AutoOCR JavaScript Binding for Alfresco

With the JavaScript API you have a direct access and can use the AutoOCR service from Alfresco Scripts. From the Repository JavaScripts (WebScript-Controller Scripts, Scripted Actions) all functions of the AutoOCR API´s can be called and used. The API is completely independent from the integration of the AutoOCR Service as Alfresco transformer.

Script Object and Configuration

The name of the script object is “autoocr”. After the installation the object is automatic available for all scripts. The configuration can be done using a JavaScript Object-Literal '{ ... }' like it is known from many JavaScript Frameworks and is always the first parameter of each method.

Use-pattern as sample:

```
var config = {
    endpoint : "https://{AUTOOCRHOST UND PORT}/AutoOCRService",
    username : "user",
    password : "password",
    apikey   : 12455
};

autoocr.{APIMETHODE1}(config, param1, param2, param2, ...);
autoocr.{APIMETHODE2}(config, param1, param2, param2, ...);

```

Get parameter

```
var config = {
    endpoint : "https://{AUTOOCRHOST UND PORT}/AutoOCRService",
    username : "user",
    password : "password",
    apikey   : 12455
};

var nrdocs = autoocr.getNrOfDocumentsInQueue(config);

var nrpages = autoocr.getNrOfPagesInQueue(config);

var availpages = autoocr.getAvailablePages(config);

var avgpages = autoocr.getAvgSecPerPage(config);
```


X KEY GmbH, Galvanigasse 2, 1210 Wien ● UID: ATU 64070267 ● T: +43 1 278 43 18 0 ● F: +43 1 278 43 18 10 ● office@xkey.at ● www.xkey.at
Bankverbindung: Raiffeisenlandesbank NÖ-Wien AG ● BLZ: 32000 ● Konto: 201.590 ● IBAN: AT413200000000201590 ● BIC: RLNWATWW

Upload and Download of Files (transformations)

```
1
2  // 1. upload file from repository
3  var sourceNode = companyhome.childByNamePath("tester.pdf");
4  var jobDesc = {
5      sourcenode : sourceNode,
6      autoocrext : 'PDF',
7      settingsname : 'AbbyyFR10',
8          label: "this is my upload of document ...."
9  };
10
11 var job = autoocr.upload(config, jobDesc);
12
13 // 2. wait for job completion
14 var result = autoocr.waitForCompletion(config, job.guid);
15 if(result != autoocr.STATUS_CONVERTED) {
16   throw "transformation timeout or transformation error";
17 }
18
19 // 3. download transformed file to repository
20 var targetNode = companyhome.createFile("transformationtarget.pdf");
21
22 var dlDesc = {
23     guid : job.guid,
24     targetnode : targetNode
25 };
26 autoocr.download(config, dlDesc);
27
```

Upload using mimetypes

```
1 var sourceNode = companyhome.childByNamePath("tester.pdf");
2 var jobDescMimetype = {
3     sourcenode : sourceNode,
4     targetmimetype : 'application/pdf',
5     settingsname : 'AbbyyFR10',
6         label: "upload per mimetype"
7 };
8 ...
```

Read settings

```
1  var settings = autoocr.getSettingsCollection(config);
2  for(var i = 0; i < settings.length; i++) {
3    var setting = settings[i];
4    logger.log(setting.settingsname);
5    logger.log(setting.description);
6    logger.log(setting.enginename);
7    for(var k = 0; k < setting.additionaloutputs.length; k++) {
8      logger.log(setting.additionaloutputs[k]);
9    }
10   for(var k = 0; k < setting.inputformats.length; k++) {
11     logger.log(setting.inputformats[k]);
12   }
13   for(var k = 0; k < setting.inputmimetypes.length; k++) {
14     logger.log(setting.inputmimetypes[k]);
15   }
16   for(var k = 0; k < setting.outputformats.length; k++) {
17     logger.log(setting.outputformats[k]);
18   }
19   for(var k = 0; k < setting.outputmimetypes.length; k++) {
20     logger.log(setting.outputmimetypes[k]);
21   }
22 }
```


X KEY GmbH, Galvanigasse 2, 1210 Wien ● UID: ATU 64070267 ● T: +43 1 278 43 18 0 ● F: +43 1 278 43 18 10 ● office@xkey.at ● www.xkey.at
Bankverbindung: Raiffeisenlandesbank NÖ-Wien AG ● BLZ: 32000 ● Konto: 201.590 ● IBAN: AT413200000000201590 ● BIC: RLNWATWW

Job- and Status - query

```
var job = autoocr.getJob(config, "b525042c-0789-4e67-ad65-991dea92fa35");

logger.log(job.guid);
logger.log(job.jobid);
logger.log(job.pagecount);
logger.log(job.status);
logger.log(job.error);
logger.log(job.owner);
logger.log(job.label);

var status = autoocr.getStatus(config, job.guid);

if(status == 1)
  logger.log("CREATED");

