

API Guide

General Info

Service	Method	URL	Encoding
XML	POST	252.231.51.53/api/ocr/xml/	UTF-8
OCR	POST	252.231.51.53/api/ocr/ocr/	UTF-8

Request Parameter (form-data)

- 252.231.51.53/api/ocr/xml/

Key	Type	Necessary	Description
job_id	UUID4	Y	Running Job ID
lang	language	Y	Target language to do OCR
file	file (file name should be well formatted)	Y	Input file
extension	'pdf' (fixed argument)	Y	Output file extension
xml_type	'page' or 'letter' (only one)	Y	Output xml file main hierarchy

- 252.231.51.53/api/ocr/ocr/

Key	Type	Necessary	Description
job_id	UUID4	Y	Running Job ID
lang	language	Y	Target language to do OCR
file	file (file name should be well formatted)	Y	Input file
extension	extension	Y	Output file extension

Response

- Header

Key	Description
job_id	Output Job ID
status	Result status
message	Result message corresponds with status

- Body

Type	Description
File (XML)	Output XML File

Status/Message List

Status	Message
000	Done

Status	Message
001	Input value error
010	Empty input file
100	Runtime error
300	URL error
500	Request type error

Extension List

- docx
- doc
- rtf
- odt
- pdf
- htm
- txt
- xlsx
- xlx
- pptx
- csv
- fb2
- epub
- djvu

Language List

- [Natural Languages](#)
- [Artificial Languages](#)
- [Formal Languages](#)

Natural languages

- Abkhaz
- Adyghe
- Afrikaans
- Agul
- Albanian
- Altai
- Arabic (Saudi Arabia)
- Armenian (Eastern, Western, Grabar)*
- Avar
- Aymara
- Azeri (Cyrillic), Azeri (Latin)
- Bashkir*
- Basque
- Belarusian
- Bemba
- Blackfoot
- Breton
- Bugotu
- Bulgarian*
- Buryat
- Catalan*
- Cebuano
- Chamorro
- Chechen

- Chinese Simplified, Chinese Traditional
- Chukchee
- Chuvash
- Corsican
- Crimean Tatar
- Croatian*
- Crow
- Czech*
- Dakota
- Danish*
- Dargwa
- Dungan
- Dutch, Dutch (Belgian)*
- English*
- Eskimo (Cyrillic), Eskimo (Latin)
- Estonian*
- Even
- Evenki
- Faroese
- Fijian
- Finnish*
- French*
- Frisian
- Friulian
- Gagauz
- Galician
- Ganda
- German (New Spelling), German*
- German (Luxembourg)
- Greek*
- Guarani
- Hani
- Hausa
- Hawaiian
- Hebrew*
- Hungarian*
- Icelandic
- Indonesian*
- Ingush
- Irish
- Italian*
- Japanese
- Jingpo
- Kabardian
- Kalmyk
- Karachay-balkar
- Karakalpak
- Kashubian
- Kawa
- Kazakh
- Khakass
- Khanty
- Kikuyu
- Kirghiz
- Kongo
- Korean, Korean (Hangul)
- Koryak
- Kpelle
- Kumyk

- Kurdish
- Lak
- Latin*
- Latvian*
- Lezgi
- Lithuanian*
- Luba
- Macedonian
- Malagasy
- Malay (Malaysian)
- Malinke
- Maltese
- Mansi
- Maori
- Mari
- Maya
- Miao
- Minangkabau
- Mohawk
- Moldavian
- Mongol
- Mordvin
- Nahuatl
- Nenets
- Nivkh
- Nogay
- Norwegian (Bokmal), Norwegian (Nynorsk)*
- Nyanja
- Polish*
- Portuguese, Portuguese (Brazilian)*
- Ojibway
- Ossetian
- Papiamentu
- Occitan
- Quechua (Bolivia)
- Rhaeto-Romanic
- Romanian*
- Romany
- Rundi
- Russian*
- Russian (Old Spelling)
- Russian with accent
- Rwanda
- Sami (Lappish)
- Samoan
- Scottish Gaelic
- Selkup
- Serbian (Cyrillic), Serbian (Latin)
- Shona
- Slovak*
- Slovenian*
- Somali
- Sorbian
- Sotho
- Spanish*
- Sunda
- Swahili
- Swazi
- Swedish*

- Tabasaran
- Tagalog
- Tahitian
- Tajik
- Tatar*
- Thai*
- Tok Pisin
- Tongan
- Tswana
- Tun
- Turkish*
- Turkmen (Cyrillic), Turkmen (Latin)
- Tuvinian
- Udmurt
- Uighur (Cyrillic), Uighur (Latin)
- Ukrainian*
- Uzbek (Cyrillic), Uzbek (Latin)
- Vietnamese*
- Welsh
- Wolof
- Xhosa
- Yakut
- Yiddish
- Zapotec
- Zulu

Artificial Languages

- Esperanto
- Interlingua
- Ido
- Occidental

Formal Languages

- Basic
- C/C++
- COBOL
- Digits
- Fortran
- Java
- Pascal
- Simple chemical formulas

* Dictionary support is available for this language, enabling ABBYY FineReader to identify unreliably recognized characters and detect spelling errors in texts written in this language.