FCS@CLARIN-AT

status 2013-04

2013-04-24, FCS-Workshop, Copenhagen
Matej Ďurčo, Charly Mörth, ICLTT, Vienna

Table of Contents

- CLARIN-AT situation

- corpus_shell

  The technological framework

- SADE/cr-xq

  The cooperation

- Plans and Wishes for further development

CLARIN-AT - Activities and situation

ICLTT

- CCV – CLARIN Center Vienna [CenterProfile CMD record]

  [http://clarin.aac.ac.at/ccv/index.html]

  expected ready by: 2013-06

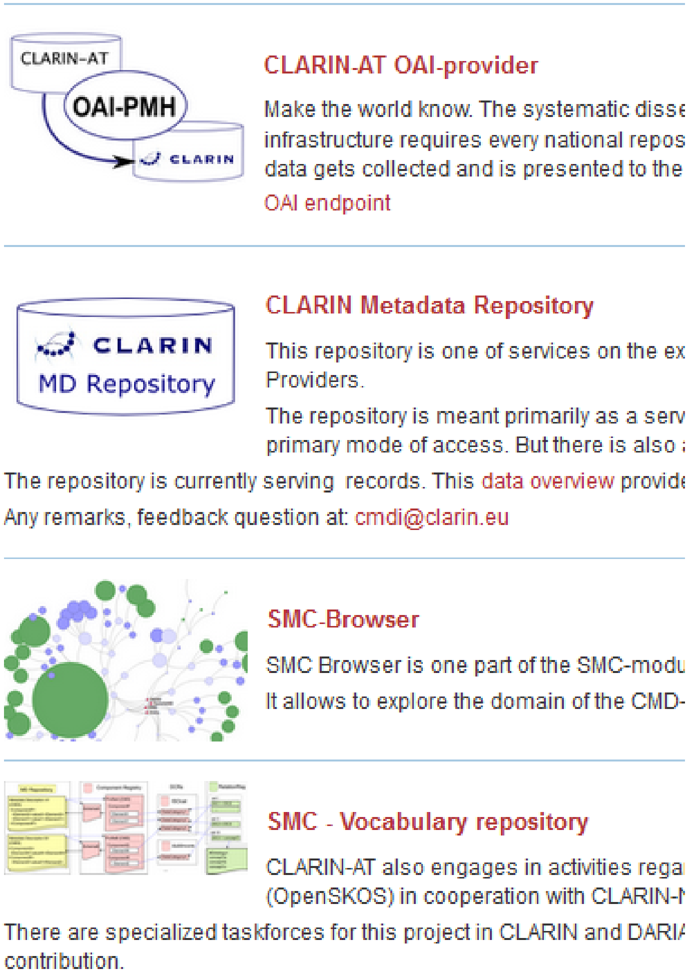
  Infrastructure services:

  - CLARIN Metadata Repository
  - SMC
  - OAI endpoint

- PID + OAI

  [http://clarin.aac.ac.at/oai/provider]

- Controlled Vocabularies
  Task Force (CLARIN + DARIAH)



CLARIN-AT OAI-provider

Make the world know. The systematic disse
infrastructure requires every national repos
data gets collected and is presented to the
OAI endpoint

CLARIN Metadata Repository

This repository is one of services on the ex
Providers.

The repository is meant primarily as a serv
primary mode of access. But there is also a

The repository is currently serving records. This data overview provide
Any remarks, feedback question at: cmdi@clarin.eu

SMC-Browser

SMC Browser is one part of the SMC-modu
It allows to explore the domain of the CMD-

SMC - Vocabulary repository

CLARIN-AT also engages in activities rega
(OpenSKOS) in cooperation with CLARIN-N

There are specialized taskforces for this project in CLARIN and DARIA
contribution.

FCS at ICLTT - corpus_shell

- corpus_shell
  a modular framework for publishing heterogeneous language resource in a distributed environment
- on github: https://github.com/vronk/corpus_shell
- wrapper
  - **php** - implementations accessing MySQL-db for Dictionaries
  - **xquery** implementation for eXist-based content repository (now integrated with SADE!)
  - **perl** implementation mapping to ddc-api (corpus search engine) allowing to access our corpora
- proto-Aggregator: switch.php

  can ask different endpoints, but only one at a time.

  distributed ! (IDS Goethe, TextGrid Library)
- ui in development:

  http://corpus3.aac.ac.at/cs2/corpus_shell/index.html

  multiple query-panels, alternative (full) views if available: full, image, external
- cs-xsl

  set of stylesheets converting FCS/SRU responses (explain, scan, searchRetrieve) to HTML

cr-xq = corpus_shell + SADE

SADE = Scalable Architecture for Digital Editions

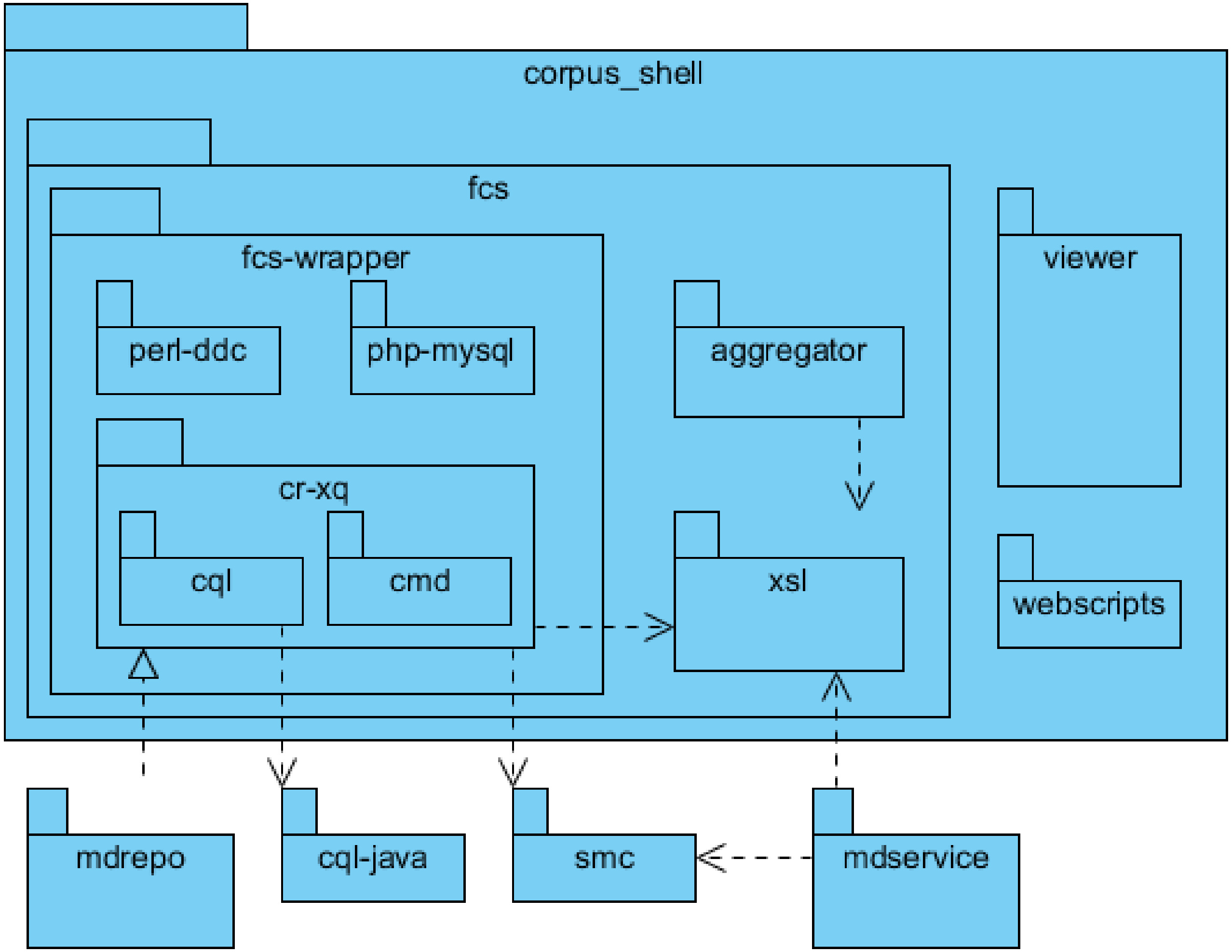
- https://github.com/tharman/SADE
- integrated in TextGrid
- in XQuery
- for existDB
- any XML but TEI as base format

cr-xq-branch

- cooperation between Berlin, Göttingen, Wien
- CMD as default metadata format
- modules:
  - `fcs+cql`
    http …clarin.aac.ac.at/cr/dict-gate/fcs
  - `cmd+resource` for pid and cmd management
  - …

corpus_shell - architecture

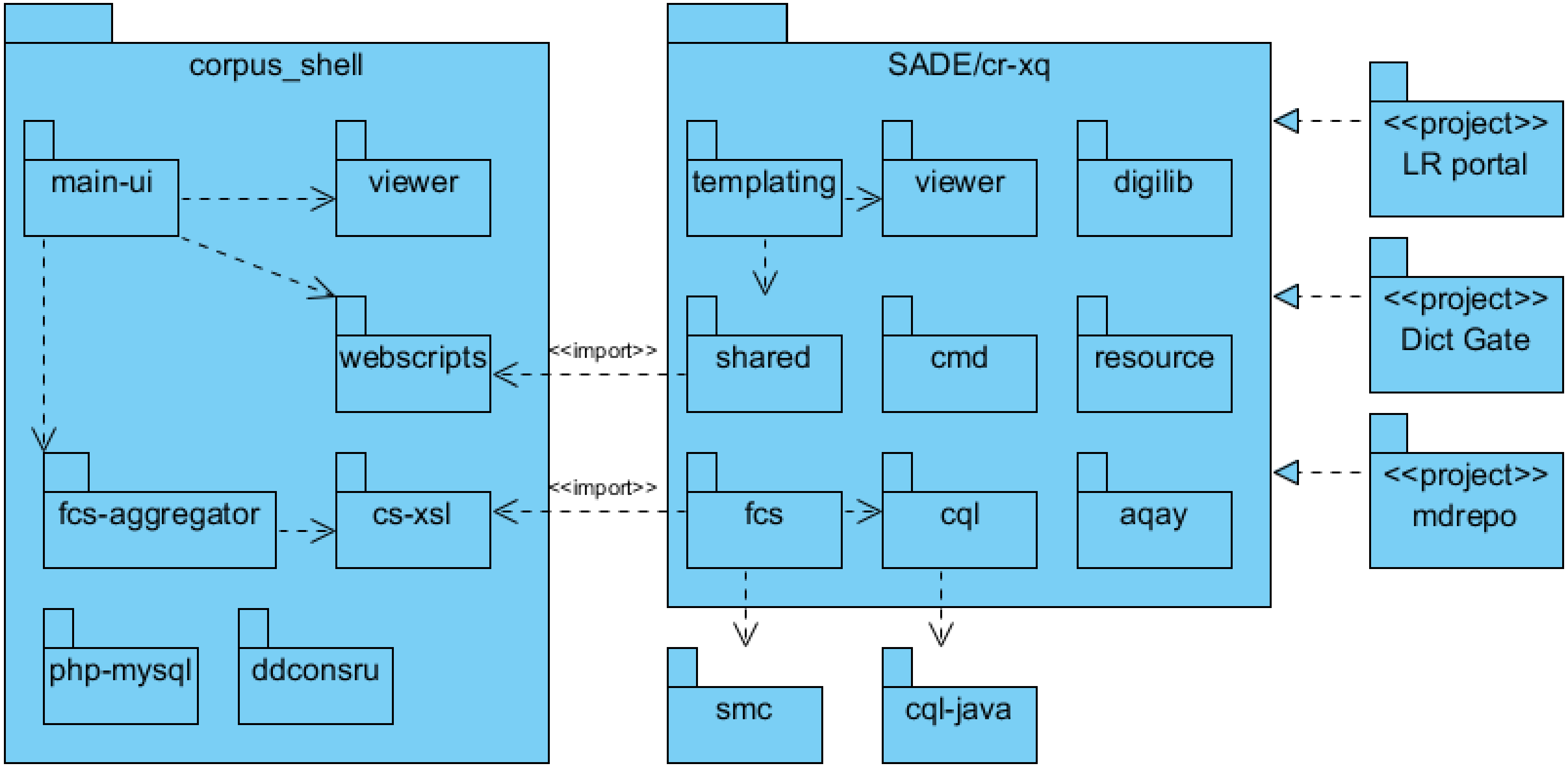
status 2012-06



corpus_shell – relation to SADE

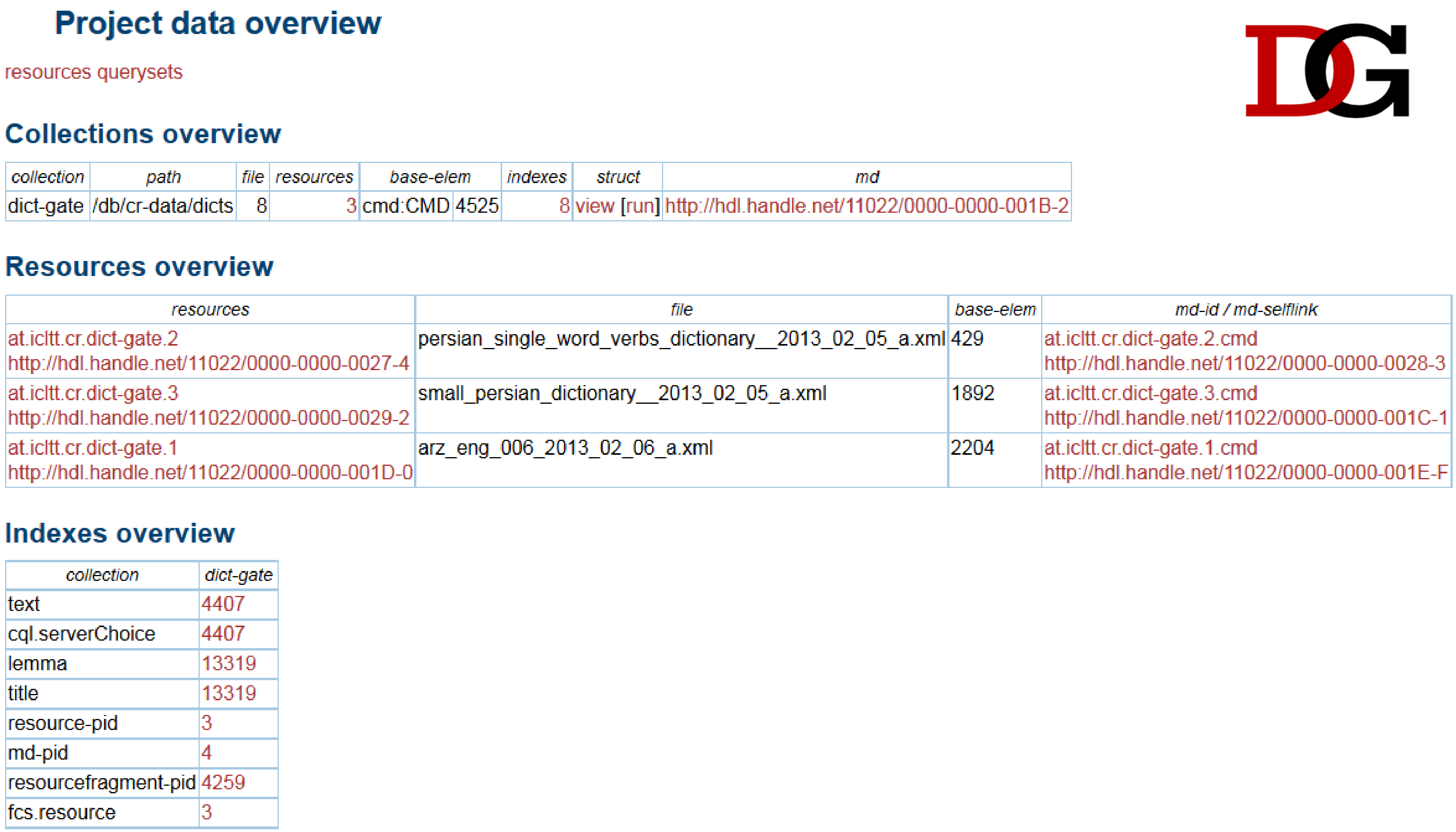
ICLTT

status 2013-04



cr-xq – project overview – simple project

- overview of the resources and indexes with links to fcs-scan
  http://clarin.aac.ac.at/exist/apps/cr-xq/dict-gate/resource



Project data overview

resources querysets

Collections overview

collection	path	file	resources	base-elem	indexes	struct	md
dict-gate	/db/cr-data/dicts	8	3	cmd:CMD 4525	8	view [run]	http://hdl.handle.net/11022/0000-0000-001B-2

