



CLARIN-D



CLARIN-D FEDERATED CONTENT SEARCH AGGREGATOR

Yana Panchenko, Thomas Zastrow



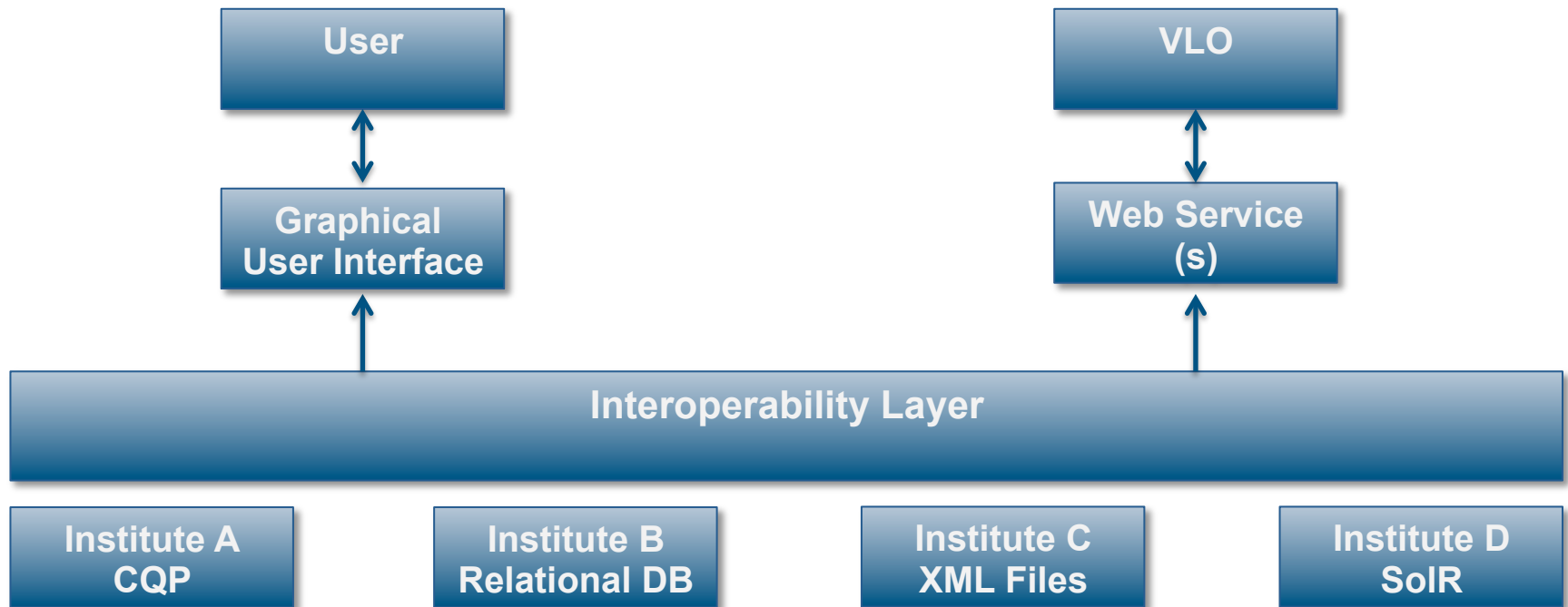
1010111001101010101010

SPONSORED BY THE

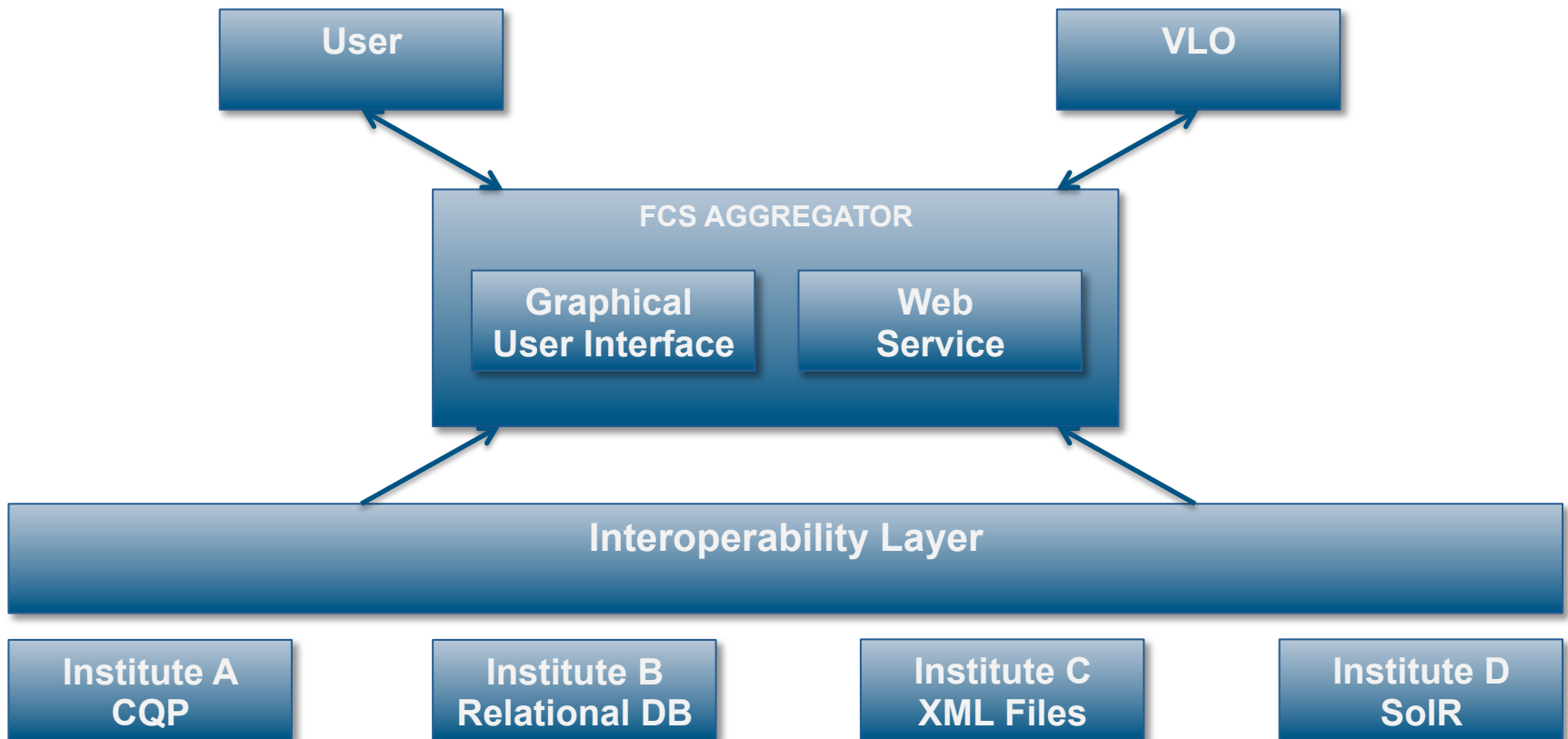


Federal Ministry
of Education
and Research

Aggregator in FCS infrastructure



Aggregator in FCS infrastructure





CLARIN-D FEDERATED CONTENT SEARCH

Actions Help

Elefant

Search

Welcome to the CLARIN-D Federated Content Search!

Corpora

+ Select All - Deselect All

- ▲ Eberhard Karls Universität Tübingen
 - ▶ <http://weblicht.sfs.uni-tuebingen.de/rws/sru/>
- ▲ ASV Leipzig
 - ▶ <http://clarinws.informatik.uni-leipzig.de:8080/CQL>
- ▲ Bayerisches Archiv für Sprachsignale
 - <http://clarin.phonetik.uni-muenchen.de/BASSRU/>
- ▲ Institut für Deutsche Sprache
 - ▶ <http://clarin.ids-mannheim.de/cosmassru>
 - ▶ <http://clarin.ids-mannheim.de/digibibsrु>
- ▲ MPI for Psycholinguistics
 - ▶ <http://cqiservlet.mpi.nl/>
- ▲ Universität des Saarlandes
 - ▶ <http://fedora.clarin-d.uni-saarland.de/sru/>
- ▲ Berlin-Brandenburg Academy of Sciences and Humanities
 - http://dspin.dwds.de/cgi-bin/FCS-Endpoint/DTA_SRU

All 9 Clarin-D Centers support FCS via SRU/CQL

```
▼<Centers>
  ▼<CenterProfile>
    <Centername>Eberhard Karls Universität Tübingen</Centername>
    <Center_id>1</Center_id>