```

Perform multiple transformations

The AutoOCR service is able to perform multiple transformations of the source file with one processing step (but only if the selected and configured OCR engine = OCR profile – supports this feature). It is not possible to upload multiple source files in one processing step. It is only possible to process and transform one input file to one or multiple output files. If multiple destination formats are needed the upload-parameter **`additionaloutputs`** can be used. The result files must be downloaded one after the other, and the download-parameter **`index`** can be used to get the different results. The parameter **`index`** is 0-based and is counted up to the max. number of results. The number of results and also the assignment to the formats must be checked with the calls of **`getResultCount`** and **`getResultExt.`** The last has to be performed, because the result list is not sorted in the same way like the input list from **`additionaloutputs`**. Because it also can be that errors happen during the conversion, it can happen that the result is lesser than the input list.

```
var jobDesc, job, resultCount, i, config, ext, targetNode, someContentNode, dlDesc;

jobDesc = {
  sourceNode: someContentNode,
  autoocrext: 'PDF', // alternativ kann hier auch targetmimetype verwenden.
  additionaloutputs: ['DOCX','PPT'], // alternativ können hier auch mimetypes verwendet werden
  settingsname: 'AbbyyFR10'
};

job = autoocr.upload(config, jobDesc);

autoocr.waitForCompletion(config, job.guid);

resultCount = autoocr.getResultCount(config, job.guid);

for(i = 0; i < resultCount; i++) {
  ext = autoocr.getResultExt(config, job.guid, i);
  targetNode = companyhome.createFile("autoocrtransformed-" + i + "." + ext);
  dlDesc = {
    guid : job.guid,
    index: i,
    targetnode : targetNode
  };
  autoocr.download(config, dlDesc);
}

```


X KEY GmbH, Galvanigasse 2, 1210 Wien ● UID: ATU 64070267 ● T: +43 1 278 43 18 0 ● F: +43 1 278 43 18 10 ● office@xkey.at ● www.xkey.at
Bankverbindung: Raiffeisenlandesbank NÖ-Wien AG ● BLZ: 32000 ● Konto: 201.590 ● IBAN: AT413200000000201590 ● BIC: RLNWATWW

Query AutoOCR user

This call returns a complete list of all available accounts that are registered at the AutoOCR Server. The permission property is a bit-field.

```
var users = autoocr.getUsers(config);
for(var i = 0; i < users.length; i++) {
  logger.log("user #" + i + ": name=" + users[i].username + ", permissions=" + users[i].permissions);
}
```

Start & Stop AutoOCR Server

```
autoocr.stopServer(config);
autoocr.startServer(config);
```

Check availability of the AutoOCR Server

```
autoocr.isServerStarted(config);
```

Get AutoOCR Server version

```
var version = autoocr.getServerVersion(config);
```

Get all jobs from AutoOCR

```
var allJobs = autoocr.getAllJobs(config);
for(var i = 0; i < allJobs.length; i++) {
    logger.log("job #" + i + ": label=" + allJobs[i].label);
}
```

Gat all jobs from AutoOCR filtered by status

```
var statusJobs = autoocr.getAllJobs(config, autoocr.STATUS_CONVERTED);
for(var i = 0; i < statusJobs.length; i++) {
    logger.log("job #" + i + ": label=" + statusJobs[i].label);
}
```

Get all own jobs from AutoOCR with status filter

```
var ownJobs = autoocr.getOwnJobs(config, autoocr.STATUS_CONVERTED);
for(var i = 0; i < ownJobs.length; i++) {
    logger.log("job #" + i + ": label=" + ownJobs[i].label);
}
```


X KEY GmbH, Galvanigasse 2, 1210 Wien ● UID: ATU 64070267 ● T: +43 1 278 43 18 0 ● F: +43 1 278 43 18 10 ● office@xkey.at ● www.xkey.at
Bankverbindung: Raiffeisenlandesbank NÖ-Wien AG ● BLZ: 32000 ● Konto: 201.590 ● IBAN: AT413200000000201590 ● BIC: RLNWATWW

Overview of all configuration options

```
var config = {
    endpoint : "https://212.41.224.95:8001/AutoOCRService", // required, this is the http or https endpoint of the AutoOCR REST service
    username : "OCR", // optional, if set enables Basic Auth
    password : "****", // only used if username is set
        apikey: 12345, // required to talk to the AutoOCR Remote API
    credentialsencoding : "UTF-8", // optional defaults to UTF-8
    urlencoding : "UTF-8", // optional, defaults to UTF-8
    connectiontimeout: 10000, // optional, defaults to infinite
    jobtimeout : 60000, // optional, defaults to 120000msec (2min)
    sleeptime : 1000 // optional, defaults to 1000msec (1sec)
};
```