Resources overview

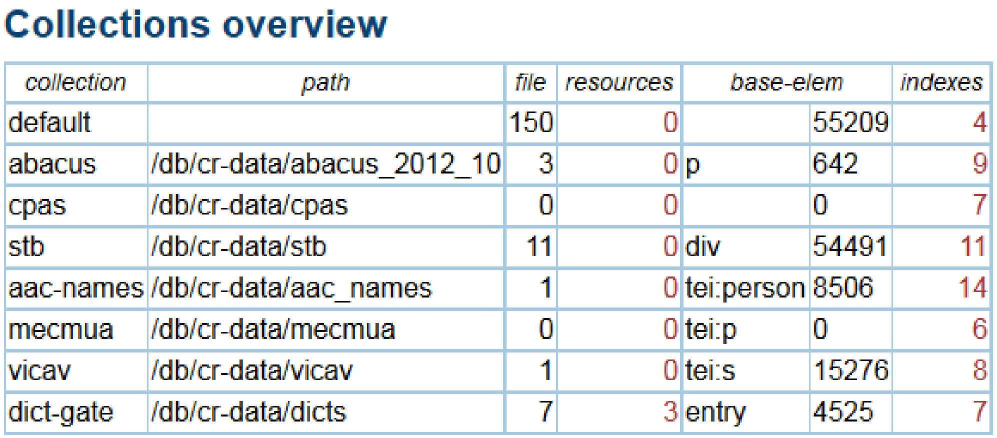
resources	file	base-elem	md-id / md-selflink
at.icltt.cr.dict-gate.2 http://hdl.handle.net/11022/0000-0000-0027-4	persian_single_word_verbs_dictionary__2013_02_05_a.xml	429	at.icltt.cr.dict-gate.2.cmd http://hdl.handle.net/11022/0000-0000-0028-3
at.icltt.cr.dict-gate.3 http://hdl.handle.net/11022/0000-0000-0029-2	small_persian_dictionary__2013_02_05_a.xml	1892	at.icltt.cr.dict-gate.3.cmd http://hdl.handle.net/11022/0000-0000-001C-1
at.icltt.cr.dict-gate.1 http://hdl.handle.net/11022/0000-0000-001D-0	arz_eng_006_2013_02_06_a.xml	2204	at.icltt.cr.dict-gate.1.cmd http://hdl.handle.net/11022/0000-0000-001E-F

Indexes overview

collection	dict-gate
text	4407
cql.serverChoice	4407
lemma	13319
title	13319
resource-pid	3
md-pid	4
resourcefragment-pid	4259
fcs.resource	3

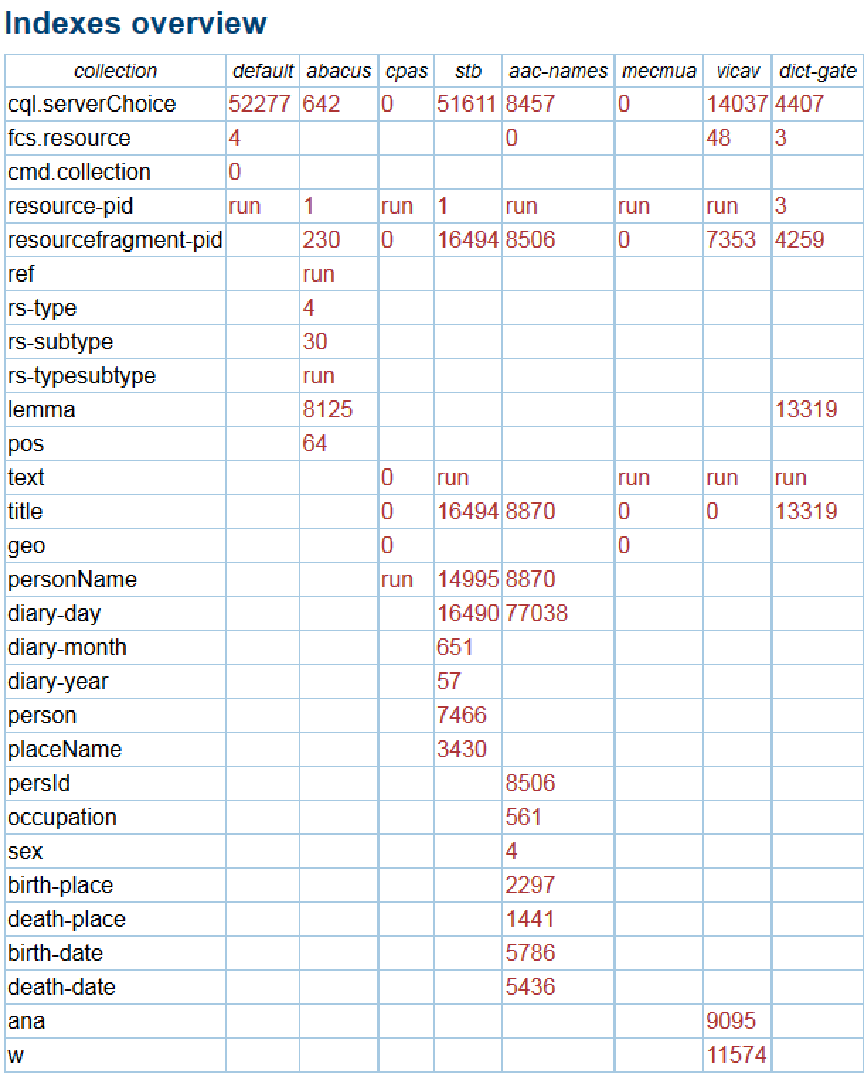
cr-xq – project overview – compound project

overview of Collections + Indexes
different collections support different indexes



Collections overview

collection	*path*	*file*	*resources*	*base-elem*		*indexes*
default		150	0		55209	4
abacus	/db/cr-data/abacus_2012_10	3	0	p	642	9
cpas	/db/cr-data/cpas	0	0		0	7
stb	/db/cr-data/stb	11	0	div	54491	11
aac-names	/db/cr-data/aac_names	1	0	tei:person	8506	14
mecmua	/db/cr-data/mecmua	0	0	tei:p	0	6
vicav	/db/cr-data/vicav	1	0	tei:s	15276	8
dict-gate	/db/cr-data/dicts	7	3	entry	4525	7

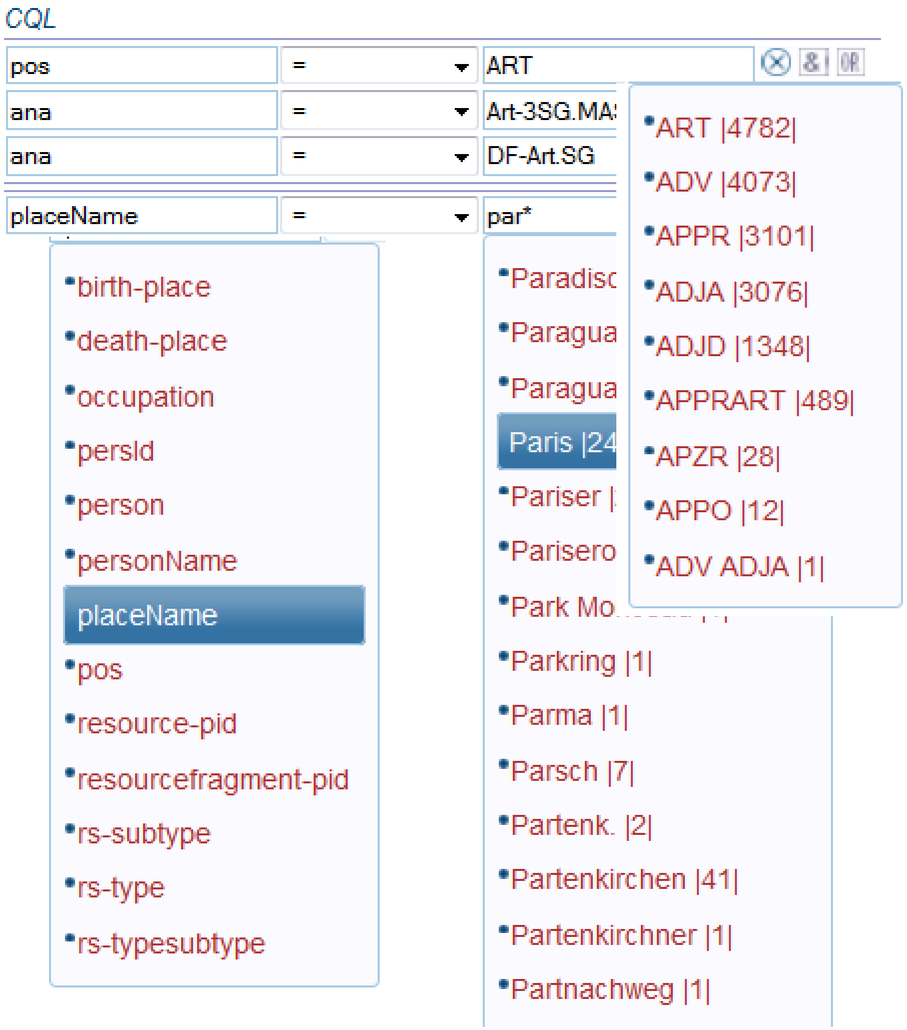


Indexes overview

collection	*default*	*abacus*	*cpas*	*stb*	*aac-names*	*mecmua*	*vicav*	*dict-gate*
cql.serverChoice	52277	642	0	51611	8457	0	14037	4407
fcs.resource	4				0		48	3
cmd.collection	0							
resource-pid	run	1	run	1	run	run	run	3
resourcefragment-pid		230	0	16494	8506	0	7353	4259
ref		run						
rs-type		4						
rs-subtype		30						
rs-typesubtype		run						
lemma		8125						13319
pos		64						
text			0	run		run	run	run
title			0	16494	8870	0	0	13319
geo			0			0		
personName			run	14995	8870			
diary-day				16490	77038			
diary-month				651				
diary-year				57				
person				7466				
placeName				3430				
persId					8506			
occupation					561			
sex					4			
birth-place					2297			
death-place					1441			
birth-date					5786			
death-date					5436			
ana							9095	
w							11574	

query_input – js lib

- js-library within corpus_shell
- implemented as a complex jQuery-widget
- simple json configuration
- customizable widgets
- cql widget (AND (OR searchClauses))
  - configuration by sru:explain
  - contextual term suggestion by sru:scan
- in progress: cql-parsing widget
  input-field validating input on the fly via js-based cql-parser feeding contextual autocomplete



CQL

pos = ART

ana = Art-3SG.MA:

ana = DF-Art.SG

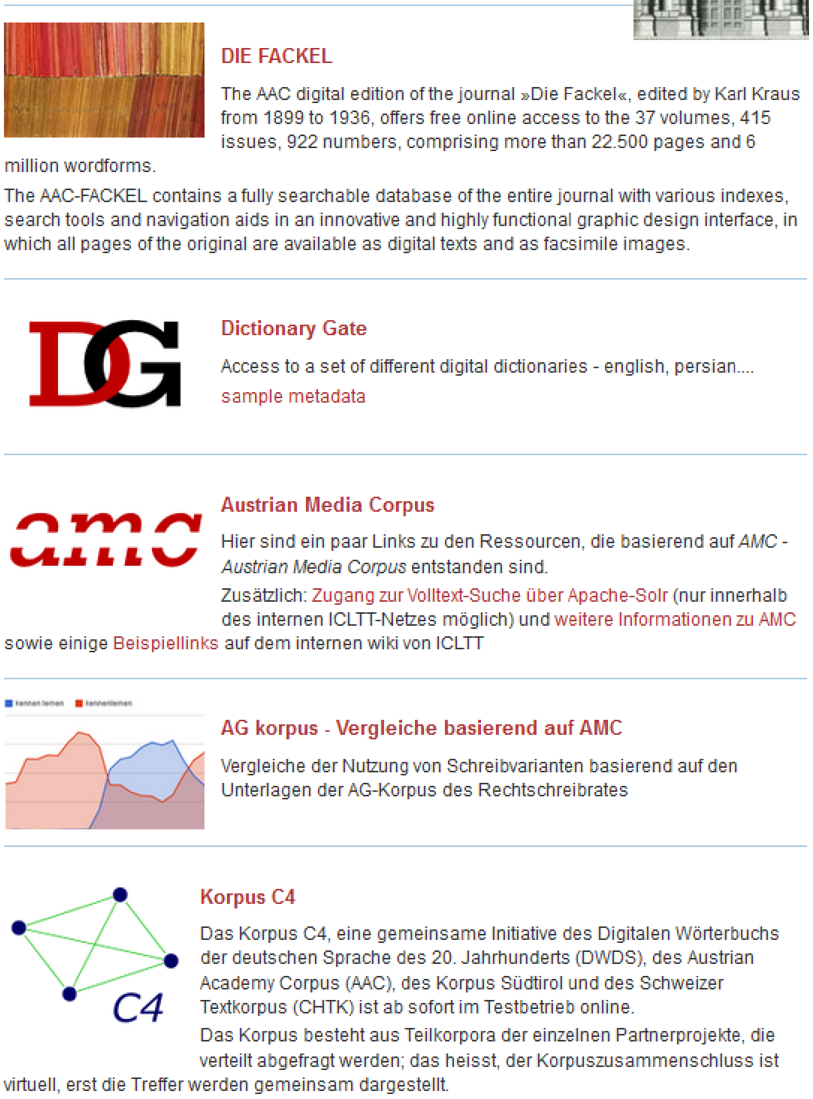
placeName = par*

ICLTT LR Portal

a heterogeneous selection of ICLTT Resources
is available via corpus_shell:

- Dictionaries
  - project VICAV - arabic dialects
  - Wiktionary in TEI
- Corpora
  - parallel corpus: Freud, Traumdeutung
  - C4 – distributed corpus of german
- Schnitzler Tagebuch online
  - full-text + semantic indexes (names, places)
- Barock texts

  with full-text and facsimile
- external:
  - TextGrid Digilib
  - IDS Goethe



DIE FACKEL

The AAC digital edition of the journal »Die Fackel«, edited by Karl Kraus from 1899 to 1936, offers free online access to the 37 volumes, 415 issues, 922 numbers, comprising more than 22.500 pages and 6 million wordforms.

The AAC-FACKEL contains a fully searchable database of the entire journal with various indexes, search tools and navigation aids in an innovative and highly functional graphic design interface, in which all pages of the original are available as digital texts and as facsimile images.

Dictionary Gate

Access to a set of different digital dictionaries - english, persian....

sample metadata

Austrian Media Corpus

Hier sind ein paar Links zu den Ressourcen, die basierend auf *AMC - Austrian Media Corpus* entstanden sind.

Zusätzlich: Zugang zur Volltext-Suche über Apache-Solr (nur innerhalb des internen ICLTT-Netzes möglich) und weitere Informationen zu AMC sowie einige Beispiellinks auf dem internen wiki von ICLTT

AG korpus - Vergleiche basierend auf AMC

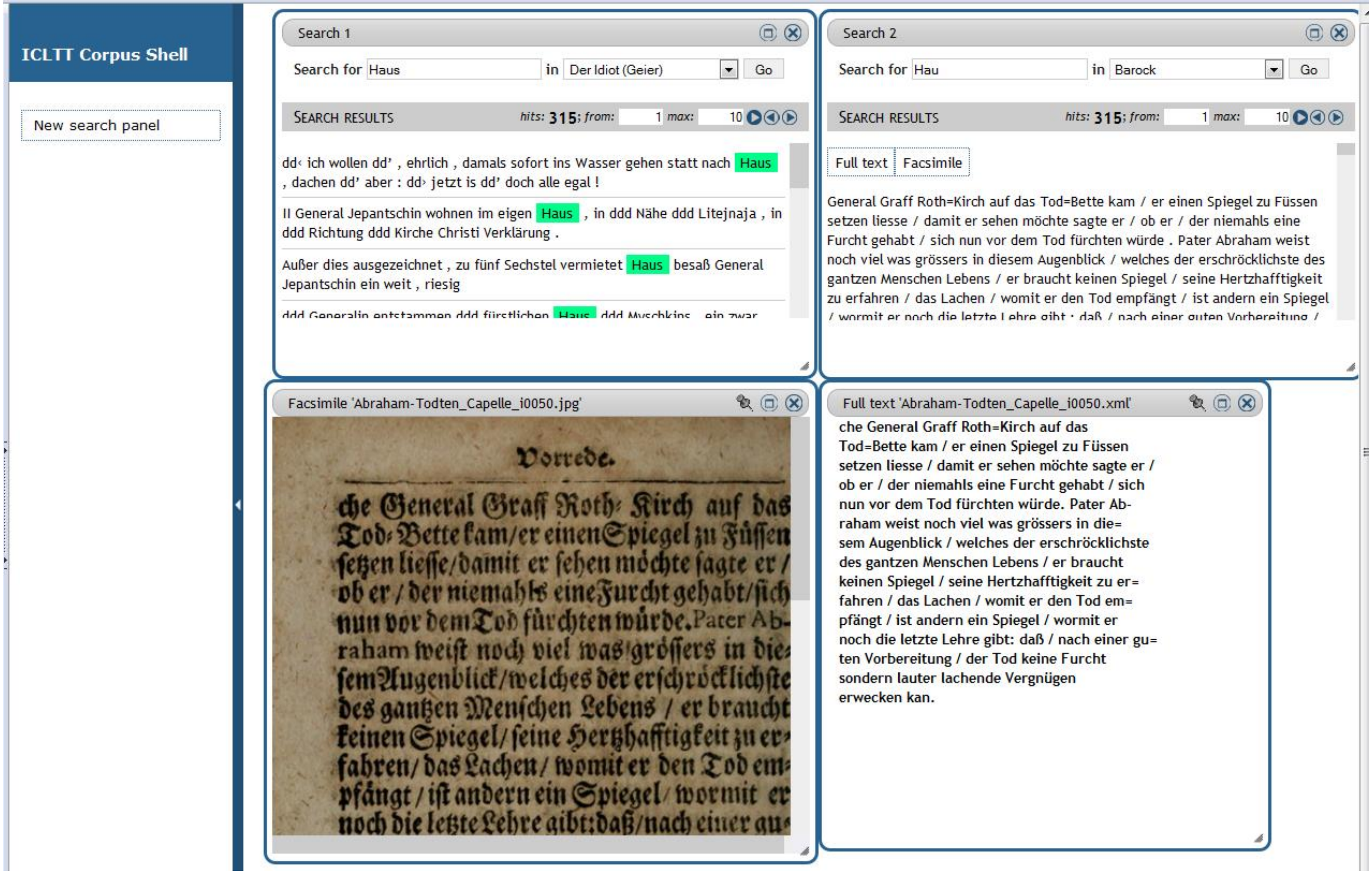
Vergleiche der Nutzung von Schreibvarianten basierend auf den Unterlagen der AG-Korpus des Rechtschreibrates

Korpus C4

Das Korpus C4, eine gemeinsame Initiative des Digitalen Wörterbuchs der deutschen Sprache des 20. Jahrhunderts (DWDS), des Austrian Academy Corpus (AAC), des Korpus Südtirol und des Schweizer Textkorpus (CHTK) ist ab sofort im Testbetrieb online.