    <Center_id_link>http://130.183.206.32/restxml/1</Center_id_link>
  </CenterProfile>
  ▼<CenterProfile>
    <Centername>ASV Leipzig</Centername>
    <Center_id>2</Center_id>
    <Center_id_link>http://130.183.206.32/restxml/2</Center_id_link>
  </CenterProfile>
  ▼<CenterProfile>
    <Centername>Bayerisches Archiv für Sprachsignale</Centername>
    <Center_id>3</Center_id>
    <Center_id_link>http://130.183.206.32/restxml/3</Center_id_link>
  </CenterProfile>
  ▼<CenterProfile>
    <Centername>Institut für Deutsche Sprache</Centername>
    <Center_id>4</Center_id>
    <Center_id_link>http://130.183.206.32/restxml/4</Center_id_link>
  </CenterProfile>
  ▼<CenterProfile>
    <Centername>MPI for Psycholinguistics</Centername>
    <Center_id>5</Center_id>
    <Center_id_link>http://130.183.206.32/restxml/5</Center_id_link>
  </CenterProfile>
  ▼<CenterProfile>
    <Centername>Universität des Saarlandes</Centername>
    <Center_id>6</Center_id>
    <Center_id_link>http://130.183.206.32/restxml/6</Center_id_link>
  </CenterProfile>
  ▼<CenterProfile>
    <Centername>
      Berlin-Brandenburg Academy of Sciences and Humanities
    </Centername>
    <Center_id>7</Center_id>
    <Center_id_link>http://130.183.206.32/restxml/7</Center_id_link>
  </CenterProfile>
</Centers>
```