Overview of all uploads options

```
var jobDesc = {
    sourcenode : sourceNode, // required, the node which content is uploaded for transformation to AutoOCR
    sourcecontentproperty : 'cm:content', // optional, if not the standard content, stored in the cm:content property, needs to be uploaded
    autoocrext : 'PDF', // required if 'targetmimetype' is not given, the AutoOCR Extension names the target transformation format
        targetmimetype: 'application/pdf', // required if 'autoocrext' is not given, the mimetype of the requested transformation target
    settingsname : 'AbbyyFR10', // required, the name of the AutoOCR setting which has to do the transformation
    additionaloutputs : [] // optional, an empty list is the default. The AutoOCR service can produce additional outputformats if required and supported by the given AutoOCR set
};
```

Overview of all downloads options

```
var dlDesc = {
    guid : job.guid, // required, the guid of the uploaded job
    index: 0, // optional, zero is the default value. If there where more than one transformations requested, the index tells which of the transformations has to be downloaded
    targetnode : targetNode, // required, at which node the downloaded content should be stored.
    targetcontentproperty : 'cm:content' // optional, if the content should not be stored in the default property cm:content.
};
```

Handling of transformation errors

The status **`autoocr.STATUS_CONVERSION_ERROR`** is sent back (Integer value 6).

Handling of transformations-timeouts

If the transformation takes longer than configured by the **`jobtimeout`** parameter, with the method **`waitForCompletion`** the status **`autoocr.STATUS_CONVERSION_TIMED_OUT`** is sent back (Integer value 129).

Error handling of miscellaneous network- and runtime errors

For network and other runtime errors all methods of the script-client throw exceptions. These can handles from JavaScript with a try-catch block:

```
try {
   autoocr.upload(...);
} catch(ex) {
   logger.log("network problems: " + ex);
}
```


X KEY GmbH, Galvanigasse 2, 1210 Wien ● UID: ATU 64070267 ● T: +43 1 278 43 18 0 ● F: +43 1 278 43 18 10 ● office@xkey.at ● www.xkey.at
Bankverbindung: Raiffeisenlandesbank NÖ-Wien AG ● BLZ: 32000 ● Konto: 201.590 ● IBAN: AT413200000000201590 ● BIC: RLNWATWW

Debug support

To be able to analyze problems on the connection level, the AutoOCR script allows performing a client debugging output based on the http-protocol level. Enabling and disabling can be done dynamic during the runtime. A restart is not necessary.

```
// Aktivierung der HTTP Debug Meldungen
autoocr.enableWireDebug();

// Deaktivierung
autoocr.disableWireDebug();
```

Sample of log-messages:

```
2011-12-27 14:26:58,781  DEBUG [impl.conn.DefaultClientConnection] [main] Sending request: GET /AutoOCRService/Ge
2011-12-27 14:26:58,782  DEBUG [apache.http.headers] [main] >> GET /AutoOCRService/GetResultEx?jobID=b525042c-078
2011-12-27 14:26:58,783  DEBUG [apache.http.headers] [main] >> Authorization: Basic
2011-12-27 14:26:58,783  DEBUG [apache.http.headers] [main] >> Connection: close
2011-12-27 14:26:58,784  DEBUG [apache.http.headers] [main] >> Host: 212.41.224.95:8001
2011-12-27 14:26:58,784  DEBUG [apache.http.headers] [main] >> User-Agent: Apache-HttpClient/4.1.1 (java 1.5)
2011-12-27 14:26:59,209  DEBUG [impl.conn.DefaultClientConnection] [main] Receiving response: HTTP/1.1 200 OK
2011-12-27 14:26:59,209  DEBUG [apache.http.headers] [main] << HTTP/1.1 200 OK
2011-12-27 14:26:59,209  DEBUG [apache.http.headers] [main] << Content-Length: 95481
2011-12-27 14:26:59,209  DEBUG [apache.http.headers] [main] << Content-Type: application/octet-stream
2011-12-27 14:26:59,210  DEBUG [apache.http.headers] [main] << Server: Microsoft-HTTPAPI/1.0
2011-12-27 14:26:59,210  DEBUG [apache.http.headers] [main] << Date: Tue, 27 Dec 2011 13:26:55 GMT
2011-12-27 14:26:59,210  DEBUG [apache.http.headers] [main] << Connection: close
```

From this link you can download a comprehensive sample script:

http://www.may.co.at/webrr/respfile.php?file=MAYComp/Current/RO/ifresco_AutoOCR_transformer_scripttest.zip

X KEY GmbH, Galvanigasse 2, 1210 Wien ● UID: ATU 64070267 ● T: +43 1 278 43 18 0 ● F: +43 1 278 43 18 10 ● office@xkey.at ● www.xkey.at
Bankverbindung: Raiffeisenlandesbank NÖ-Wien AG ● BLZ: 32000 ● Konto: 201.590 ● IBAN: AT413200000000201590 ● BIC: RLNWATWW