
This is motivated by an email message to R-help entitled "Modifying output to Google Docs" by Ajay Ohri. The idea is to be able to interact with Google Documents from within R. Details are available from http://code.google.com/apis/documents/developers_guide_protocol.html. After we get the document, we can manipulate it with XML or whatever tools are appropriate for the format of the document. But we focus here on accessing the service. This uses HTTPs and some authentication.

```
library(RCurl)
library(XML)
```

See <http://code.google.com/apis/accounts/docs/AuthForInstalledApps.html>

```
ans =
  getForm("https://www.google.com/accounts/ClientLogin",
    accountType = "HOSTED_OR_GOOGLE",
    Email = "dtemplelang@gmail.com",
    Passwd = gpasswd,
    service = "writely",
    source = "R-GoogleDocs-0.1",
    .opts = list(ssl.verifypeer = FALSE))
```

getForm or postForm work fine. We extract the settings from this with the following code:

```
getGoogleAuth =
function(ans)
{
  x = unlist(strsplit(ans, "\\n"))
  tmp = strsplit(x, "=")
  structure(sapply(tmp, `[`, 2), names = sapply(tmp, `[`, 1))
}
```

We are looking for the Auth field. (The SID and LSID are not currently used.)

```
auth = getGoogleAuth(ans)["Auth"]
```

From now on, each HTTPS request to the Google API should have an HTTP header field

Authorization: GoogleLogin auth=yourAuthValue

So we create a new Curl handle

```
curl = getCurlHandle(httpheader = c(Authorization = paste("GoogleLogin auth=", auth)))
x = getURL("http://docs.google.com/feeds/documents/private/full", curl = curl)
doc = xmlParse(x, asText = TRUE)
```

How many entries are there?

```
entries = getNodeSet(doc, "//w:entry", "w")
length(entries)
```

What are their names?

```
xpathSApply(doc, "//g:entry/g:title", xmlValue, namespaces = "g")
```

Which are documents and which are spreadsheets?

```
xpathSApply(doc, "//g:entry/g:category/@label", namespaces = "g")
```

When were these last modified?

```
strptime(xpathSApply(doc, "//g:entry/g:updated", xmlValue, namespaces = "g"),
"%Y-%m-%dT%H:%M:%S")
```

Let's download the first one

```
gdoc = getURL(xmlGetAttr(entries[[1]][["content"]], "src"), curl = curl)
(Don't forget to use the curl object with the Authorization header.) This is an HTML document.
```

```
hdoc = htmlParse(gdoc)
```

Let's modify this document slightly. Since the HTML parsing doesn't work, we'll use a regular expression for now.

```
gdoc = gsub("This uses information from", "Working from", gdoc)
```

Now we upload the document `postForm("http://docs.google.com/feeds/documents/private/full", .opts = list(httpheader = c(c(Authorization = paste("GoogleLogin auth=", auth, sep = "")), Slug = "My Sample Doc from R")), 'x' = fileUpload("Rdummy", gdoc, "google.doc"), curl = curl)`

```
id = xmlGetAttr(entries[[1]][names(entries[[1]]) == "link"][[3]], "href")
# "http://docs.google.com/feeds/documents/private/full/document%3Adfwhmfk3_2gwrnvd"
curlPerform(customrequest = "DELETE", url = id, curl = curl)
```