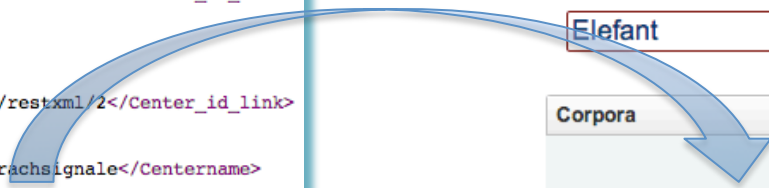
CLARIN-D FEDERATED CONTENT SEARCH

- Corpora
-
- ▶ Eberhard Karls Universität Tübingen
 - ▶ ASV Leipzig
 - ▶ Bayerisches Archiv für Sprachsignale
 - ▶ Institut für Deutsche Sprache
 - ▶ MPI for Psycholinguistics
 - ▶ Universität des Saarlandes
 - ▶ Berlin-Brandenburg Academy of Sciences and Humanities
 - ▶ Hamburger Zentrum für Sprachkorpora (HZSK)
 - ▶ IMS, Universität Stuttgart

Maximum hits

Language

Tier types



Altogether 10 SRU endpoints

```
<CenterBasicInformation>
<CenterBasicInformation>
<CenterBasicInformation>
<CenterBasicInformation>
<CenterBasicInformation>
<CenterBasicInformation>
<CenterBasicInformation>
<CenterBasicInformation>
<CenterBasicInformation>
<CenterBasicInformation>
<Name>
  <CenterBasicInformation>
    <Name>Universität des Saarlandes</Name>
    <!--...-->
    <Type>B</Type>
    <Status>Aiming for B</Status>
    <Description>...</Description>
    <Country>...</Country>
  </CenterBasicInformation>
  <CenterExtendedInformation>
    <Website>http://fedora.clarin-d.uni-saarland.de/</Website>
    <LongTermArchivingPolicy>not yet</LongTermArchivingPolicy>
    <RepositorySystem>Fedora Commons</RepositorySystem>
    <PidStatus>handle via EPIC</PidStatus>
    <StrictVersioning>>false</StrictVersioning>
    <AssessmentStatus>...</AssessmentStatus>
    <WebReference>
      <Website>http://fr46.uni-saarland.de/</Website>
      <Description>
        Website of the faculty of Applied Linguistics
      </Description>
    </WebReference>
    <WebReference>
      <Website>http://fedora.clarin-d.uni-saarland.de/</Website>
      <Description>CQL</Description>
    </WebReference>
    <AAI>
      <AaiStatus>Not available yet</AaiStatus>
      <MemberOfNationalFederation>DE</MemberOfNationalFederation>
      <MemberOfSpf>>true</MemberOfSpf>
    </AAI>
    <Metadata>
      <MetadataScheme>CMDI</MetadataScheme>
      <WebServicesSet/>
    </Metadata>
  </CenterExtendedInformation>
</CenterBasicInformation>
```