Das Korpus besteht aus Teilkorpora der einzelnen Partnerprojekte, die verteilt abgefragt werden; das heisst, der Korpuszusammenschluss ist virtuell, erst die Treffer werden gemeinsam dargestellt.

UI sample 1 – barock texte



ICLTT Corpus Shell
New search panel
Search 1
Search for Haus in Der Idiot (Geier) Go
Search results hits: 315; from: 1 max: 10
Search 2
Search for Hau in Barock Go
Search results hits: 315; from: 1 max: 10
Full text
Facsimile
Facsimile 'Abraham-Todten_Capelle_i0050.jpg'
Full text 'Abraham-Todten_Capelle_i0050.xml'

dd‹ ich wollen dd' , ehrlich , damals sofort ins Wasser gehen statt nach Haus , dachen dd' aber : dd› jetzt is dd' doch alle egal !

II General Jepantschin wohnen im eigen Haus , in ddd Nähe ddd Litejnaja , in ddd Richtung ddd Kirche Christi Verklärung .

Außer dies ausgezeichnet , zu fünf Sechstel vermietet Haus besaß General Jepantschin ein weit , riesig

ddd Generalin entstammen ddd fürstlichen Haus ddd Myschkins , ein zwar

General Graff Roth=Kirch auf das Tod=Bette kam / er einen Spiegel zu Füssen setzen liesse / damit er sehen möchte sagte er / ob er / der niemahls eine Furcht gehabt / sich nun vor dem Tod fürchten würde . Pater Abraham weist noch viel was grössers in diesem Augenblick / welches der erschröcklichste des gantzen Menschen Lebens / er braucht keinen Spiegel / seine Hertzhafftigkeit zu erfahren / das Lachen / womit er den Tod empfängt / ist andern ein Spiegel / wormit er noch die letzte Lehre gibt : daß / nach einer guten Vorbereitung /

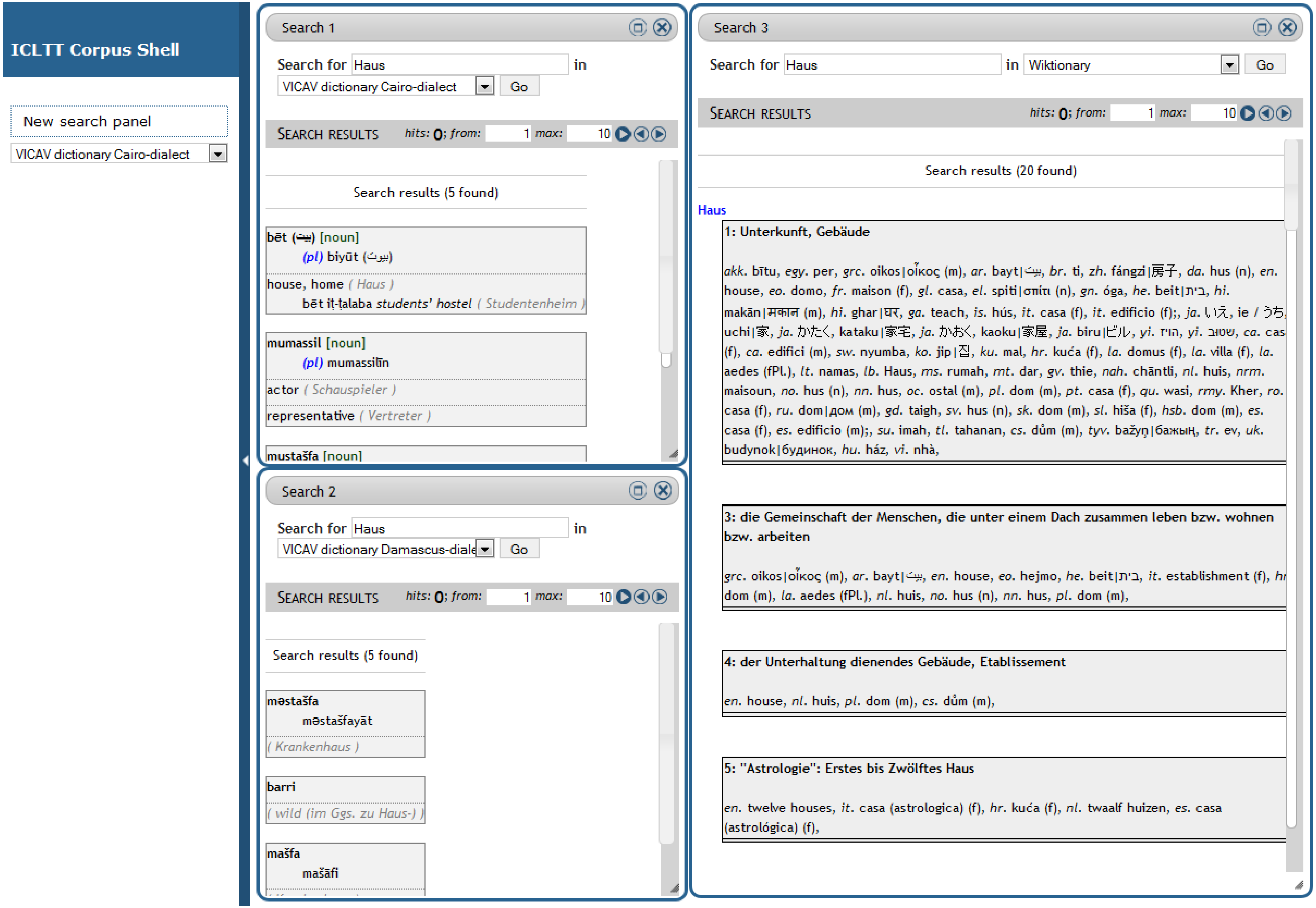
Vorrede.

che General Graff Roth⸗Kirch auf das
Tod⸗Bette kam/er einen Spiegel zu Füſſen
ſetzen lieſſe/damit er ſehen möchte ſagte er /
ob er / der niemahls eine Furcht gehabt/ſich
nun vor dem Tod fürchten würde.Pater Ab-
raham weiſt noch viel was gröſſers in die⸗
ſem Augenblick/welches der erſchröcklichſte
des gantzen Menſchen Lebens / er braucht
keinen Spiegel/ſeine Hertzhafftigkeit zu er⸗
fahren/das Lachen/womit er den Tod em⸗
pfängt / iſt andern ein Spiegel/ wormit er
noch die letzte Lehre gibt:daß/nach einer gu⸗

che General Graff Roth=Kirch auf das
Tod=Bette kam / er einen Spiegel zu Füssen
setzen liesse / damit er sehen möchte sagte er /
ob er / der niemahls eine Furcht gehabt / sich
nun vor dem Tod fürchten würde. Pater Ab-
raham weist noch viel was grössers in die=
sem Augenblick / welches der erschröcklichste
des gantzen Menschen Lebens / er braucht
keinen Spiegel / seine Hertzhafftigkeit zu er=
fahren / das Lachen / womit er den Tod em=
pfängt / ist andern ein Spiegel / wormit er
noch die letzte Lehre gibt: daß / nach einer gu=
ten Vorbereitung / der Tod keine Furcht
sondern lauter lachende Vergnügen
erwecken kan.

UI sample 2 - dictionaries

ICLTT



ICLTT Corpus Shell

New search panel

VICAV dictionary Cairo-dialect

Search 1

Search for Haus in VICAV dictionary Cairo-dialect Go

SEARCH RESULTS hits: 0; from: 1 max: 10

Search results (5 found)

bēt (بيت) [noun]
(pl) biyūt (بيوت)

house, home *(Haus)*
bēt iṭ-ṭalaba *students' hostel* *(Studentenheim)*

mumassil [noun]
(pl) mumassilīn

actor *(Schauspieler)*

representative *(Vertreter)*

mustašfa [noun]

Search 2

Search for Haus in VICAV dictionary Damascus-diale Go

SEARCH RESULTS hits: 0; from: 1 max: 10

Search results (5 found)

məstašfa
məstašfayāt

(Krankenhaus)

barri

(wild (im Ggs. zu Haus-))

mašfa
mašāfi

Search 3

Search for Haus in Wiktionary Go

SEARCH RESULTS hits: 0; from: 1 max: 10

Search results (20 found)

Haus

1: Unterkunft, Gebäude

akk. bītu, *egy.* per, *grc.* oikos|οἶκος (m), *ar.* bayt|بيت, *br.* ti, *zh.* fángzi|房子, *da.* hus (n), *en.* house, *eo.* domo, *fr.* maison (f), *gl.* casa, *el.* spiti|σπίτι (n), *gn.* óga, *he.* beit|בית, *hi.* makān|मकान (m), *hi.* ghar|घर, *ga.* teach, *is.* hús, *it.* casa (f), *it.* edificio (f);, *ja.* いえ, ie / うち, uchi|家, *ja.* かたく, kataku|家宅, *ja.* かおく, kaoku|家屋, *ja.* biru|ビル, *yi.* הויז, *yi.* שטוב, *ca.* cas (f), *ca.* edifici (m), *sw.* nyumba, *ko.* jip|집, *ku.* mal, *hr.* kuća (f), *la.* domus (f), *la.* villa (f), *la.* aedes (fPl.), *lt.* namas, *lb.* Haus, *ms.* rumah, *mt.* dar, *gv.* thie, *nah.* chāntli, *nl.* huis, *nrm.* maisoun, *no.* hus (n), *nn.* hus, *oc.* ostal (m), *pl.* dom (m), *pt.* casa (f), *qu.* wasi, *rmy.* Kher, *ro.* casa (f), *ru.* dom|дом (m), *gd.* taigh, *sv.* hus (n), *sk.* dom (m), *sl.* hiša (f), *hsb.* dom (m), *es.* casa (f), *es.* edificio (m);, *su.* imah, *tl.* tahanan, *cs.* dům (m), *tyv.* bažyŋ|бажың, *tr.* ev, *uk.* budynok|будинок, *hu.* ház, *vi.* nhà,

3: die Gemeinschaft der Menschen, die unter einem Dach zusammen leben bzw. wohnen bzw. arbeiten

grc. oikos|οἶκος (m), *ar.* bayt|بيت, *en.* house, *eo.* hejmo, *he.* beit|בית, *it.* establishment (f), *hr* dom (m), *la.* aedes (fPl.), *nl.* huis, *no.* hus (n), *nn.* hus, *pl.* dom (m),

4: der Unterhaltung dienendes Gebäude, Etablissement

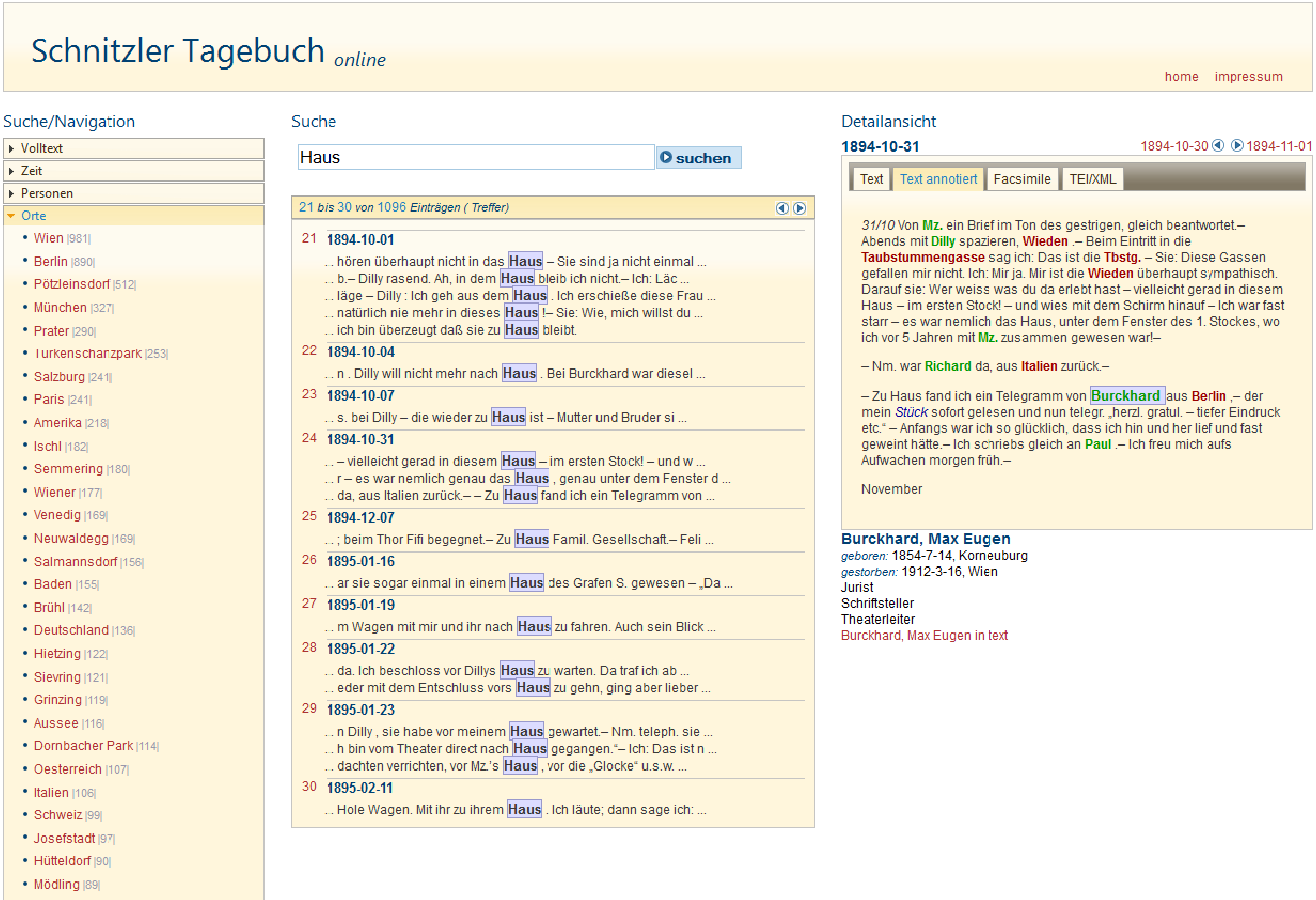
en. house, *nl.* huis, *pl.* dom (m), *cs.* dům (m),

5: "Astrologie": Erstes bis Zwölftes Haus

en. twelve houses, *it.* casa (astrologica) (f), *hr.* kuća (f), *nl.* twaalf huizen, *es.* casa (astrológica) (f),

UI sample 3 – Schnitzler Tagebuch

ICLTT



Schnitzler Tagebuch *online*

home impressum

Suche/Navigation

- Volltext
- Zeit
- Personen
- Orte
  - Wien |981|
  - Berlin |890|
  - Pötzleinsdorf |512|
  - München |327|
  - Prater |290|
  - Türkenschanzpark |253|
  - Salzburg |241|
  - Paris |241|
  - Amerika |218|
  - Ischl |182|
  - Semmering |180|
  - Wiener |177|
  - Venedig |169|
  - Neuwaldegg |169|
  - Salmannsdorf |156|
  - Baden |155|
  - Brühl |142|
  - Deutschland |136|
  - Hietzing |122|
  - Sievring |121|
  - Grinzing |119|
  - Aussee |116|
  - Dornbacher Park |114|
  - Oesterreich |107|
  - Italien |106|
  - Schweiz |99|
  - Josefstadt |97|
  - Hütteldorf |90|
  - Mödling |89|

Suche

Haus

suchen

21 bis 30 von 1096 Einträgen (Treffer)

21 **1894-10-01**

... hören überhaupt nicht in das **Haus** – Sie sind ja nicht einmal ...
... b.– Dilly rasend. Ah, in dem **Haus** bleib ich nicht.– Ich: Läc ...
... läge – Dilly : Ich geh aus dem **Haus** . Ich erschieße diese Frau ...
... natürlich nie mehr in dieses **Haus** !– Sie: Wie, mich willst du ...
... ich bin überzeugt daß sie zu **Haus** bleibt.

