



TexLang Reco Engine API

Introduction

TexLang Reco api is build to extract text from image documents and pdf. This Api will return both extracted text and Error marked text of the particular document. TexLang supports a wide range of image file types in form of base64 string and returns a html output with error tagging. The service can we consumed across any platform.

URL

<http://texlangrecoapi.cloudapp.net/recoervice.svc>

Method

Post

Parameters

- Input file in Base64 string format.
- File type ("pdf", "jpg", "png")
- Unique key which is provided by Megdap for accessing the API
- Language of the Input file(Kannada,Marathi,Telugu, Urdu,English,Malayalam,Hindi,Japanese).

Success Response

Json format

Example:-

```
{
  "ExtractedText": "This is just a Testing"

  "HTMLData": "\n\n<html xmlns=\"http://www.w3.org/1999/xhtml\" xml:lang=\"en\"
lang=\"en\">\n <head>\n <title></title>\n<meta http-equiv=\"Content-Type\"
content=\"text/html; charset=utf-8\" />\n <strong><em>This</em></strong></span> <span>
class='ocrx_word' id='word_2_132' title='bbox 806 2890 844 2912;
x_wconf=91'><strong><em>ls</em></strong></span> <span class='ocrx_word'
id='word_2_133' title='bbox 866 2890 922 2912;
x_wconf=96'><strong><em>Just</em></strong></span> \n </span>\n </p>\n\n <p
class='ocr_par' id='par_2_28' lang='eng' title=\"bbox 456 2928 897 2950\">\n <span
class='ocr_line' id='line_2_45' title=\"bbox 456 2928 897 2950; baseline 0 -3; x_size 30.5;
x_descenders 5.5; x_ascenders 5.5\"><span class='ocrx_word' id='word_2_134' title='bbox
456 2928 609 2947; x_wconf=91'><strong><em>Testing</em></strong></span> </span>\n
</p>\n </div>\n </div>\n </body>\n</html>\n"
```

API Call Sample Code

C#

```
Service1Client sc = new Service1Client();

Byte[] bytes = File.ReadAllBytes("path to your file");

String file = Convert.ToBase64String(bytes);

string result =sc.GetExtratedText("key",file,Formats.jpg,Language.English);
```

Java

```
File f = new File("path to your file ");

String encodstring = encodeFileToBase64Binary(f);

String str = Webtest.getExtratedText("key",encodstring Formats.jpg,Language.English);
```

