Document Management Storage System (DMSS)

Overview

DMSS is intended to be a service that provides storage to an indexing system. Its entire function is to receive files, securely place them on the server and assign a URI that can be used to retrieve it or, receive a URI and securely transfer the file to the client computer.

DMSS comprises two programs. One is a server application (docServer) that receives network transfers of files and parameters. The other is a client application (docClient) that resides on a user's computer and conducts the actual sending and receiving of files. docClient is a normal computer application installed as many computers for as many users as needed. docServer is a special program installed on a single computer.

DMSS does not provide any services to help find files other than with the URI. To find out the URI for a particular file requires a separate application called an indexing agent that provides context, data and indexes to help a user identify stored document files.

Because DMSS is a service, controlling access to files is done with special parameters sent from the indexing agent. To store a file in the docServer, the indexing agent sends parameters, including the file's disk location, to the docClient application. docClient conveys the file to the docServer which assigns a URI and reports it to the indexing agent. To retrieve a file, the indexing agent is used to find the URI of a file and then send it to the docClient program which forwards it to the docServer. As a result, the file referenced by the URI is put on the client computer.

The only functions of the docClient are to send files to docServer (and thereby update the indexing agent with a URI) or to send a URI to docServer and get files. The docServer, in addition to receiving files and creating a URI and receiving a URI and sending files, also maintains a comprehensive log of every interaction with the system, ie, what computers made requests and other details intended to represent an auditable log that allows tracking of the location and access of every file.

Indexing Agent Protocols

Seven parameters are defined for the interaction between an indexing agent and a docClient. Two represent the specific file (ds_uri and ds_filepath) and one uniquely identifies this transaction (ds_refid). The rest are security and user permission information. (These are, for now, static data based on encryption technologies. In the future, more elaborate, one-time-use parameters will be used to increase security.)

When a file is loaded into the docServer, data is sent to the indexing agent. The main point of this data is to send the URI (ds_uri) but, for the indexing agent to be able to provide context, it needs a unique identifier, a reference ID (ds_refid). For convenience, the name of the file (ds_filename) is also sent (though, in principle, the indexing agent should know this).

For flexibility, the DMSS protocol allows the indexing agent to send additional data in a storage request. This is automatically returned to the indexing agent along with the URI.

Interactions with the docServer are always conducted through the docClient.

DMSS Service Request File

To send or provide access to a file, the indexing agent sends a file with a specific name to a specific location on the client computer disk drive. This file is temporary (the docClient deletes it as soon as it is processed) and is located at the file path

```
/tmp/docStoreRequest.tsv
```

This location is built into the docClient program and cannot be changed.

The file is intended to be exported by the indexing agent as tab separated data with a header row comprising these fields:

```
ds_operation                 upload or download only
ds_uri                       a URI for download operations (blank otherwise)
ds_filepath                  a local file path for upload operations (blank otherwise)
ds_refid                     a unique identifier for this file transaction
ds_authusername              the agent authenticated name of the user
ds_rsakeyserverstaticpublic  security information supplied at setup time
ds_authuserrequesttoken      security information supplied at setup time
```

Any other fields the agent wants sent from the docServer along with the URI.

All of the fields are carried as strings.

Sample files for upload and download are available.

DMSS URI Index Agent Update

When a file has been successfully loaded into the docServer storage and a URI has been generated, it sent to one or more indexing agents. For flexibility, the indexing agent protocol is not built into docServer. Instead it is a plug-in model that has access to all of the details of the file and associated information. Presently, there is only one plug-in.

The Helix Standard Plugin

The helix standard protocol relies on an HTTP POST. Three standard fields and an arbitrary set of custom fields are sent in the body of the post. The standard ones are:

```
ds_uri                   the unique identifier supplied in docStoreRequest.tsv
ds_refid                 the URI generated for the uploaded file
ds_fileName              the name (and extension) of the uploaded file
```

All three of these are strings of arbitrary length. Custom fields are also strings and no error checking is done on them.