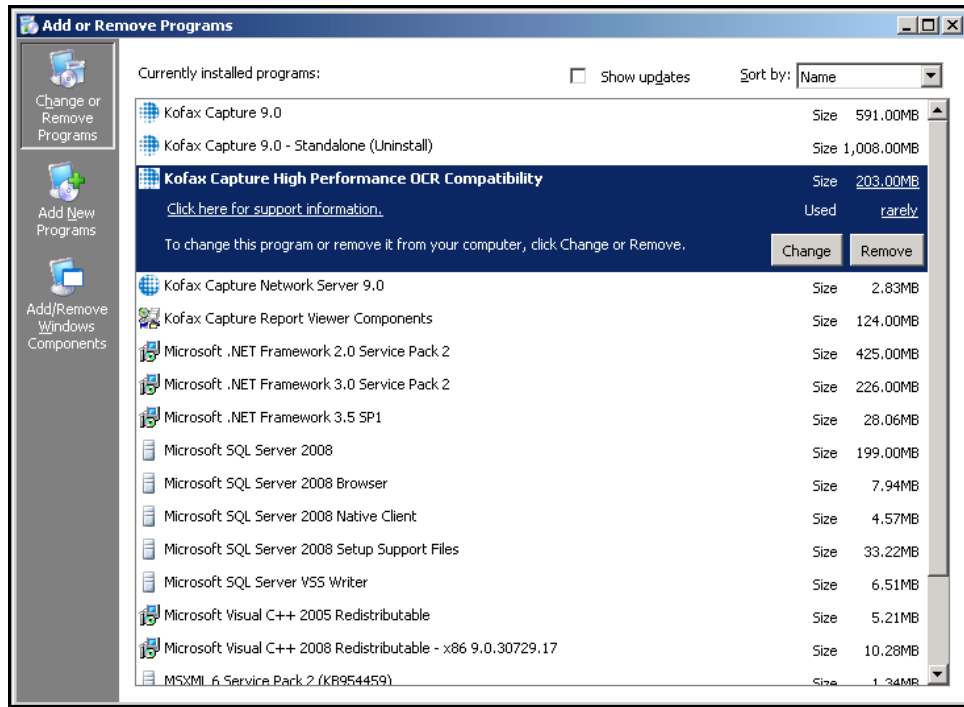
In Kofax Capture 9.0, the OCE RecoStar engine is included on the DVD as an optional installation. Additional license is required to enable this feature.



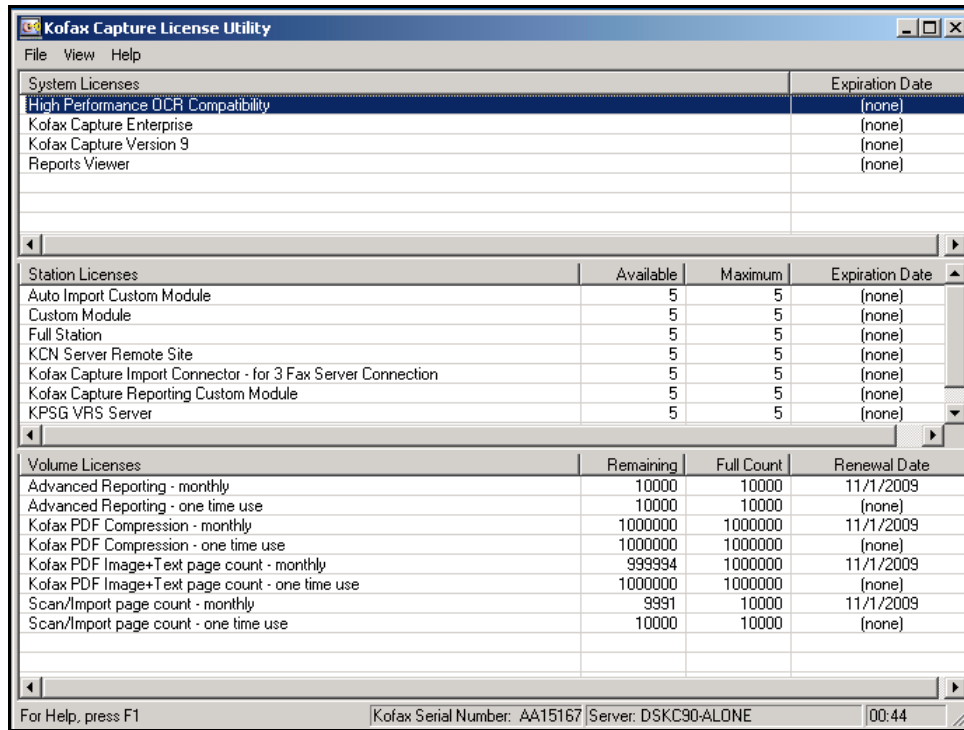
Kofax Capture 9.0 ABBYY Engine

Application Note

OCE optional engine can be uninstalled from Control Panel | Add or Remove programs.



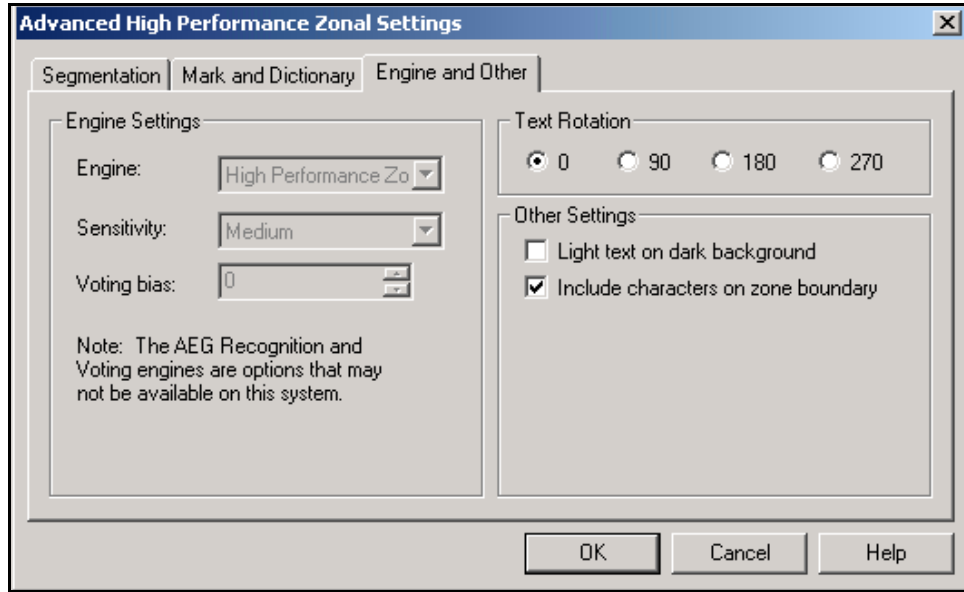
OCE optional engine licensing will be controlled by the Kofax License server. It needs to be licensed with High Performance OCR Compatibility System License.



Kofax Capture 9.0 ABBYY Engine

Application Note

If OCE is not installed nor licensed, the Engine Settings in Advanced High Performance Zonal Settings is disabled / grayed out.



Application Note Summary

Both ABBYY and OCE engines have their unique characteristics. Because of this, Kofax Capture 9.0 users who are using ABBYY may need to readjust their batch class recognition profiles settings, especially for batch classes that are already optimized for OCE, in order to achieve the most precise results.