Corpora

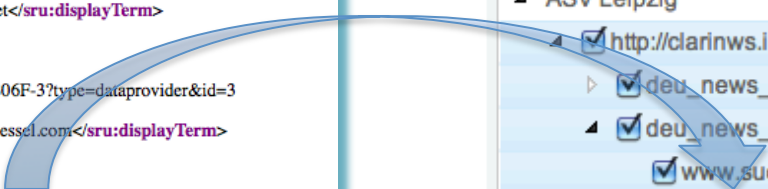
+ Select All - Deselect All

- ▲ Eberhard Karls Universität Tübingen
 - ▶ <http://weblicht.sfs.uni-tuebingen.de/rws/sru/>
- ▲ ASV Leipzig
 - ▶ <http://clarinws.informatik.uni-leipzig.de:8080/CQL>
- ▲ Bayerisches Archiv für Sprachsignale
 - ▶ <http://clarin.phonetik.uni-muenchen.de/BASSRU/>
- ▲ Institut für Deutsche Sprache
 - ▶ <http://clarin.ids-mannheim.de/cosmassru>
 - ▶ <http://clarin.ids-mannheim.de/digibibru>
- ▲ MPI for Psycholinguistics
 - ▶ <http://cqlservlet.mpi.nl/>
- ▲ Universität des Saarlandes
 - ▶ <http://fedora.clarin-d.uni-saarland.de/sru/>
- ▲ Berlin-Brandenburg Academy of Sciences and Humanities
 - ▶ http://dspin.dwds.de/cgi-bin/FCS-Endpoint/DTA_SRU
- ▲ Hamburger Zentrum für Sprachkorpora (HZSK)
 - ▶ <http://virt-fedora.multilingua.uni-hamburg.de:8080/HZSKsru/>
- ▲ IMS, Universität Stuttgart
 - ▶ <http://clarin01.ims.uni-stuttgart.de/SRUCQIBridge>

[endpoint url]/?version=1.2&operation=scan&scanClause=fcs.resource=root

[endpoint url]/?version=1.2&operation=scan&scanClause=fcs.resource=[resource identifier]

```
<?xml version="1.0" encoding="UTF-8" ?>
<sru:scanResponse>
  <sru:scanResponse>
    <sru:version>1.2</sru:version>
    <sru:terms>
      <sru:term>
        <sru:value>
          11858/00-229C-0000-0001-B06F-3?type=dataproducer&id=1
        </sru:value>
        <sru:displayTerm>www.shortnews.de</sru:displayTerm>
      </sru:term>
      <sru:term>
        <sru:value>
          11858/00-229C-0000-0001-B06F-3?type=dataproducer&id=2
        </sru:value>
        <sru:displayTerm>www.faz.net</sru:displayTerm>
      </sru:term>
      <sru:term>
        <sru:value>
          11858/00-229C-0000-0001-B06F-3?type=dataproducer&id=3
        </sru:value>
        <sru:displayTerm>www.brennessel.com</sru:displayTerm>
      </sru:term>
      <sru:term>
        <sru:value>
          11858/00-229C-0000-0001-B06F-3?type=dataproducer&id=4
        </sru:value>
        <sru:displayTerm>www.morgenpost.de</sru:displayTerm>
      </sru:term>
      <sru:term>
        <sru:value>
          11858/00-229C-0000-0001-B06F-3?type=dataproducer&id=5
        </sru:value>
        <sru:displayTerm>tagesschau.sf.tv</sru:displayTerm>
      </sru:term>
      <sru:term>
        <sru:value>
          11858/00-229C-0000-0001-B06F-3?type=dataproducer&id=6
        </sru:value>
        <sru:displayTerm>www.winfuture.de</sru:displayTerm>
      </sru:term>
      <sru:term>
        <sru:value>
          11858/00-229C-0000-0001-B06F-3?type=dataproducer&id=7
        </sru:value>
        <sru:displayTerm>rss.feedsportal.com</sru:displayTerm>
      </sru:term>
    </sru:terms>
  </sru:scanResponse>
</sru:scanResponse>
```



Corpora

+ Select All - Deselect All

- ▲ Eberhard Karls Universität Tübingen
 - ▲ http://weblicht.sfs.uni-tuebingen.de/rws/sru/
 - TuebaDDC
- ▲ ASV Leipzig
 - ▲ http://clarinws.informatik.uni-leipzig.de:8080/CQL
 - ▶ deu_news_2008_10K
 - ▲ deu_news_2008_100K
 - www.sueddeutsche.de
 - www.bverwg.de
 - www.jungewelt.de
 - www.heise.de
 - www.neue-oz.de
 - www.faz.net
 - www.frankenpost.de
 - www.berlinerliteraturkritik.de
 - www.laut.de
 - www.finanztreff.de
 - www.taz.de
 - www.baden-online.de