22 **1894-10-04**

... n . Dilly will nicht mehr nach **Haus** . Bei Burckhard war diesel ...

23 **1894-10-07**

... s. bei Dilly – die wieder zu **Haus** ist – Mutter und Bruder si ...

24 **1894-10-31**

... – vielleicht gerad in diesem **Haus** – im ersten Stock! – und w ...
... r – es war nemlich genau das **Haus** , genau unter dem Fenster d ...
... da, aus Italien zurück.– – Zu **Haus** fand ich ein Telegramm von ...

25 **1894-12-07**

... ; beim Thor Fifi begegnet.– Zu **Haus** Famil. Gesellschaft.– Feli ...

26 **1895-01-16**

... ar sie sogar einmal in einem **Haus** des Grafen S. gewesen – „Da ...

27 **1895-01-19**

... m Wagen mit mir und ihr nach **Haus** zu fahren. Auch sein Blick ...

28 **1895-01-22**

... da. Ich beschloss vor Dillys **Haus** zu warten. Da traf ich ab ...
... eder mit dem Entschluss vors **Haus** zu gehn, ging aber lieber ...

29 **1895-01-23**

... n Dilly , sie habe vor meinem **Haus** gewartet.– Nm. teleph. sie ...
... h bin vom Theater direct nach **Haus** gegangen.“– Ich: Das ist n ...
... dachten verrichten, vor Mz.’s **Haus** , vor die „Glocke“ u.s.w. ...

30 **1895-02-11**

... Hole Wagen. Mit ihr zu ihrem **Haus** . Ich läute; dann sage ich: ...

Detailansicht

1894-10-31

1894-10-30 ◀ ▶ 1894-11-01

Text | Text annotiert | Facsimile | TEI/XML

31/10 Von **Mz.** ein Brief im Ton des gestrigen, gleich beantwortet.– Abends mit **Dilly** spazieren, **Wieden** .– Beim Eintritt in die **Taubstummengasse** sag ich: Das ist die **Tbstg.** – Sie: Diese Gassen gefallen mir nicht. Ich: Mir ja. Mir ist die **Wieden** überhaupt sympathisch. Darauf sie: Wer weiss was du da erlebt hast – vielleicht gerad in diesem Haus – im ersten Stock! – und wies mit dem Schirm hinauf – Ich war fast starr – es war nemlich das Haus, unter dem Fenster des 1. Stockes, wo ich vor 5 Jahren mit **Mz.** zusammen gewesen war!–

– Nm. war **Richard** da, aus **Italien** zurück.–

– Zu Haus fand ich ein Telegramm von **Burckhard** aus **Berlin** ,– der mein *Stück* sofort gelesen und nun telegr. „herzl. gratul. – tiefer Eindruck etc.“ – Anfangs war ich so glücklich, dass ich hin und her lief und fast geweint hätte.– Ich schriebs gleich an **Paul** .– Ich freu mich aufs Aufwachen morgen früh.–

November

Burckhard, Max Eugen

geboren: 1854-7-14, Korneuburg
gestorben: 1912-3-16, Wien
Jurist
Schriftsteller
Theaterleiter
Burckhard, Max Eugen in text

Wishes – next steps

- **more complex CQL-queries**
  CQL-indexes, boolean, sequential tier search
  - **isocat** as new context set
  - fcs.resourcefragment-id, cmd.pid as new indexes? (next to fcs.resource?)

```
isocat.DC-1324  isocat.lemma
isocat.DC-1403  isocat.token
…
```

- **ResourceType/DataViews**
  x-format, x-dataview
  what and how to deliver
  - Tiers / Annotation Layers
  - Geo-data (KML)
  - Dataset (list, table, matrix)
  - Graph
- unified json-serialization ? (x-format=json)
- **CDMDC** - Combined Distributed Metadata Content Search

– Result Format – DataViews I

basic/minimal/default(?) currently „defined“ DataViews:

- kwic svn-repo:/FederatedSearch/Resource-KWIC.xsd

  moved to separate schema/ns

- title
- metadata (CMD)

```
<kwic:kwic>
 <kwic:c type="left">Also Paris, der damals Gast im
</kwic:c>
 <kwic:kw>Haus</kwic:kw>
 <kwic:c type="right"> der Atriden Frech den gastlichen
Tisch entweiht,der die Gattin entführt hat.</kwic:c>
</kwic:kwic>
```

new param needed:

- **x-dataview**

  to allow select DataViews in the result

```
<DataView type=“metadata” schema=“{cmd}”
     ref=“{md-handle}” />
/* OR */
<DataView type=“metadata” schema=“{cmd}”>
   <CMD>…</CMD>
```

+ another param would be handy:

- **x-format**

  to allow say how the result shall be delivered
  (SRU only has `resultStylesheet`-parameter with URL to a XSLT-stylesheet to be applied)

  ?? again value domain – starting with [XML, HTML, JSON] ?

– Result Format – DataViews II

needed further

- full
- image
- existing formats

  TEI, EAF, TCF, ???

- geographic data [KML]

more difficult ones:

- multi-tiers + alignment between tiers

  one tier per DataView? or (TCF/EAF / ANNEX?)

- parallel corpus

  each language one Resource? or ResourceFragment, alignment

- summary

  nested frequency list, matrix – JSON?

```
[<key, number, link?>]
```

- graph

– Result Format – DataViews III

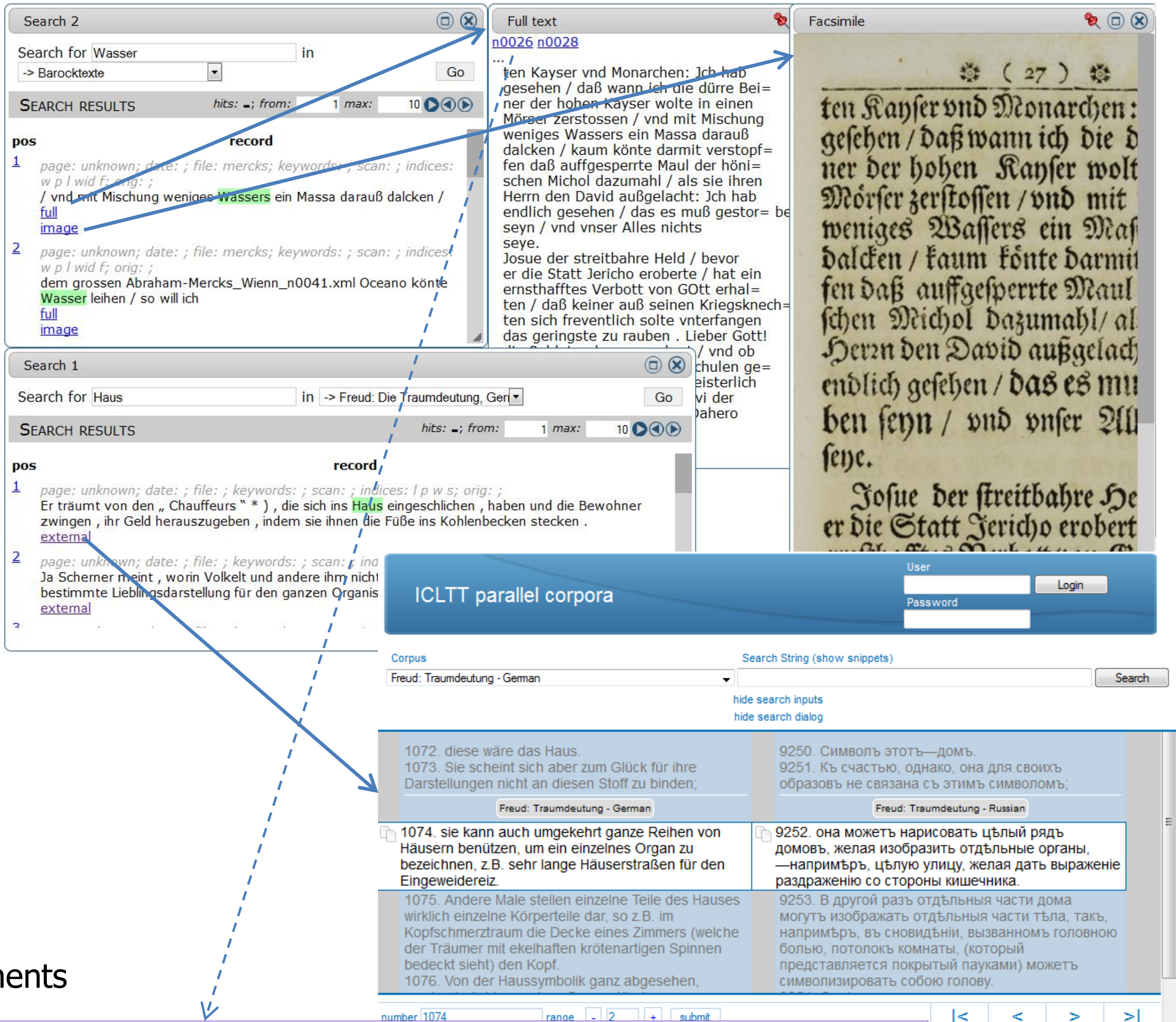
ICLTT

links already in the FCS-response as separate `DataView@ref`

handled generically by FCS-XSL

„specialized“ viewer:

- TEI-stylesheets
- full, image
- ParCorp
- navigation
  given an ordered sequence of ResourceFragments



Search 2

Search for Wasser in -> Barocktexte Go

SEARCH RESULTS hits: -; from: 1 max: 10

pos record

1 page: unknown; date: ; file: mercks; keywords: ; scan: ; indices: w p l wid f; orig: ;
/ vnd mit Mischung weniges Wassers ein Massa darauß dalcken /
full
image

2 page: unknown; date: ; file: mercks; keywords: ; scan: ; indices: w p l wid f; orig: ;
dem grossen Abraham-Mercks_Wienn_n0041.xml Oceano könte Wasser leihen / so will ich
full
image

Full text

n0026 n0028

...
ten Kayser vnd Monarchen: Jch hab
gesehen / daß wann ich die dürre Bei=
ner der hohen Kayser wolte in einen
Mörser zerstossen / vnd mit Mischung
weniges Wassers ein Massa darauß
dalcken / kaum könte darmit verstopf=
fen daß auffgesperrte Maul der höni=
schen Michol dazumahl / als sie ihren
Herrn den David außgelacht: Jch hab
endlich gesehen / das es muß gestor= be
seyn / vnd vnser Alles nichts
seye.
Josue der streitbahre Held / bevor
er die Statt Jericho eroberte / hat ein
ernsthafftes Verbott von GOtt erhal=
ten / daß keiner auß seinen Kriegsknech=
ten sich freventlich solte vnterfangen
das geringste zu rauben . Lieber Gott!
/ vnd ob
chulen ge=
eisterlich
vi der
Dahero

Facsimile

(27)
ten Kayſer vnd Monarchen:
geſehen / daß wann ich die d
ner der hohen Kayſer wolt
Mörſer zerſtoſſen / vnd mit
weniges Waſſers ein Maſ
dalcken / kaum kōnte darmit
fen daß auffgeſperrte Maul
ſchen Michol dazumahl / al
Herꝛn den David außgelach
endlich geſehen / das es mu
ben ſeyn / vnd vnſer All
ſeye.
Joſue der ſtreitbahre He
er die Statt Jericho erobert

Search 1

Search for Haus in -> Freud: Die Traumdeutung, Ger Go

SEARCH RESULTS hits: -; from: 1 max: 10

pos record

1 page: unknown; date: ; file: ; keywords: ; scan: ; indices: l p w s; orig: ;
Er träumt von den „ Chauffeurs “ *) , die sich ins Haus eingeschlichen , haben und die Bewohner zwingen , ihr Geld herauszugeben , indem sie ihnen die Füße ins Kohlenbecken stecken .
external

2 page: unknown; date: ; file: ; keywords: ; scan: ; ind
Ja Scherner meint , worin Volkelt und andere ihm nicht
bestimmte Lieblingsdarstellung für den ganzen Organis
external

ICLTT parallel corpora

User Login
Password

Corpus: Freud: Traumdeutung - German
Search String (show snippets) Search

hide search inputs
hide search dialog

1072. diese wäre das Haus.
1073. Sie scheint sich aber zum Glück für ihre Darstellungen nicht an diesen Stoff zu binden;

Freud: Traumdeutung - German

1074. sie kann auch umgekehrt ganze Reihen von Häusern benützen, um ein einzelnes Organ zu bezeichnen, z.B. sehr lange Häuserstraßen für den Eingeweidereiz.

1075. Andere Male stellen einzelne Teile des Hauses wirklich einzelne Körperteile dar, so z.B. im Kopfschmerztraum die Decke eines Zimmers (welche der Träumer mit ekelhaften krötenartigen Spinnen bedeckt sieht) den Kopf.
1076. Von der Haussymbolik ganz abgesehen,

9250. Символъ этотъ—домъ.
9251. Къ счастью, однако, она для своихъ образовъ не связана съ этимъ символомъ;

Freud: Traumdeutung - Russian

9252. она можетъ нарисовать цѣлый рядъ домовъ, желая изобразить отдѣльные органы, —напримѣръ, цѣлую улицу, желая дать выраженіе раздраженію со стороны кишечника.

9253. В другой разъ отдѣльныя части дома могутъ изображать отдѣльныя части тѣла, такъ, напримѣръ, въ сновидѣніи, вызванномъ головною болью, потолокъ комнаты, (который представляется покрытый пауками) можетъ символизировать собою голову.

number 1074 range - 2 + submit

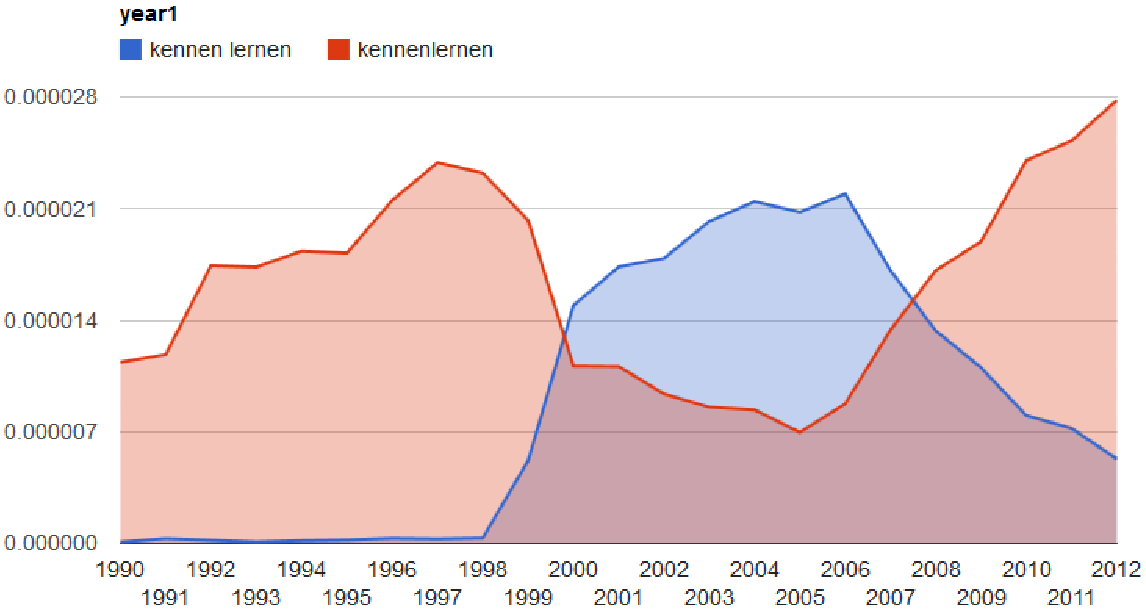
```
<fcs:ResourceFragment type="prev" pid="n0026" ref="?query=toc=n0026"/>
<fcs:ResourceFragment type="next" pid="n0028" ref="?query=toc=n0028"/>
```

– Result Format – DataViews IV

- Lists
- Dataset
- Matrix

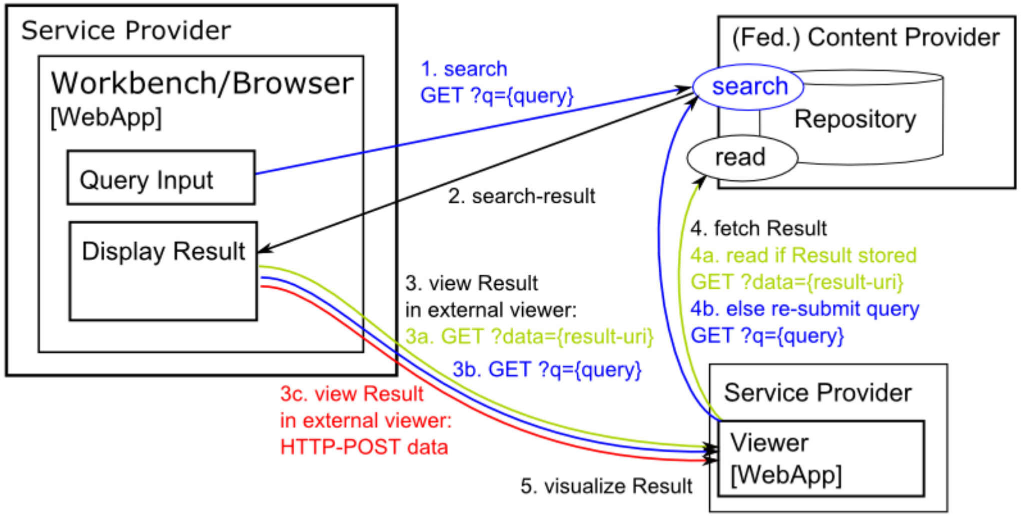
But how to map it to sru:records?

