## Updating Korp-FCS endpoint

### Running Korp-FCS services on Alf

Korp: <https://alf.hum.ku.dk/korp/>

FCS endpoint: [https://alf.hum.ku.dk/korp/fcs/2.0/endpoint/sru](https://alf.hum.ku.dk/korp/fcs/2.0/endpoint/sru?operation=explain)

### Configurations

* Nginx proxy server (config: /etc/nginx/includes/korp.locations)
* Korp frontend running on port 8079 (config: /etc/apache2/conf-available/korp.conf)
* CGI server (uwsgi) running on port 9001 (config: /etc/uwsgi/cgi\_config.ini)
* Tomcat running on port 8088 with deployed endpoint (*fcs-korp-endpoint.war*) in default *webapps* (config: /opt/apache-tomcat-9.0.14/conf/server.xml), context: *fcs-korp-endpoint*

Note: *tomcat* user owns the /opt/apache-tomcat-9.0.14/ directory and runs the local server.

### Source project

Git project: <https://git.sc.ku.dk/kuhumcst/fcs-korp-endpoint/tree/clarin-dk> (branch: *clarin-dk*)

Source code installed at location (Alf): /opt/fcs-korp-endpoint

Note: *master* is upstream branch - <https://github.com/clarin-eric/fcs-korp-endpoint/>

### FCS operations

[See CLARIN-FCS to SRU/CQL binding](https://trac.clarin.eu/wiki/Taskforces/FCS/FCS-Specification-Draft#CLARIN-FCStoSRUCQLbinding) (FCS 2.0 Specification)

#### explain

The *explain* operation of the SRU protocol serves to announce server capabilities and to allow clients to configure themselves automatically. This operation is used similarly. *x-fcs-endpoint-description* (default: false) - optional parameter to retrieve the [Endpoint Description](https://trac.clarin.eu/wiki/Taskforces/FCS/FCS-Specification-Draft#EndpointDescription)

<https://alf.hum.ku.dk/korp/fcs/2.0/endpoint/sru?operation=explain&x-fcs-endpoint-description=true>

#### searchRetrieve

The *searchRetrieve* operation of the SRU protocol is used for searching in the Resources that are provided by the Endpoint.

[https://alf.hum.ku.dk/korp/fcs/2.0/endpoint/sru?operation=searchRetrieve&query=%5Bword%3D%22man\*%22%5D&queryType=fcs](https://alf.hum.ku.dk/korp/fcs/2.0/endpoint/sru?operation=searchRetrieve&query=%5Bword%3D%22man*%22%5D&queryType=fcs)

Note: queryType parameter can be *fcs* or *cql* (two supported querying formats)

### Customising Korp-FCS Endpoint to allow new corpora

Note: Any edits to the endpoint should be maintained on the separate Git branch (*clarin-dk*).

A list of *MODERN\_CORPORA* is maintained in the source code (./src/main/java)[[1]](#footnote-1).

This list is a pair string match to the list of available corpora in Korp see <https://alf.hum.ku.dk/korp/cgi/korp.cgi?command=info> for a full list.

The endpoint relies on a clear match in corpus names[[2]](#footnote-2), and will query all corpora made available in the list in their supported layers (i.e. lemma, pos)

#### Adding or removing corpora

Update the corpora list in *ServiceInfo.java* (package: se.gu.spraakbanken.fcs.endpoint.korp) and re-build (deploy) the project to Tomcat as described below.

#### Editing Endpoint Description

The *Endpoint Description*[[3]](#footnote-3) configured in <https://git.sc.ku.dk/kuhumcst/fcs-korp-endpoint/blob/clarin-dk/src/main/webapp/WEB-INF/endpoint-description.xml> should be edited to include the list of available corpora in the endpoint i.e. corpora in Korp made available to the endpoint.

The *Resources* description is a hierarchical true of corpora and sub-corpora (*Resource*). Include elements Title, Description, Languages. Include a valid landing-page URL in the Resource description if it exists. Include AvailableDataViews and AvailableLayers (as defaults) as they apply to the query layers in Korp.

Note: *Resource@pid* could be a valid PID handle or a stub-value.

### 

### Building or Updating the Korp-FCS Endpoint Project

The project is using the [Maven WAR plugin](https://maven.apache.org/plugins/maven-war-plugin/).

To build the source code run from the source root directory:

mvn *war:war*

A new WAR artifact is created at: *./target/fcs-korp-endpoint-0.1.1.war*

Note: project *version* defined in *pom.xml* (Maven) is appended to the filename

Change the filename before copying to the Tomcat *webapps* deploy directory.

Copy the new WAR file to the Tomcat webapps (/opt/tomcat/webapps/fcs-korp-endpoint.war).

Restart the Tomcat service using:

sudo systemctl *restart tomcat*

Refresh the page at <https://alf.hum.ku.dk/korp/fcs/2.0/endpoint/sru?operation=explain&x-fcs-endpoint-description=true>

Review the *Endpoint Description* and retry a search query.

Alternatively, the query results can be tested and viewed in the CLARIN FCS Aggregator but note that it may take some time before the Aggregator is updated with the new *Endpoint Description*.

### CLARIN FCS Aggregator

[https://spraakbanken.gu.se/ws/fcs/2.0/aggregator/#](https://spraakbanken.gu.se/ws/fcs/2.0/aggregator/)

Use the web application to search and query FCS endpoints and/or filter by selected corpora/sub-corpora.

Further details can be viewed in the Statistics log page [https://spraakbanken.gu.se/ws/fcs/2.0/aggregator/stats#](https://spraakbanken.gu.se/ws/fcs/2.0/aggregator/stats)

1. <https://git.sc.ku.dk/kuhumcst/fcs-korp-endpoint/blob/clarin-dk/src/main/java/se/gu/spraakbanken/fcs/endpoint/korp/data/json/pojo/info/ServiceInfo.java#L40> [↑](#footnote-ref-1)
2. <https://git.sc.ku.dk/kuhumcst/fcs-korp-endpoint/blob/clarin-dk/src/main/java/se/gu/spraakbanken/fcs/endpoint/korp/data/json/pojo/info/ServiceInfo.java#L203> [↑](#footnote-ref-2)
3. <https://trac.clarin.eu/wiki/Taskforces/FCS/FCS-Specification-Draft#EndpointDescription> [↑](#footnote-ref-3)