http://weblicht.sfs.uni-tuebingen.de/rws/sru/?operation=searchRetrieve&query=kind&version=1.2&x-cmd-context=%22http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D%22

```

<sru:searchRetrieveResponse
  xmlns:sru="http://www.loc.gov/zing/srw/"
  <sru:version>1.2</sru:version>
  <sru:numberOfRecords>27</sru:numberOfRecords>
  <sru:records>
    <sru:record>
      <sru:recordSchema>http://clarin.eu/fcs/1.0</sru:recordSchema>
      <sru:recordPacking>xml</sru:recordPacking>
      <sru:recordData>
        <fcs:Resource xmlns:fcs="http://clarin.eu/fcs/1.0"
          pid="http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D"
          ref="http://www.sfs.uni-tuebingen.de/corpora.shtml">
          <fcs:DataView type="application/x-clarin-fcs-kwic+xml">
            <kwic:kwic xmlns:kwic="http://clarin.eu/fcs/1.0/kwic">
              <kwic:c type="left">
                Das Wort Schlag bedeutet so ziemlich alles ; es
                bedeutet unser blow , stroke , dash , hit , shock
                clip , clap , time , bar , coin , stamp ,
              </kwic:c>
              <kwic:kw>kind</kwic:kw>
              <kwic:c type="right">
                , sort , manner , way , apoplexy , woodcutting ,
                enclosure , field , forest-clearing .
              </kwic:c>
            </fcs:DataView>
          </fcs:Resource>
        </sru:recordData>
        <sru:recordPosition>1</sru:recordPosition>
      </sru:record>
    </sru:record>
    <sru:record>
      <sru:recordSchema>http://clarin.eu/fcs/1.0</sru:recordSchema>
      <sru:recordPacking>xml</sru:recordPacking>
      <sru:recordData>
        <fcs:Resource xmlns:fcs="http://clarin.eu/fcs/1.0"
          pid="http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D"
          ref="http://www.sfs.uni-tuebingen.de/corpora.shtml">
          <fcs:DataView type="application/x-clarin-fcs-kwic+xml">
            <kwic:kwic xmlns:kwic="http://clarin.eu/fcs/1.0/kwic">
              <kwic:c type="left">Sie beschreiben das as a</kwic:c>
  
```

Das Wort Schlag bedeutet so ziemlich alles ; es bedeutet unser blow , stroke , dash , hit , shock clip , clap , time , bar , coin , stamp ,	kind	, sort , manner , way , apoplexy , woodcutting , enclosure , field , forest-clearing .
Sie beschreiben das as a	kind	of touchsight , als eine Art Tastsehen , was also so zu verstehen wäre , daß der Geist in sein Sehen gewissermaßen auch das Tasten mit hereinnimmt ; oder wir müßten annehmen , daß er die Zeit ausschalten kann .
Er hat dieses diorama of a very eccentric	kind	nachgezeichnet , Galtons Buch bringt die Zeichnungen .
excuse me , now , Mr. Kerseker , will you be so	kind	, as to give us the boundaries of Ohio ?
Be so	kind	to cut , Sir – seien Sie so freundlich und heben Sie ab !
Ergeben steh ich vor des rätsels macht Wie er mein	kind	ich meines Kindes kind .. Wie sein gesetz ist dass aus erdenstoff Der Hohe wird und eh ihn tat versehrt Mit schmerz und lächeln seinen heimweg nimmt .. Wie sein gesetz ist dass sich der erfüllt Der sich und allen sich zum opfer gibt Und dann die tat mit seinem tod gebiert .

Parallel search in all selected resources



CLARIN-D FEDERATED CONTENT SEARCH

Actions Help

Hallo

Corpora

- ▲ Eberhard Karls Universität Tübingen
 - ▾ <http://weblicht.sfs.uni-tuebingen.de/rws/sru/>
 - TuebaDDC
- ▲ ASV Leipzig
 - ▾ <http://clarinws.informatik.uni-leipzig.de:8080/CQL>
- ▲ Bayerisches Archiv für Sprachsignale
 - <http://clarin.phonetik.uni-muenchen.de/BASSRU/>
- ▲ Institut für Deutsche Sprache
 - ▾ <http://clarin.ids-mannheim.de/cosmassru>
 - Goethe Korpus
 - ▾ <http://clarin.ids-mannheim.de/digibibsr>
 - TextGrid Digital Library (Literature)
- ▲ MPI for Psycholinguistics
 - ▾ [http://cqlservlet.mpi.nl/](http://cqlservlet.mpi.nl)
- ▲ Universität des Saarlandes
 - ▾ <http://fedora.clarin-d.uni-saarland.de/sru/>
- ▲ Berlin-Brandenburg Academy of Sciences and Humanities
 - http://dspin.dwds.de/cgi-bin/FCS-Endpoint/DTA_SRU
- ▲ Hamburger Zentrum für Sprachkorpora (HZSK)
 - <http://virt-fedora.multilingua.uni-hamburg.de:8080/HZSKsru/>
- ▲ IMS, Universität Stuttgart
 - <http://clarin01.ims.uni-stuttgart.de/SRUCQIBridge>

Maximum hits

Language

Tier types

Institut für Deutsche Sprache <http://clarin.ids-mannheim.de/digibibsr> TextGrid Digital Library (Literature) <http://clarin.ids-mannheim.de/digilib/lit> <http://clarin.ids-mannheim.de/digilib/lit>

SACCO . Wahrhaftig ! Winkt .	Hallo	, hierher
	Hallo	, aber vier leere Patronenhülsen habe ich noch .
Die Eingeweide sind wie ausgebrannt . Poltert an die Tür . He ,	hallo	!
Am andern Morgen ging das	Hallo	von neuem los .
	Mit Hallo	wanderte die » schwarze Suse « ins Moor hinaus .
Und gleich darauf ertönte es von der Freitreppe her : » He ,	hallo	, da ist er ja schon - Elsbeth , sieh , da ist er ja schon ! «
	und dann - Hallo	!
	Hallo	, jetzt schwant mir's - !
Leuchtkugeln flogen steil auf , hielten sich ein Weilchen flackernd in		
schwebender Helle , sanken nieder und erloschen blakend neben uns im		
Sande ; sie wurden mit	Hallo	als Bereicherung des nächtlichen Feuerwerks begrüßt .
	Hallo	, Junge ! « schrie er zu einem Fenster hinauf , hinter welchem er einst
	Hallo	geschlafen hatte .