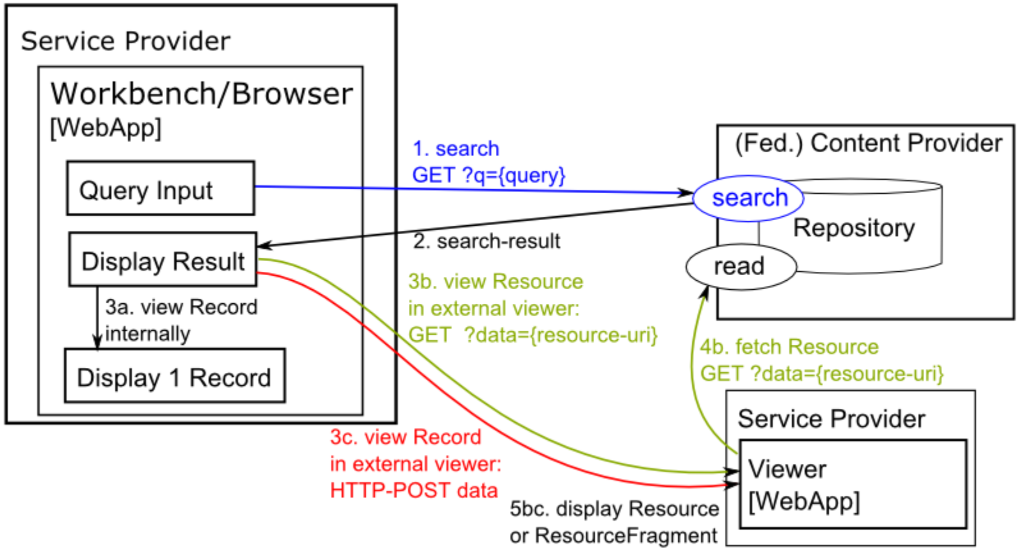
key	tokens-sum	kennen lernen		kennenlernen	
all	5.696.309.303	68.496	12,02	90.457	15,88
1990	35.263.029	3	0,09	401	11,37
1991	54.206.194	16	0,3	642	11,84
1992	99.053.385	19	0,19	1.727	17,44
1993	98.958.468	9	0,09	1.716	17,34
1994	119.010.601	20	0,17	2.183	18,34
1995	99.289.667	21	0,21	1.809	18,22
1996	150.462.594	47	0,31	3.239	21,53
1997	196.564.137	54	0,27	4.697	23,9
1998	211.973.459	70	0,33	4.928	23,25
1999	232.216.317	1.204	5,18	4.709	20,28
		3.838	14,93	2.862	11,13
		4.713	17,35	3.015	11,1
		5.674	17,89	2.975	9,38
		6.787	20,21	2.870	8,55
		7.489	21,46	2.925	8,38
		7.320	20,79	2.454	6,97
		7.941	21,93	3.169	8,75
		6.595	17,14	5.149	13,38
		5.297	13,34	6.801	17,13
		4.204	11,03	7.218	18,94
		3.244	8,02	9.715	24,03
		3.081	7,22	10.785	25,27
		850	5,29	4.468	27,83



Resource Viewer

- specialized for specific data types
- can be fed with data
  `POST` or `GET?data={url}`
- Process
  1. query: Browser -> Content Provider
  2. result: CP -> Browser
  3. Browser -> Viewer
     either:
     - POST result
     - GET with resourceID
     - GET with recordID
     - GET with resultset ID
     - GET resending the query
  4. Viewer -> Content Provider



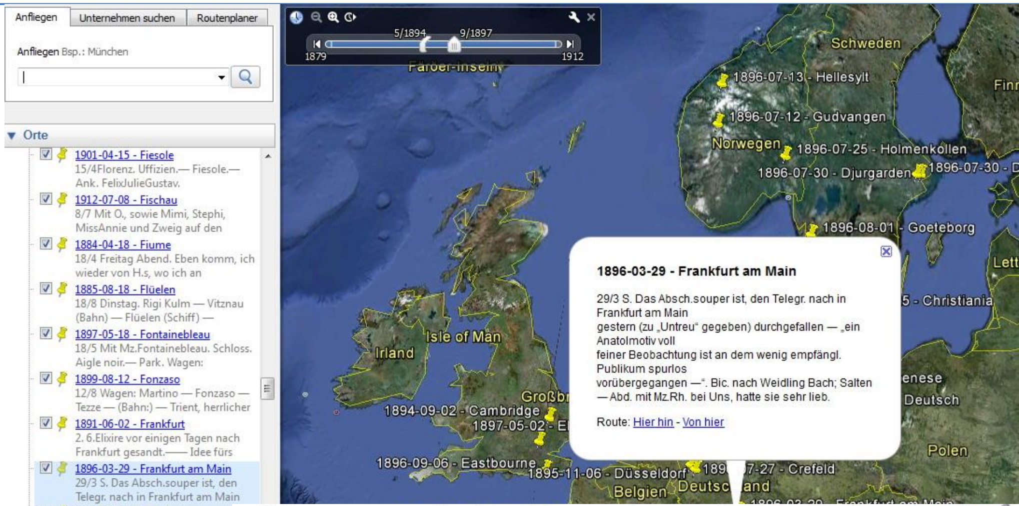


Viewer

ICLTT

different types:

- generic charts
  histogram, treemap
- special Linguistic Visualizations
  Linfovis (DoubleTree, CorpusClouds)
- Space & Time
  TimeLine, TimeMap
- interactive Graphs



☑ 1901-04-15 - Fiesole
15/4Florenz. Uffizien.— Fiesole.— Ank. FelixJulieGustav.

☑ 1912-07-08 - Fischau
8/7 Mit O., sowie Mimi, Stephi, MissAnnie und Zweig auf den

☑ 1884-04-18 - Fiume
18/4 Freitag Abend. Eben komm, ich wieder von H.s, wo ich an

☑ 1885-08-18 - Flüelen
18/8 Dinstag. Rigi Kulm — Vitznau (Bahn) — Flüelen (Schiff) —

☑ 1897-05-18 - Fontainebleau
18/5 Mit Mz.Fontainebleau. Schloss. Aigle noir.— Park. Wagen:

☑ 1899-08-12 - Fonzaso
12/8 Wagen: Martino — Fonzaso — Tezze — (Bahn:) — Trient, herrlicher

☑ 1891-06-02 - Frankfurt
2. 6.Elixire vor einigen Tagen nach Frankfurt gesandt.—— Idee fürs

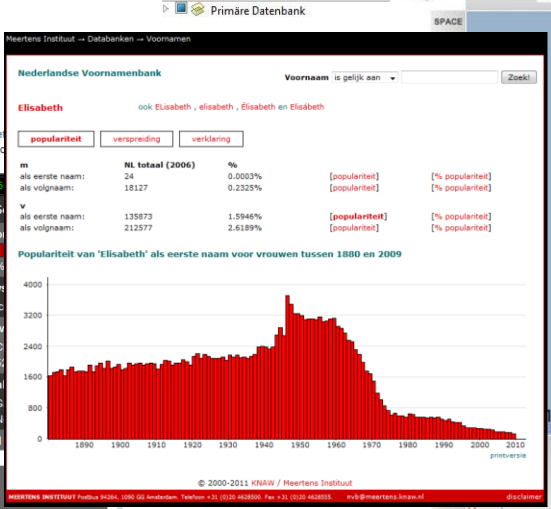
☑ 1896-03-29 - Frankfurt am Main
29/3 S. Das Absch.souper ist, den Telegr. nach in Frankfurt am Main

☑ 1895-09-23 - Frankreich
23/9 „Zur Urne geschritten." Von

1896-03-29 - Frankfurt am Main

29/3 S. Das Absch.souper ist, den Telegr. nach in
Frankfurt am Main
gestern (zu „Untreu" gegeben) durchgefallen — „ein
Anatolmotiv voll
feiner Beobachtung ist an dem wenig empfängl.
Publikum spurlos
vorübergegangen —". Bic. nach Weidling Bach; Salten
— Abd. mit Mz.Rh. bei Uns, hatte sie sehr lieb.

Route: Hier hin - Von hier



Meertens Instituut → Databanken → Voornamen

Nederlandse Voornamenbank

Voornaam is gelijk aan ▾ Zoek!

Elisabeth ook ELisabeth , elisabeth , Élisabeth en Elisábeth

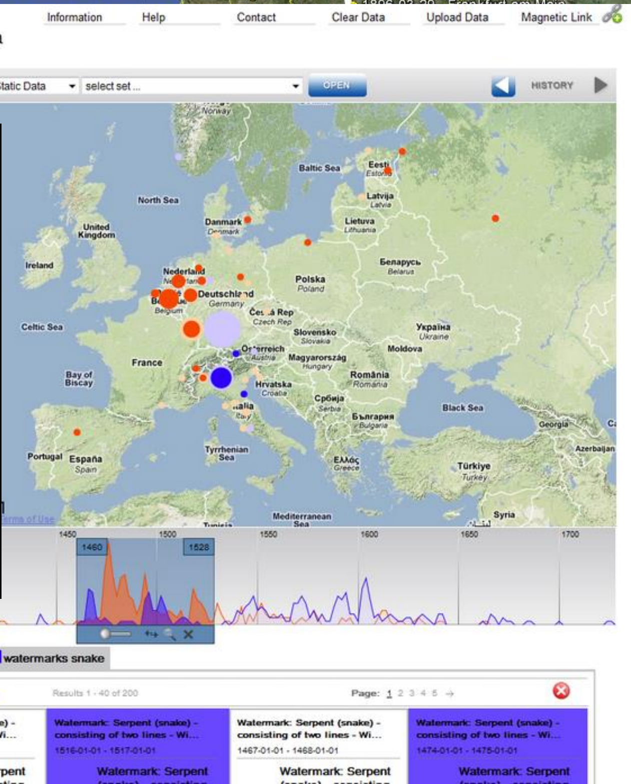
popularitiet | verspreiding | verklaring

m	NL totaal (2006)	%		
als eerste naam:	24	0.0003%	[populariteit]	[% populariteit]
als volgnaam:	18127	0.2325%	[populariteit]	[% populariteit]
v				
als eerste naam:	135873	1.5946%	[populariteit]	[% populariteit]
als volgnaam:	212577	2.6189%	[populariteit]	[% populariteit]

Populariteit van 'Elisabeth' als eerste naam voor vrouwen tussen 1880 en 2009

© 2000-2011 KNAW / Meertens Instituut

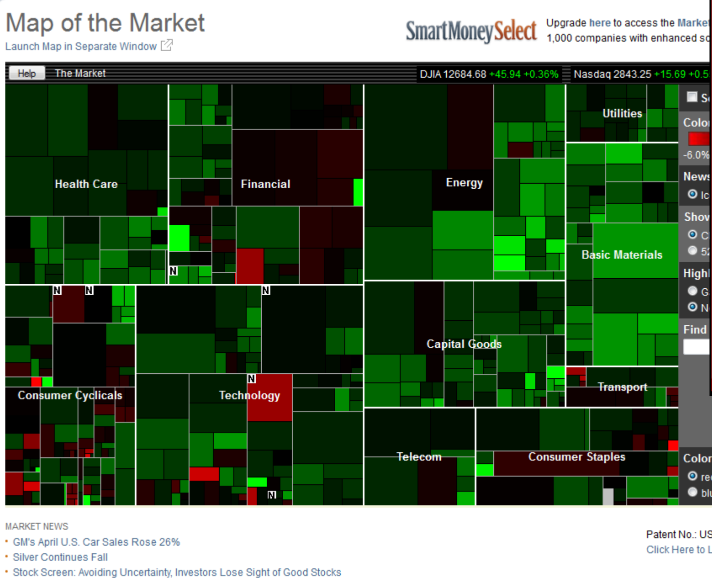
MEERTENS INSTITUUT Postbus 94264, 1090 GG Amsterdam. Telefoon +31 (0)20 4628500. Fax +31 (0)20 4628555. nvb@meertens.knaw.nl disclaimer



🟥 watermarks lily 🟦 watermarks snake

Results 1 - 40 of 200

Watermark: Serpent (snake) - consisting of two lines - Wi... 1541-01-01 - 1542-01-01	Watermark: Serpent (snake) - consisting of two lines - Wi... 1516-01-01 - 1517-01-01	Watermark: Serpent (snake) - consisting of two lines - Wi... 1467-01-01 - 1468-01-01	Watermark: Serpent (snake) - consisting of two lines - Wi... 1474-01-01 - 1475-01-01
Watermark: Serpent	Watermark: Serpent	Watermark: Serpent	Watermark: Serpent



Map of the Market

Launch Map in Separate Window

SmartMoneySelect Upgrade here to access the Market 1,000 companies with enhanced sc

MARKET NEWS
- GM's April U.S. Car Sales Rose 26%
- Silver Continues Fall
- Stock Screen: Avoiding Uncertainty, Investors Lose Sight of Good Stocks

Patent No.: US 6,583,794 B1
Click Here to License the Map Applet

Combined Distributed Metadata Content Search

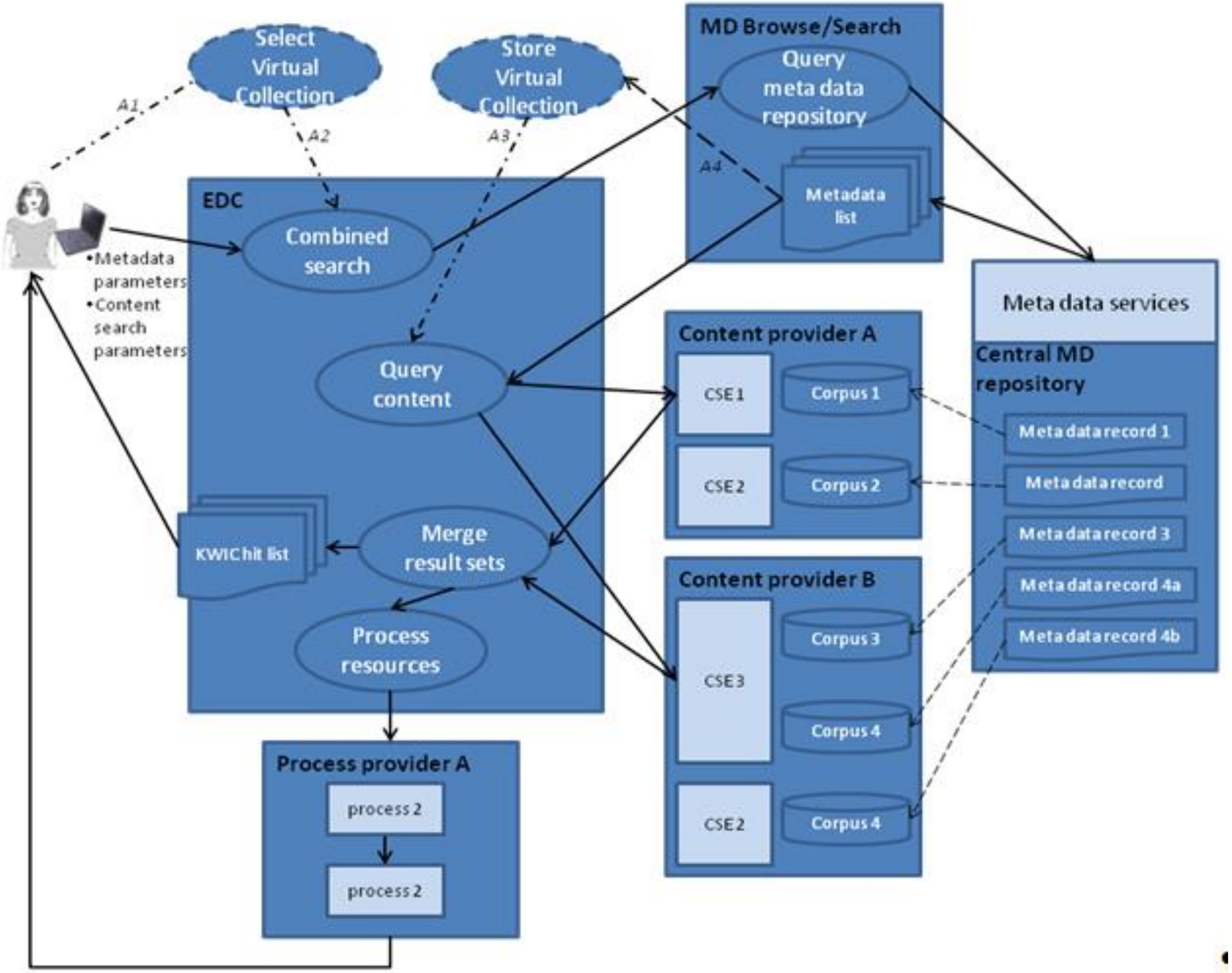
ICLTT

2 phases:

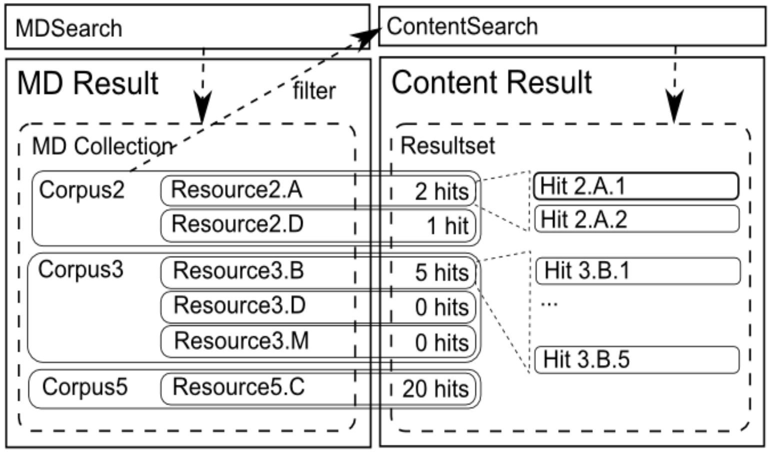
1. MD Search
   find candidate resources (collections) based on the MD

   (+ wrap with
   up to a parent-collection
   with specified endpoint)

2. FCS
   query repositories with candidate resources



multi-result



- resultID
- status = running|finished
- TTL!
- how to restrict
- variants for the (multi)result-set:
  - one flat list - every hit from every endpoint is one result-item (sru:record) `fcs:Resouce` identifying the endpoint

```
<sru:record><sru:recordData>
   <fcs:Resource pid="endpoint1" >
      <fcs:RF><fcs:DataView type="kwic"><kwic:kwic>
<sru:record><sru:recordData>
   <fcs:Resource pid="endpoint1" >
      <fcs:RF><fcs:DataView type="kwic"><kwic:kwic>
<sru:record><sru:recordData>
   <fcs:Resource pid="endpoint2" >
      <fcs:RF><fcs:DataView type="kwic"><kwic:kwic>
```

  - one record per endpoint (with summary) + pointer to the result from one endpoint

```
<sru:resultID>r1
<sru:record><sru:recordData>
   <fcs:Resource ref="search?resultId=r1&x-context=endpoint1" >
<sru:record><sru:recordData>
   <fcs:Resource ref="search?resultId=r1&x-context=endpoint2" >
/* OR */
<sru:record><sru:recordData>
   <sru:recordIdentifer>search?resultId=r1&x-context=endpoint2
```

multi-result

- facetedResult
  SRU 2.0,
  summary over data sources
  AND facets (=indexes)
- searchResultAnalysis ?
  meant to indicate results for
  parts of complex query

```
<facetedResults xmlns="http://docs.oasis-
open.org/ns/search-ws/sru-facetedResults" >
<datasource>
 <!-- first data source -->
 <datasourceDisplayLabel>LC</datasourceDisplayLabel>
 <datasourceDescription>Library of
Congress</datasourceDescription>
 <baseURL> http://z3950.loc.gov:7090/voyager</baseURL>
 <facets>
  <facet>
    <facetDisplayLabel> subject</facetDisplayLabel>
    <facetDescription> Dublin Core
Subject</facetDescription>
    <index> dc.subject</index>
    <relation>=</relation>
    <terms>
      <term><actualTerm>birds</actualTerm>
       <query>nuthaches AND dc.subject=birds</query>
       <requestUrl> http://z3950.loc.gov:7090/voyager
          ?query=”nuthaches%20AND%20dc.subject=birds”
       </requestUrl>
       <count>12 </count>
      </term>
```

Thank you, questions?

Questions

- scan with ?x-context
    - fcs.resource?x-context= vs. fcs.resource?x-context=dict-gate

.http://weblicht.sfs.uni-tuebingen.de/rws/sru/?operation=scan&scanClause=fcs.resource&version=1.1
operation with 'scanClause' with value 'fcs.resource' is deprecated within CLARIN-FCS

```
<xsd:element ref="recordPacking" minOccurs="0"/>
<xsd:element ref="recordSchema" minOccurs="0"/>
```

Older extra slides on extensions

FCS Extensions – overview

- **fcs:Resource.xsd**
  generic schema for recordData
- **x-context + fcs.resource**
  extension parameter to restrict search domain, with corresponding index providing the values
- x-format, x-dataview
  what and how to deliver
- new context sets
  - **isocat**
  - **fcs**
  - **cmd**
- dynamic Indices
  not defined statically in the context-set, but every endpoint announces it's indices individually
- nested scan-response
- Sequential Tier Search
- binding Indices

Extension – Result Format – Resource.xsd

- generic schema to go inside `sru:recordData`
- data inline or by reference
- 3 Elements: **`Resource/ResourceFragment/DataView`**
- 3 attributes:
  - **`@pid/@ref`** := identify/reference resources and their parts
  - **`@type`**~~`/@schema`~~ := indicate type of the data

```
<fcs:Resource pid="123“>
  <fcs:ResourceFragment pid="123#a“>
    <fcs:DataView type="text/xml"><meertens:any/>
    </fcs:DataView>
    <fcs:DataView type="image/jpeg" ref=”{URI}”></fcs:DataView>
  </fcs:ResourceFragment>
</fcs:Resource>
```

- http://www.clarin.eu/system/files/Resource.xsd, (also in svn:/FederatedSearch)
- `@schema`-attribute dropped (namespace should be enough)
- ?? value domain of `@type`-attribute?
  (mime-type + „other“? some resource-type taxonomy?)

Extension – fcs.resource + x-context

- restrict the request (explain, scan, searchRetrieve) to a set of "resources" identified by PID:
  - **repositories** – basic federated search
  - **collections** – even in non-federated search
  - single **resource** – within a repository
  - **virtual collection**
- announcing repositories Center Registry (or explain (F&N) ?)
- announcing collections (misuse scan)

```
?operation=scan
&scanClause=fcs.resource (={resource-handle})?
```

- specifying Collections in the request

```
param: ?x-context={resource-id}   /* OR */
CQL:   {search-term} AND fcs.resource= {resource-handle}

CQL:   fcs.resource= {resource-handle} /* should return the resource */
```

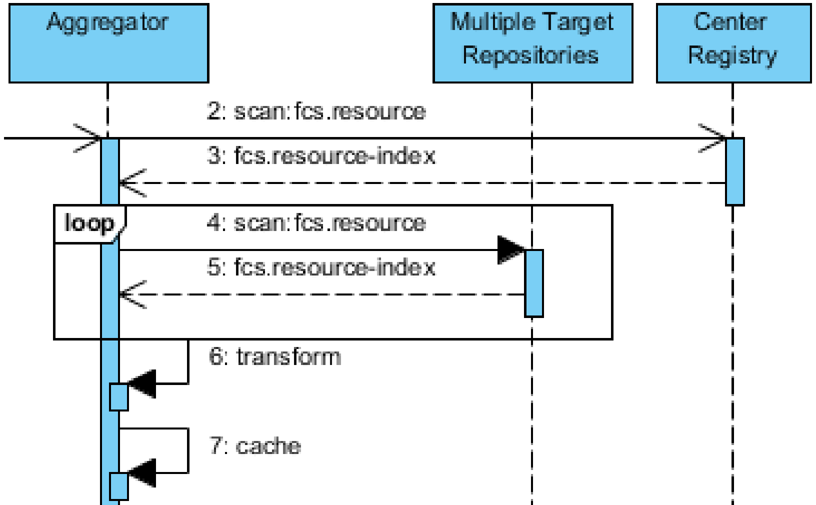
```
+ MPI
  + ESF
    + ...
      > Res1
  + Childes

+ C4
  + Basel
  + Bozen
  + Berlin
  + Wien

+ MIMORE
  + DynaSAND
  + DiDDD
  + GTRP
+ INL
```

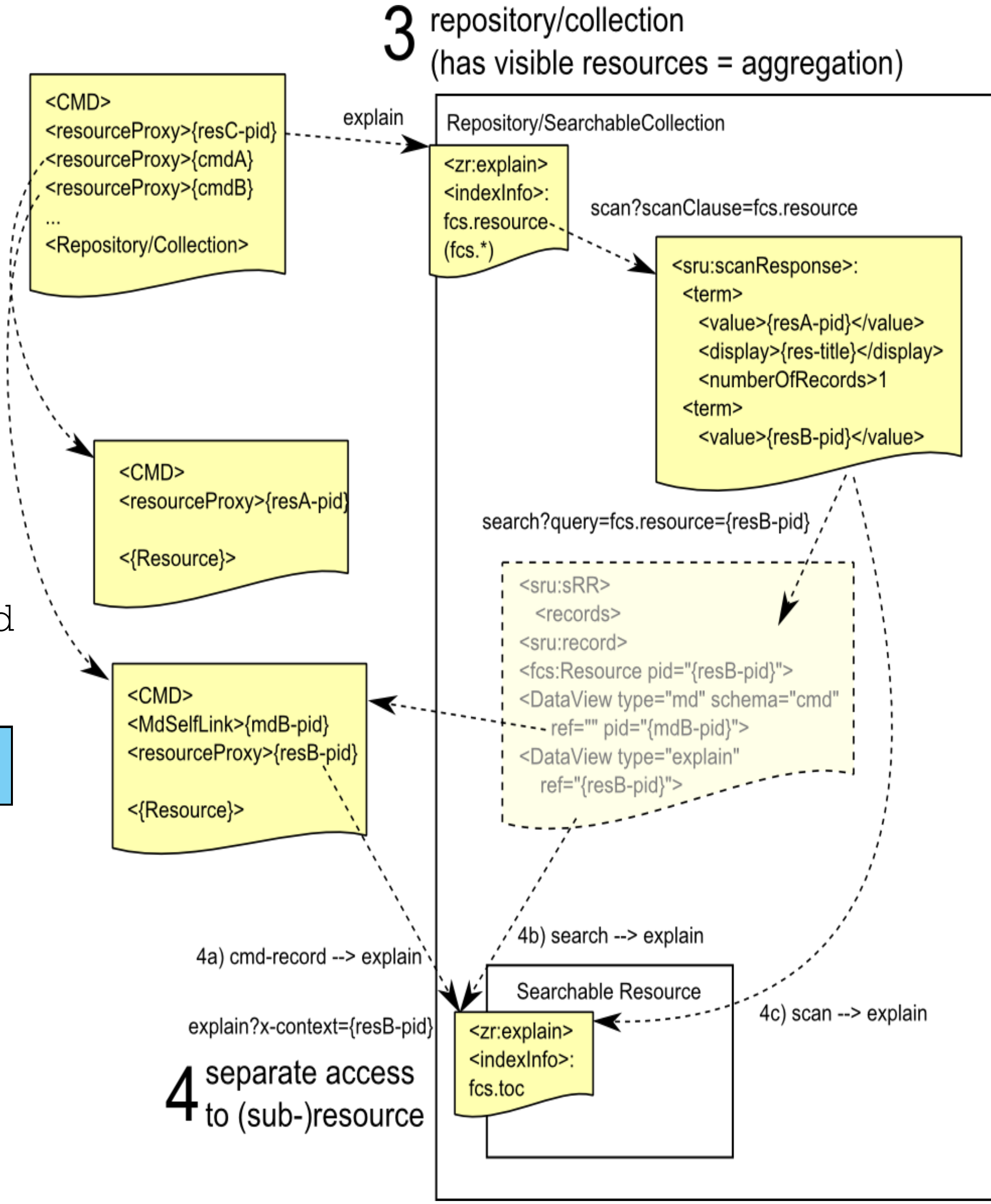
– fcs.resource + x-context

ICLTT

index of searchable resources

- 1st level = repositories in CenterRegistry
- repositories optionally expose resources via `scan:fcs.resource`
- by crawling `scan:fcs.resource` a recursive index is built up. (in practice 2 levels ?)
- aggregator has inverted map `res-id -> providing repository` so confronted with `?x-context=res-id` it knows where to root the query to.





3 repository/collection
(has visible resources = aggregation)

<CMD>
<resourceProxy>{resC-pid}
<resourceProxy>{cmdA}
<resourceProxy>{cmdB}
...
<Repository/Collection>

explain

Repository/SearchableCollection

<zr:explain>
<indexInfo>:
fcs.resource
(fcs.*)

scan?scanClause=fcs.resource

<sru:scanResponse>:
<term>
<value>{resA-pid}</value>
<display>{res-title}</display>
<numberOfRecords>1
<term>
<value>{resB-pid}</value>

<CMD>
<resourceProxy>{resA-pid}
<{Resource}>

search?query=fcs.resource={resB-pid}

<sru:sRR>
<records>
<sru:record>
<fcs:Resource pid="{resB-pid}">
<DataView type="md" schema="cmd"
ref="" pid="{mdB-pid}">
<DataView type="explain"
ref="{resB-pid}">

<CMD>
<MdSelfLink>{mdB-pid}
<resourceProxy>{resB-pid}
<{Resource}>

4a) cmd-record --> explain

4b) search --> explain

4c) scan --> explain

explain?x-context={resB-pid}

Searchable Resource

<zr:explain>
<indexInfo>:
fcs.toc

4 separate access
to (sub-)resource

Extension – nested scan

- needed for collections/resources

  cmd.collection, fcs.resource – index

- needs another extra parameter: `x-maximumDepth`

```
<?xml version="1.0" encoding="utf-8"?><sru:scanResponse
xmlns:sru="http://www.loc.gov/zing/srw/">  <sru:version>1.2</sru:version>
  <sru:terms>
    <sru:term>
      <sru:value>clarin.at:icltt:ddc</sru:value>
      <sru:numberOfRecords>3</sru:numberOfRecords>
      <sru:displayTerm>Text corpora by ICLTT on DDC</sru:displayTerm>
      <sru:extraTermData>
        <sru:terms>
          <sru:term>
            <sru:value>clarin.at:icltt:ddc:traum_deu</sru:value>
            <sru:numberOfRecords>1</sru:numberOfRecords>
            <sru:displayTerm>Freud: Die Traumdeutung, German</sru:displayTerm>
          </sru:term>
          <sru:term>
            <sru:value>clarin.at:icltt:ddc:barock</sru:value>
            <sru:numberOfRecords>1</sru:numberOfRecords>
            <sru:displayTerm>Barocktexte</sru:displayTerm>
          </sru:term>
        </sru:terms>
      </sru:extraTermData>
    </sru:term>
    <sru:term>
      <sru:value>clarin.at:icltt:ddc</sru:value>
      <sru:numberOfRecords>1</sru:numberOfRecords>
      <sru:displayTerm>Text corpora by ICLTT on DDC</sru:displayTerm>
      <sru:extraTermData>
        <sru:terms>
          <sru:term>
            <sru:value>clarin.at:icltt:ddc:c4</sru:value>
            <sru:numberOfRecords>1</sru:numberOfRecords>
            <sru:displayTerm>C4 Vienna</sru:displayTerm>
          </sru:term>
        </sru:terms>
```

Extension – new context sets

Define new Context Sets:

- **isocat**
  preferred way, every endpoint should try to map internally and expose indexes as isocat data categories

```
isocat.DC-1324  isocat.lemma
isocat.DC-1403  isocat.token
…
```

- **fcs**
  only if isocat does not provide an equivalent, but again what would be allowed indexes?
  index for every aspect of a Tier/AnnotationLayer: **TierType**, TierName, Participant ?

```
fcs.TierType.w                         fcs.TierName.? /*open domain!*/
 =? fcs.w                               fcs.TierName.V40069-Spch

fcs.TierType.PoS                       fcs.Participant.?
 =? fcs.pos     =? isocat.PoS           fcs.Participant.V40069
=?fcs.TierType.isocat.partOfSpeech
```

- (**cmd**)
  searching in MD, (path-like) index for *every Profile/Component/Element*

```
cmd.Project.Name                    cmd.Collection.Project.Title
cmd.Actor.Name                      cmd.title
cmd.Name  /* delib ambig */         cmd.Actor.Role
```