Eberhard Karls Universität Tübingen <http://weblicht.sfs.uni-tuebingen.de/rws/sru/> TuebaDDC <http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D> <http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D>

	Hallo	!
Männer sollten sich an allen Punkten der Welt sagen :	Hallo	, alter Bursche , was hast du gegen mich ?
	Hallo	– old Jack !
	" Hallo	!
	Hallo	, Warrington !
	Hallo	!
	Hallo	, ihr Kerle !
	" Hallo	, Caballero ! " gebot er , mit einem Satz in den Kahn springend .
	" Hallo	, Prinz ! " tönte es von unten .
Jetzt schlug er die letzte Stufe , jetzt schwang er sich über den Rand auf die		
Oberfläche des Gletschers , und aufatmend tönte sein	Hallo	herüber .

Institut für Deutsche Sprache <http://clarin.ids-mannheim.de/cosmassru> Goethe Korpus <http://clarin.ids-mannheim.de/cosmassru> goethe goethe

Sorry, there were no results!

Save the results of the search...

 **CLARIN-D FEDERATED CONTENT SEARCH**



Actions ▾ Help ▾

Institut für Deutsche Sprache <http://clarin.ids-mannheim.de/digilib/lit>

Personal Workspace Sign In

Username:

Password:

Die Eingeweide sind wie aus

Am andern Morgen ging das Hallo von neuem los .

Mit Hallo wanderte die » schwarze Suse « ins Moor hinaus .

Und gleich darauf ertönte es von der Freitreppe her : » He , hallo , da ist er ja schon - Elsbeth , sieh , da ist er ja schon ! «

und dann - Hallo !

Hallo , jetzt schwant mir's - !

Leuchtkugeln flogen steil auf , hielten sich ein Weilchen flackernd in schwebender Helle , sanken nieder und erloschen blakend neben uns im Sande ; sie wurden mit Hallo als Bereicherung des nächtlichen Feuerwerks begrüßt .

Hallo , Junge ! « schrie er zu einem Fenster hinauf , hinter welchem er einst geschlafen hatte .

Eberhard Karls Universität Tübingen <http://weblicht.sfs.uni-tuebingen.de/rws/sru/TuebaDDC> <http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D>

Hallo !

Männer sollten sich an allen Punkten der Welt sagen : Hallo , alter Bursche , was hast du gegen mich ?

... to process with linguistic tools

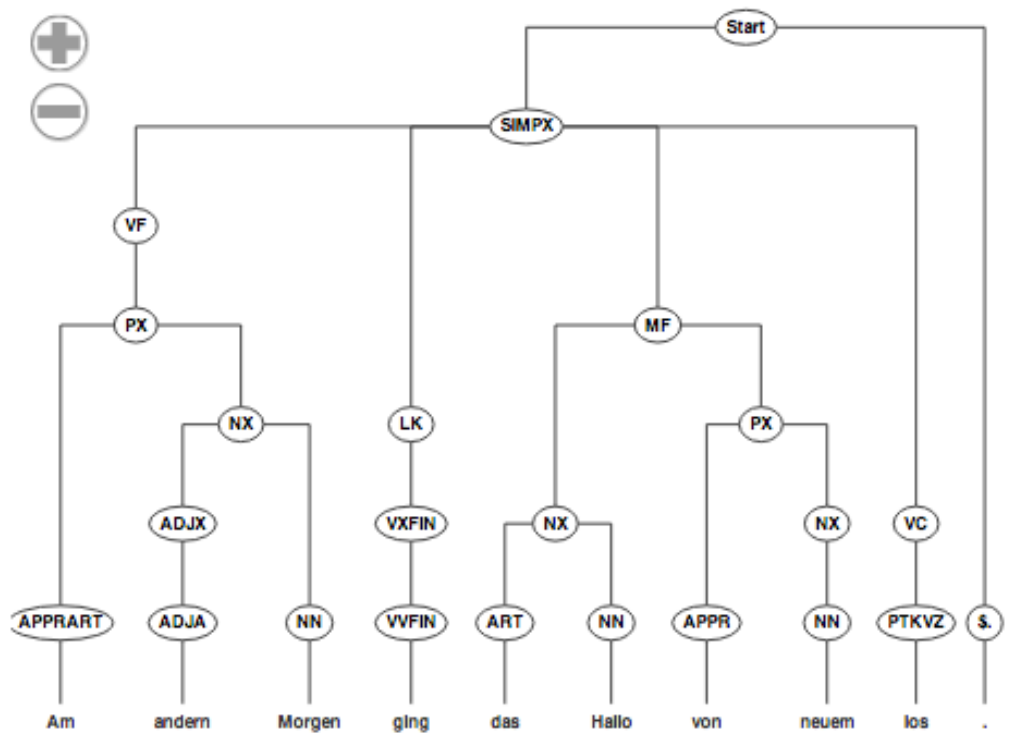
ClarinDFederatedContentSearch12974790022711081410.xml