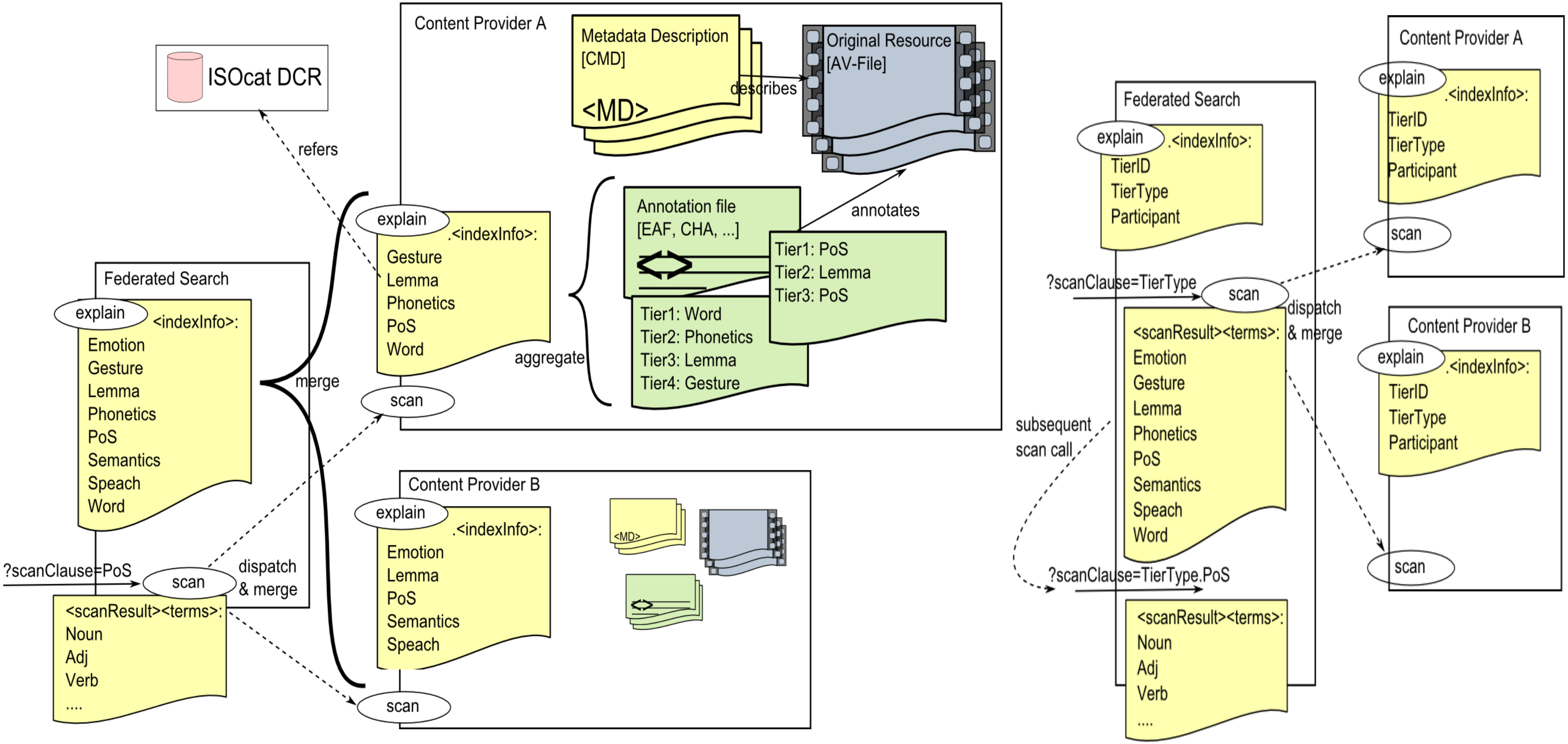
- requires dynamic Indices
  not supported by SRU, but formal syntax for Context Sets undefined(?) anyhow

– dynamic indices – federated announce

A) all indices in explain (possibly unbearable bloating of the explain-response)

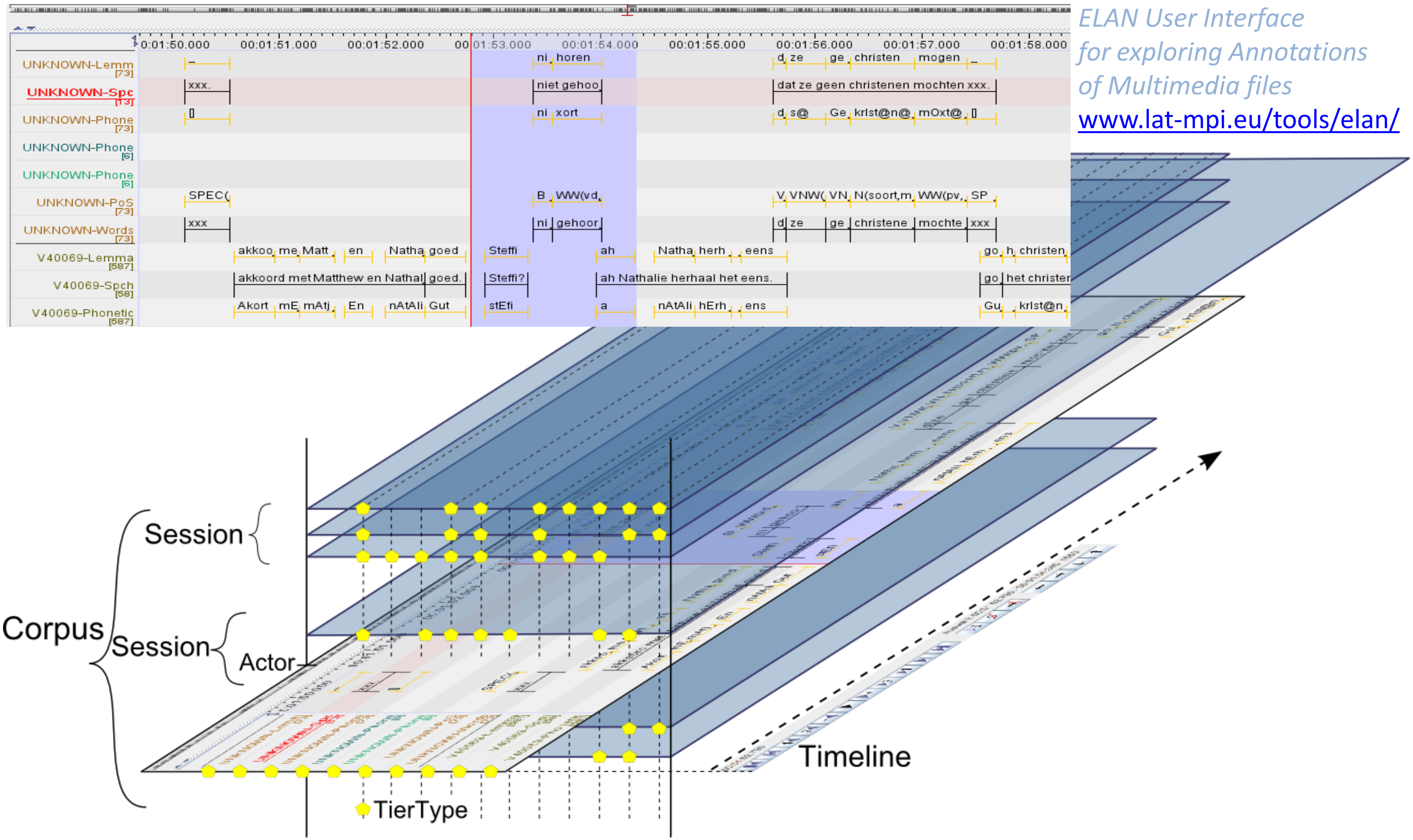
```
TierType:English, TierType:PoS, TierType:Word, TierType:Gesture
TierName:I'sGest, TierName:Damian, TierName:Unknown.WORD,
```

B) only static explain + misuse scan: `TierType, TierName, Participant`



– dynamic indices – TierType

TierType as the most usable aspect of annotation layers, allowing to search in related Tiers across Participants, Sessions and even Corpora


0:01:50.000
00:01:51.000
00:01:52.000
00:01:53.000
00:01:54.000
00:01:55.000
00:01:56.000
00:01:57.000
00:01:58.000
UNKNOWN-Lemm [73]
UNKNOWN-Spc [73]
UNKNOWN-Phone [73]
UNKNOWN-Phone [6]
UNKNOWN-Phone [6]
UNKNOWN-PoS [73]
UNKNOWN-Words [73]
V40069-Lemma [587]
V40069-Spch [58]
V40069-Phonetic [587]
– ni horen d ze ge christen mogen –
xxx. niet gehoo dat ze geen christenen mochten xxx.
[] ni xort d s@ Ge krIst@n@ mOxt@ []
SPEC(B WW(vd, V VNW(VN N(soort,m, WW(pv, SP
xxx ni gehoor d ze ge christene mochte xxx
akkoo me Matt en Natha goed Steffi ah Natha herh eens go h christen
akkoord met Matthew en Nathal goed. Steffi? ah Nathalie herhaal het eens. go het christen
Akort mE mAtj En nAtAli Gut stEfi a nAtAli hErh ens Gu krIst@n
Session
Session
Corpus
Actor
TierType
Timeline

ELAN User Interface for exploring Annotations of Multimedia files
www.lat-mpi.eu/tools/elan/

– dynamic indices in SRU – FCS

- FCS explain:

```
<explain>
 <indexInfo>
   <set name=“fcs" identifier=“http://clarin.eu/fcs/1.0"/>
                              /*  or: info:srw/schema/102/fcs? */
   <set name=“isocat" identifier=“http://isocat.org"/>
   <set name=“dc" identifier="info:srw/cql-context-set/1/dc"/>
/* variants!: */
    <index search="true" scan=“false" sort="false">
        <title lang="en">Word</title>
        <map><name set=“fcs">w</name></map></index>
    <index><map><name set=“fcs">word</name></map></index>
    <index><map><name set=“fcs">TierType.w</name></map></index>
    <index><map><name set=“fcs">TierType.word</name></map></index>
    <index><map><name set=“isocat">token</name></map></index>
    <index><map><name set=“isocat">DC-1403</name></map></index>
    <index search="true" scan=“true" sort="false“>
        <title lang="en">Part of Speech</title>
        <map><name set=“fcs">TierType.pos</name></map></index>
  /* But!: */
    <index><map><name set=“fcs">TierName</name></map></index> /* ? */

    /* OR: */
      <index><map><name set=“fcs">TierType</name></map></index>
      <index><map><name set=“fcs">TierName</name></map></index>
      <index><map><name set=“fcs">Participant</name></map></index>
</indexInfo>
```

– dynamic indices in SRU – CMD

- CMD explain:

```
<explain>
 <indexInfo>
   <set name=“cmd" identifier="info:srw/cql-context-set/101/CMD"/>
   <set name=“dc" identifier="info:srw/cql-context-set/1/dc"/>
   <set name=“imdi" identifier="info:srw/cql-context-set/3/IMDI-Session"/>

  /* variants!: */
    <index><title lang="en">DC Title</title>
             <map><name set=“cmd">  dc.title </name></map></index>
    <index><map><name set="dc">   title  </name></map></index>
    <index><map><name set=“cmd">  title  </name></map></index>
    <index><map><name set=“cmd">  Project.Title </name></map></index>
    <index><map><name set=“cmd">  Actor.Role </name></map></index>
    <index><map><name set=“cmd">  Session.Actor.Role </name></map></index>
    <index><map><name set=“cmd">  imdi.Actor.Role </name></map></index>
    <index><map><name set=“imdi“> Actor.Role </name></map></index>
 </indexInfo>
 <schemaInfo>
    <schema name="dc" identifier="info:srw/schema/1/dc-v1.1">
             <title>Simple Dublin Core</title></schema>
    <schema name=“cmd" identifier="info:srw/schema/101/cmd">
             <title>Component Metadata</title></schema>
    ...
```

Extensions II – Sequential Tier Search

- CQL: a) provides boolean operator PROX

```
Herz PROX/unit=sentence/distance=0 zerreißen
Actor.w = „Ja“ PROX/seconds/4 Actor.emotion=laugh
```

b) proposes `window` and `element` in CQL 2.0

```
word all/windowSize=10 “hat cat rat”
bib.name ="adam smith" PROX/element=bib.author dc.date =1965

/* my unifying proposal: */
bib.name ="adam smith" PROX/unit=bib.author/0 dc.date =1965
bib.name ="adam smith" PROX/unit=bib.author/>0 dc.date =1965 /* other */
```

However this is limited either to only two operands or to simple terms

- Therefore proposal of a new boolean operator : IN or HAS

  However not CQL anymore.

```
( Q1 AND Q2 AND Q3) IN Q4?

( Actor.X.w=Ja PROX/w/4 Actor.Y.emotion =laugh
    AND Actor.Z.gesture=”clap hands”
    AND Actor.w adj “wonderful feeling”
 ) IN  Paragraph  /* or: */  ) IN PROX/min/2
```

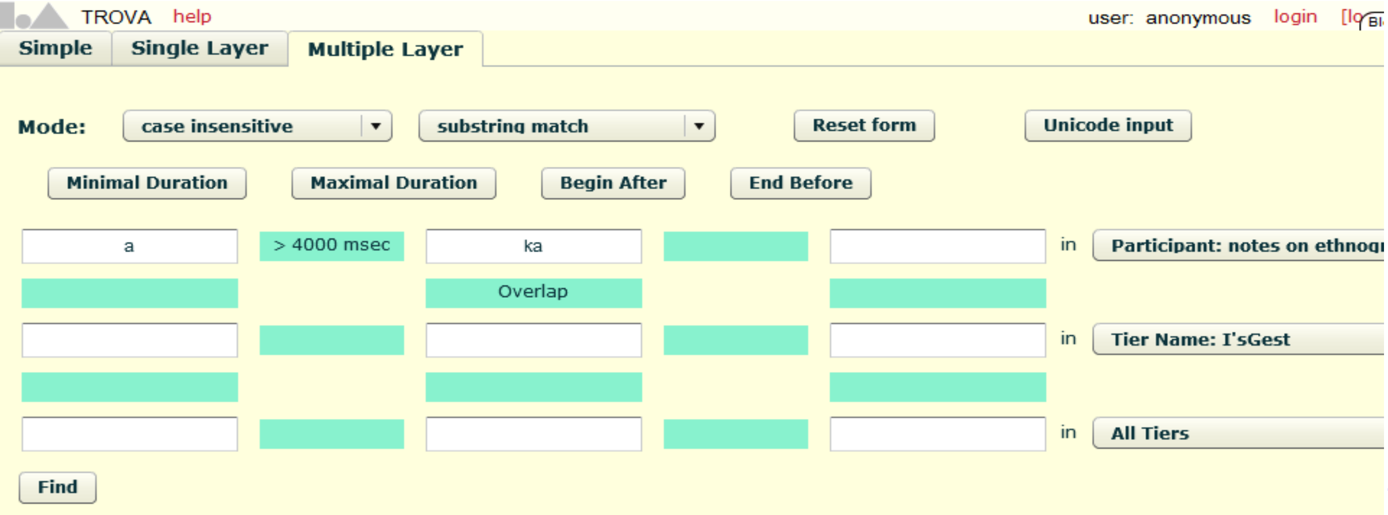
- Binary chain:

```
( Q1  PROX/{modifiers}/a  Q2  PROX//a  Q3 )

(Actor.X.w=Ja PROX/w/4 Actor.Y.emotion =laugh)
 PROX/min/2/a Actor.Z.gesture=”clap hands”
 PROX//a Actor.w adj “wonderful feeling”
```

- Other ideas?

– STS - Alignement

- Aligned! tiers (primary track/sequence: AV-file, tokens)
  → AnnotationGraph (Bird & Liberman)?



TROVA help user: anonymous login [lo

Mode: case insensitive ▾ substring match ▾ Reset form Unicode input

Minimal Duration Maximal Duration Begin After End Before

a	> 4000 msec	ka			in	Participant: notes on ethnogr
		Overlap				
					in	Tier Name: I'sGest
					in	All Tiers

Find

modifiers:

/ Fully aligned

/ Overlap

/ Left Overlap

/ Right Overlap

/ Surrounding

/ Within

/ No Constraint

/ Clear

{ All combinations of: begin/end time, and =/>/< }

TROVA multi-layer search

www.lat-mpi.eu/tools/annex/

tier:		1	2	3	4	5	6			
timecode (original-track)										
segmentation										
allocation		1	1	1	1	2	2	pause	1	1
w		Ist	sie	da	?	Ja	.		Und	?
w.pos		V	PRO	PROP	\$.	ja	\$.			
w.lemma		sein	sie	da	\$.		\$.			
s		1	1	1	1	2	2		3	3
Actor1		x	x	x	x					
Actor1.emotion		Suspension					Relief			
Actor1.gesture										
Actor2										
Actor2.emotion										
Actor2.gesture						Hand on shoulder				

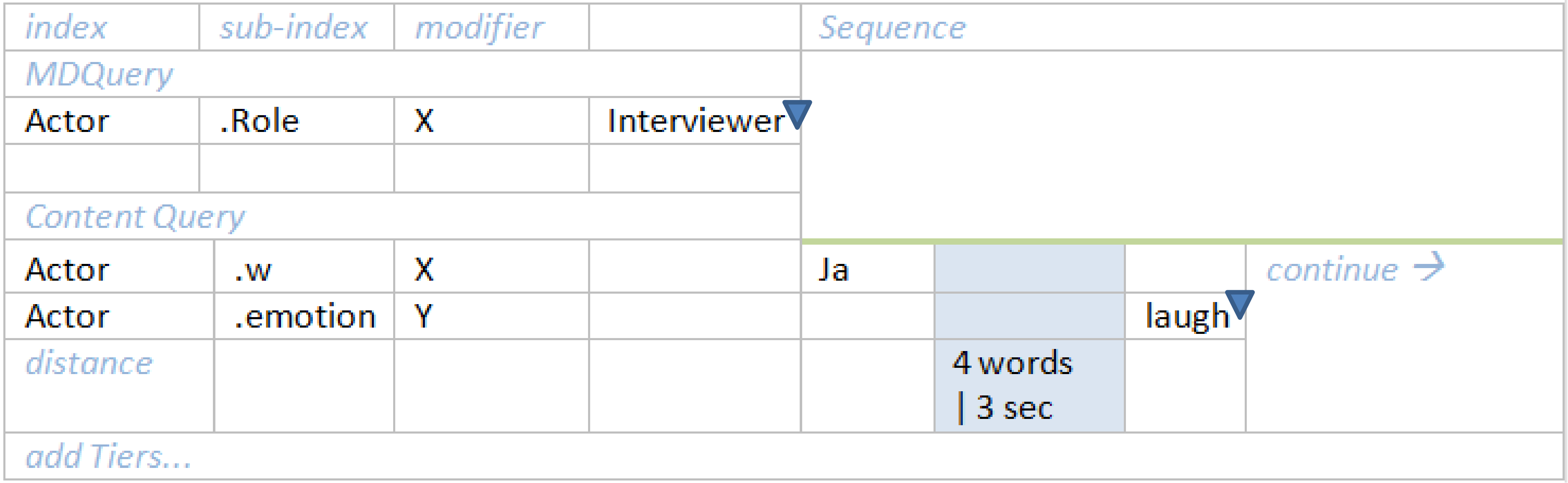
Extensions II – binding indices

- Binding Indices

```
{index} {relation}/var=(X|Y|Z,…) {term}

Actor.Role =/var=X Annotator AND Actor.Age >/var=X 40
AND Actor.Role =/var=Y Speaker   AND Actor.Sex =/var=Y Female

Actor.(X).Role   /* shorthand */

TierType.PoS =/var=X noun PROX/s/0 TierType.PoS =/var=Y verb
```

Combined Metadata Content Query with Sequence and bound Indices

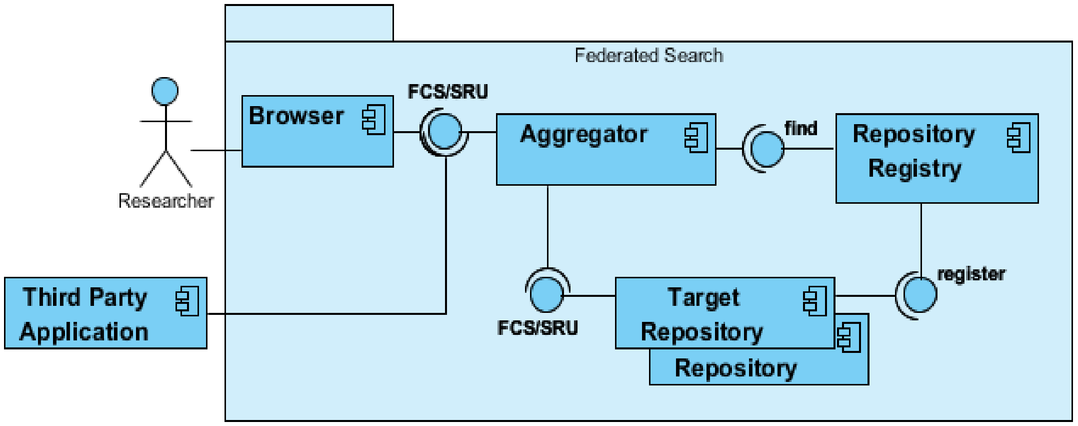
```
Actor.(X).Role = Interviewer AND
(  (Actor.(X).w = „Ja�much PROX/words/4 Actor.(Y).emotion=laugh)
OR (Actor.(X).w = „Ja‟ PROX/sec/3 Actor.(Y).emotion=laugh)  )
```

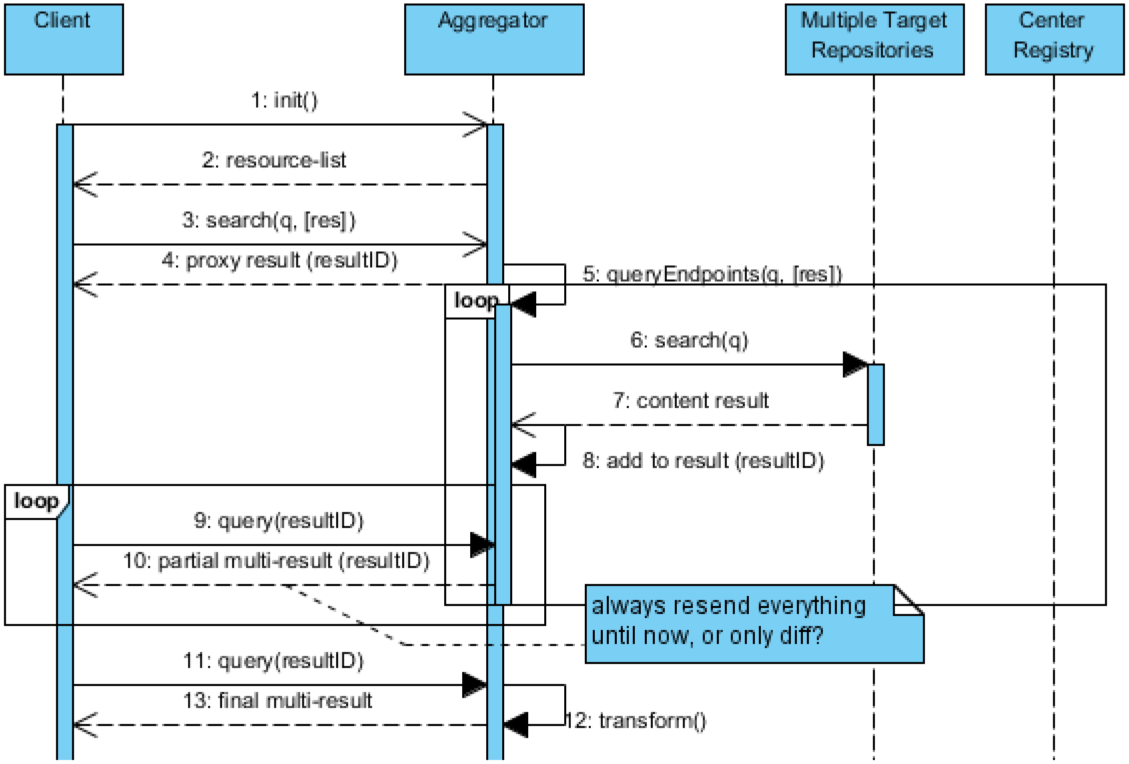


index	sub-index	modifier		Sequence			
MDQuery							
Actor	.Role	X	Interviewer				
Content Query							
Actor	.w	X		Ja			continue →
Actor	.emotion	Y				laugh	
distance					4 words | 3 sec		
add Tiers...							

Aggregator

- requirements:
  - also a FCS/SRU-endpoint
  - asynchronous (not waiting for the slowest one)
  - but still session-less
  - “multiResult”
    - “merge” not possible - retain provenance
    - summary over data sources
  - every match is one result-item (`sru:recordData`) as opposed to e.g. one Resource with many hits being one item in the result set
- solution(?):
  - resultID as ticket
  - client keeps asking
  - aggregator delivers summary of intermediate result + status

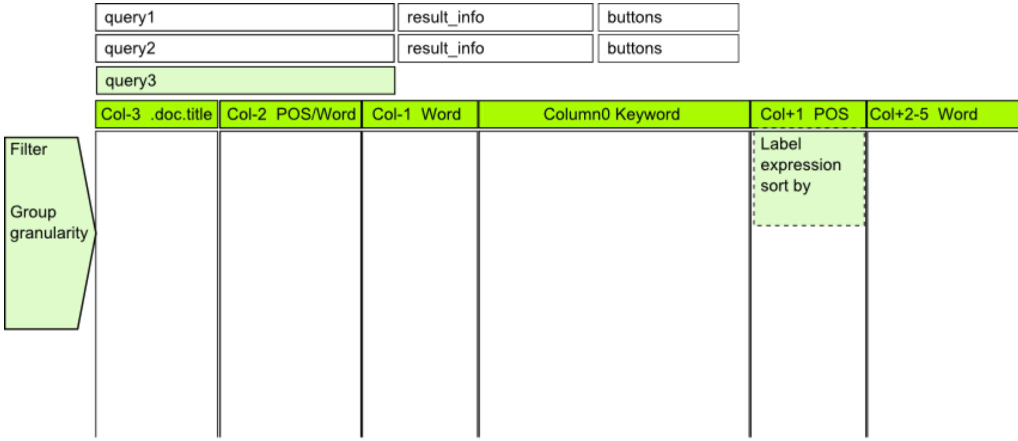




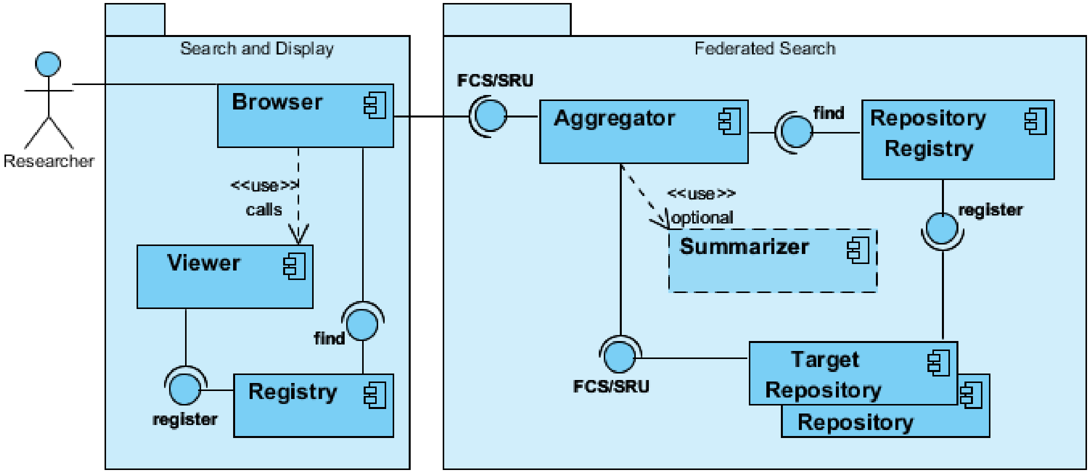
always resend everything until now, or only diff?

Aggregator – User Interface

- allow multiple queries/results
- allow selecting indices
  - in query-input,
  - to display in the result
  - for grouping/sorting result


query1 | result_info | buttons
query2 | result_info | buttons
query3
Col-3 .doc.title | Col-2 POS/Word | Col-1 Word | Column0 Keyword | Col+1 POS | Col+2-5 Word
Filter
Group granularity
Label expression sort by

- Browser knows the Viewers (per Type) OR endpoint already delivers link to Viewer
- multiple Viewers for one type possible
- Visualization may need summarized data
  either provided by target repository or a specialized Summarizer-module as fall-back (inefficient)



Summary - main issues

- `fcs.resource`
  as distributed hierarchical index
- value domain for `DataView@type`
  kwic, title, metadata,
  {mime-type}: application/tcf, application/eaf, application/kml
- multi-result
- announcing indices
- agree on new (optional) parameters:

```
  *?x-format   search?x-dataview    scan?x-maximumDepth
```

- ResourceViewer
  for various data-types
- Visualization (+ Aggregation)
  - TimeLine (different scales: for Metadata: years to days, for content: seconds)

CQL-Examples - Metadata queries

cmdIndex

```
>clarin.eu:2625   >Actor.Contact.Phone
>Session.Project.Name
```

Basic

```
>dc.title adj "open access"
>dc.date > 1900
```

Boolean operators

```
>Organisation any University
 and (dc.language=de or cmd.Country=Austria)
 and (dc.title any Liebe or cmd.Author any Trakl)
```

•Alternatives

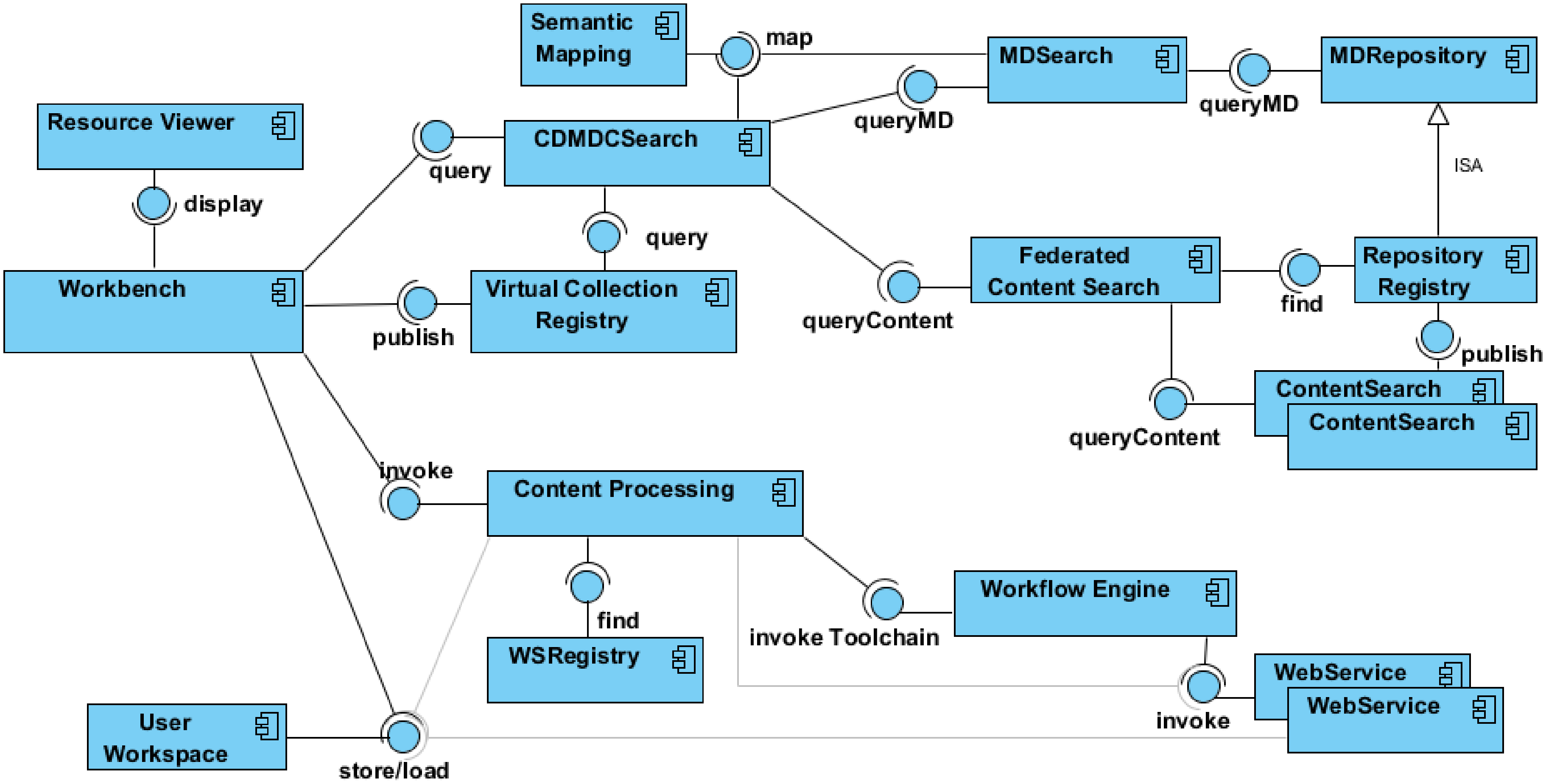
```
>cmd.genre = (opera or novel or fantasy)
>cmd.genre any "opera novel fantasy"
```

- Multiple conditions to **same** Component/**Element** -> new modifier:

```
>Actor.gender =/var=X f and Actor.age >/var=X 15 /*bind-variables */
/* CQL 2.0 proposal: */
>bib.name=”Adam Smith” PROX/element=bib.author dc.date=1965
```

Reference architecture

ICLTT


Semantic Mapping
map
MDSearch
MDRepository
queryMD
Resource Viewer
query
CDMDCSearch
queryMD
ISA
display
query
Workbench
Virtual Collection Registry
Federated Content Search
find
Repository Registry
publish
queryContent
publish
ContentSearch
ContentSearch
queryContent
invoke
Content Processing
find
Workflow Engine
WSRegistry
invoke Toolchain
WebService
WebService
User Workspace
invoke
store/load