Annotation Layers: language = de
ClarinDFederatedContentSearch12974790022711081410.xml

Annotation Layers: language = de
ClarinDFederatedContentSearch12974790022711081410.xml

- Annotation Layers:
 - Simple view
 - text
 - sentences
 - Table view
 - tokens
 - POStags
 - lemmas
 - namedEntities
 - Graphical view
 - parsing
 - Highlighted view
 - namedEntities

tagset = tigertb 8 / 27



...
hatte .
Hallo !
Männer sollten :
sagen : Hallo , :
gegen mich ?
Halo & am ...

Integration with other applications: x-aggregation-context parameter



[back to results](#) | [open in original context](#)

< Record 3 out of 7 >

Field	Value
name	TüBa-DDC
id	http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D
collection	Tübingen Language Resources
dataProvider	OLAC Metadata Providers
metadataSource	http://catalog.clarin.eu/oai-harvester/olac-and-dc-providers/harvested/results/cmdj/TDDC.xml
nationalProject	CLARIN-D
organisation	University of Tübingen
projectName	TüBa-DDC, Tübingen Treebank of Written German - Diachronic Corpus, Tübinger Baumbank des Deutschen - Diachrones Corpus

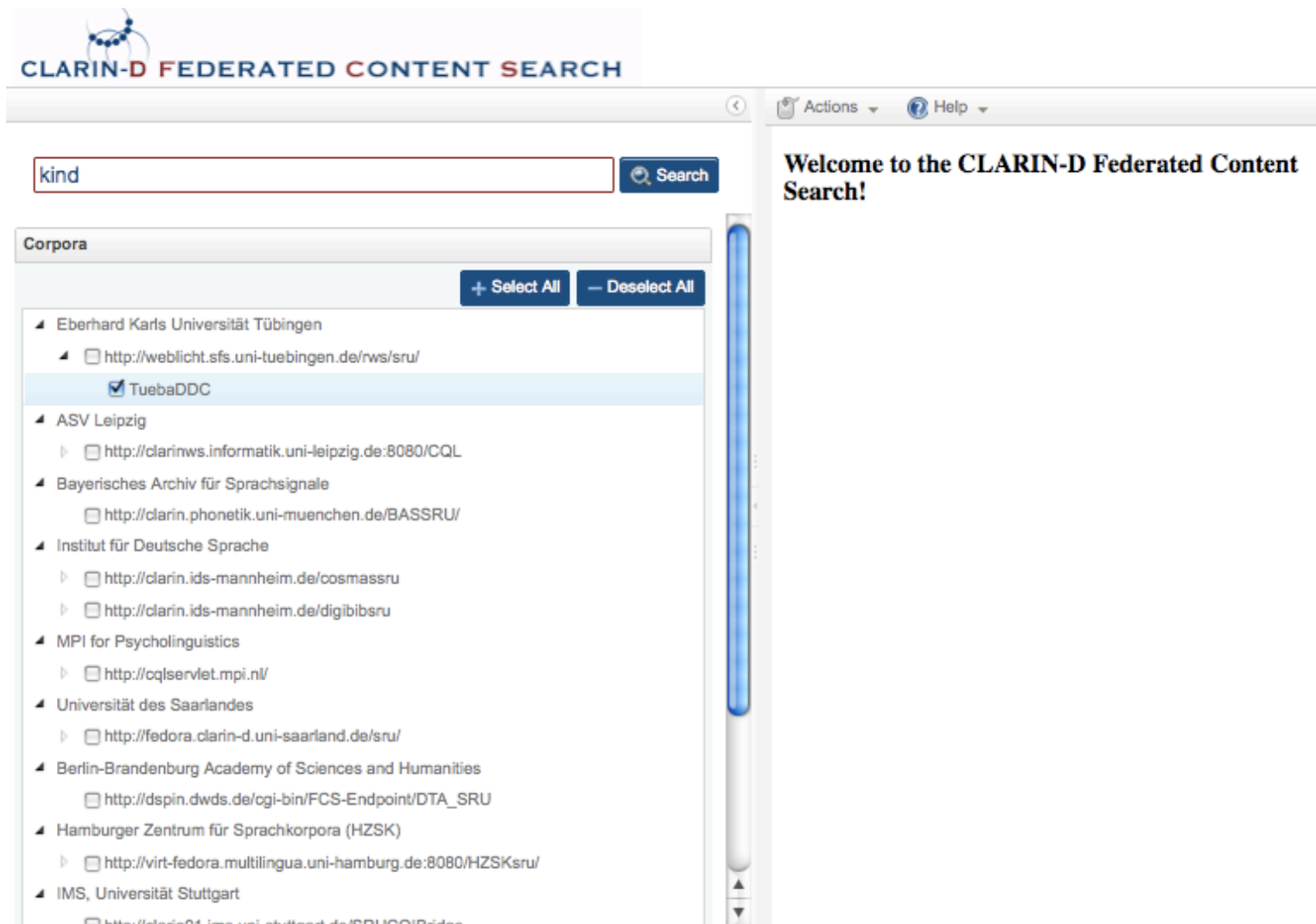
Resources:

-  <http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D@ds1>
-  <http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D@ds2>
-  <http://hdl.handle.net/11858/00-1778-0000-0001-DDAF-D@ds3>

Search in this resource:

CQL query:

Integration with other applications: x-aggregation-context parameter



The screenshot shows the CLARIN-D Federated Content Search interface. At the top left is the CLARIN-D logo and the text "CLARIN-D FEDERATED CONTENT SEARCH". Below this is a search bar containing the text "kind" and a "Search" button. To the right of the search bar is a "Welcome to the CLARIN-D Federated Content Search!" message. Below the search bar is a list of corpora under the heading "Corpora". The list includes several entries, each with a checkbox and a URL. The entry "TuebaDDC" is selected, indicated by a checked checkbox and a blue highlight. The list also includes "Eberhard Karls Universität Tübingen", "ASV Leipzig", "Bayerisches Archiv für Sprachsignale", "Institut für Deutsche Sprache", "MPI for Psycholinguistics", "Universität des Saarlandes", "Berlin-Brandenburg Academy of Sciences and Humanities", "Hamburger Zentrum für Sprachkorpora (HZSK)", and "IMS, Universität Stuttgart".

CLARIN-D FEDERATED CONTENT SEARCH

Search: kind

Welcome to the CLARIN-D Federated Content Search!

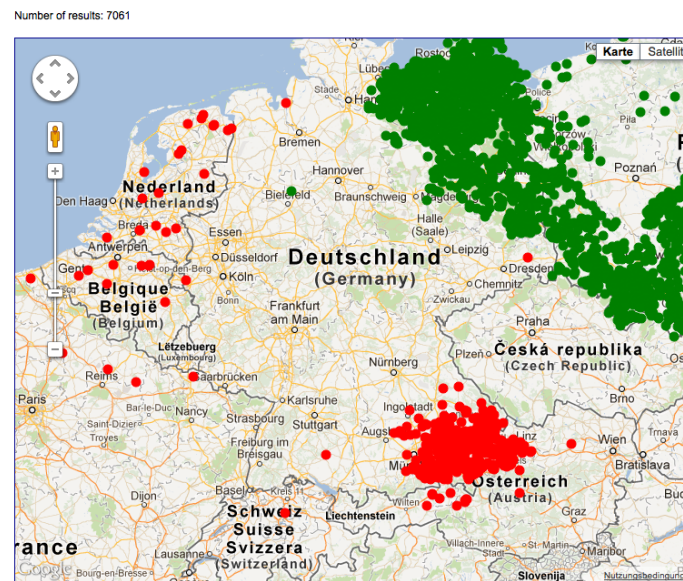
Corpora

- TuebaDDC
- http://weblicht.sfs.uni-tuebingen.de/rws/sru/
- http://clarinws.informatik.uni-leipzig.de:8080/CQL
- http://clarin.phonetik.uni-muenchen.de/BASSRU/
- http://clarin.ids-mannheim.de/cosmassru
- http://clarin.ids-mannheim.de/digibibsr
- http://cqlservlet.mpi.nl/
- http://fedora.clarin-d.uni-saarland.de/sru/
- http://dspin.dwds.de/cgi-bin/FCS-Endpoint/DTA_SRU
- http://virt-fedora.multilingua.uni-hamburg.de:8080/HZSKsru/
- http://clarin01.ims.uni-stuttgart.de/SRU/CQIBridge

To be done ...



- Improvements in GUI
- Integration of additional information about the resources: language, description, landing page (scan operation with `x-cmd-resource-info=true`)
- SRU/CQL interface to external apps
- Support for more (non-KWIC) data views (e.g. KML)





Demo



Thank you!