MTAS

Multi Tier Annotation Search

practicalities

Matthijs Brouwer

Installation

- Version number Mtas follows numbering Solr / Lucene
- Correct Solr version has to be installed
- Create a new core

Add files
- mtas-6.6.1.jar (in the lib directory)
- commons-math3-3.5.jar (in the lib directory)
- mapping files

In solrconfig.xml
- Define Mtas searchComponent
- Define the Mtas CQL queryParser
- Define a new updateRequestProcessorChain
- Define or adjust the update requestHandler

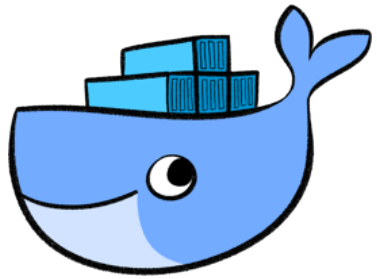
In schema.xml
- Define the Mtas fieldType(s) with reference(s) to mapping
- Add field(s) with these Mtas fieldType(s)

```
demo
demo/conf
demo/core.properties
demo/data
demo/data/index
demo/data/index/segments_1
demo/data/index/write.lock
demo/data/snapshot_metadata
demo/data/tlog
demo/lib
demo/schema.xml
demo/solrconfig.xml
```

```
demo/folia.xml
demo/lib/mtas-6.6.1.jar
demo/lib/commons-math3-3.5.jar
```

Documentation and examples on GitHub

Docker



```
docker build -t mtas https://raw.githubusercontent.com/meertensinstituut/mtas/master/docker/Dockerfile
docker run -t -i -p 8080:80 --name mtas mtas
```

Demonstration
Experiment
Installation
Latest release

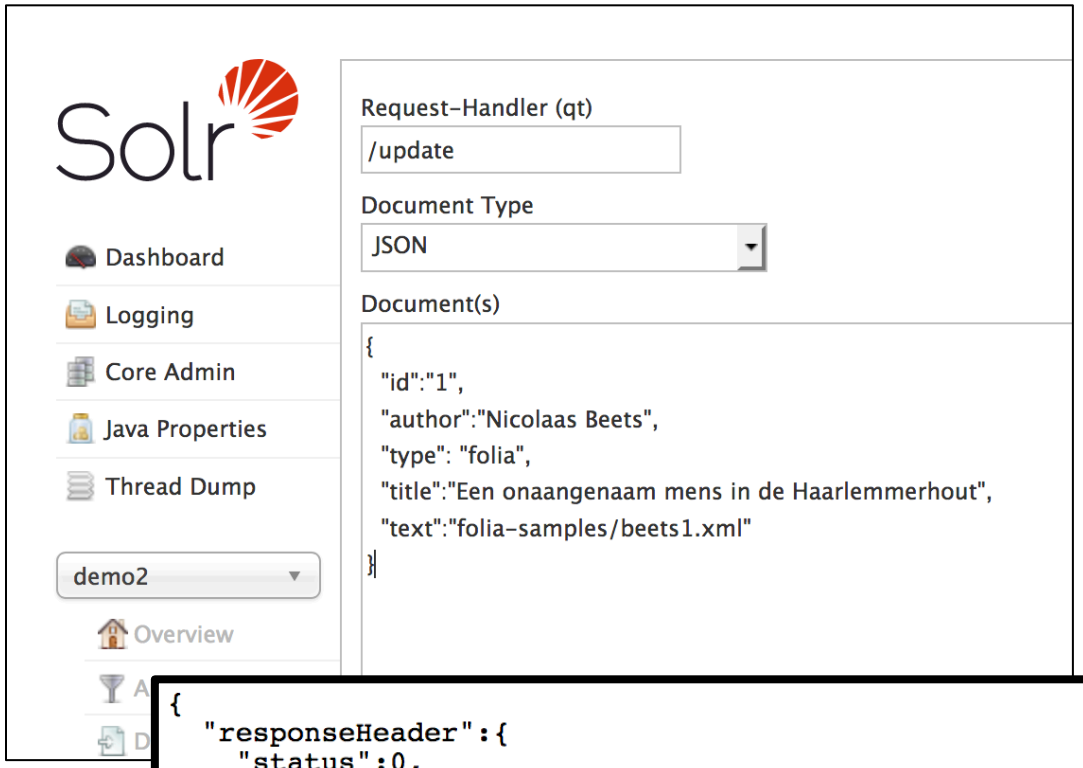
Examples

- FoLiA
- ISO-TEI
- CHAT

ubuntu

Solr

MTAS




```
{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"*:*",
      "indent":"on",
      "wt":"json"}},
  "response":{"numFound":1,"start":0,"docs":[
      {
        "id":"1",
        "author":"Nicolaas Beets",
        "type":"folia",
        "title":"Een onaangenaam mens in de Haarlemmerhout",
        "text":"folia-samples/beets1.xml",
        "size":6920072,
        "numberOfPositions":7419,
        "numberOfTokens":85924,
        "_version_":1579351387560476672}]
  }}
```



```
<text xml:id="untitled.text">
 <p xml:id="untitled.p.1">
  <t>Een onaangenaam mens in de Haarlemmerhout</t>
  <s xml:id="untitled.p.1.s.1">
   <t>Een onaangenaam mens in de Haarlemmerhout</t>
   <w xml:id="untitled.p.1.s.1.w.1" class="WORD">
    <t>Een</t>
    <pos class="LID(onbep,stan,agr)" confidence="0.981771" head="LID">
     <feat class="onbep" subset="lwtype"/>
     <feat class="stan" subset="naamval"/>
     <feat class="agr" subset="npagr"/>
    </pos>
    <morphology>
     <morpheme>
      <t offset="0">een</t>
     </morpheme>
    </morphology>
    <lemma class="een"/>
   </w>
   ....
```


/var/www/html/demo/**folia-samples/beets1.xml**
http://localhost/demo/**folia-samples/beets1.xml**

```
<mapping type="wordAnnotation" name="lemma">
 <token type="string" parent="false">
  <pre>
   <item type="name" />
  </pre>
  <post>
   <item type="attribute" name="class" />
  </post>
 </token>
 <condition>
  <item type="attribute" name="class" />
 </condition>
</mapping>
```

```
<text xml:id="untitled.text">
 <p xml:id="untitled.p.1">
  <t>Een onaangenaam mens in de Haarlemmerhout</t>
  <s xml:id="untitled.p.1.s.1">
   <t>Een onaangenaam mens in de Haarlemmerhout</t>
   <w xml:id="untitled.p.1.s.1.w.1" class="WORD">
    <t>Een</t>
    <pos class="LID(onbep,stan,agr)" confidence="0.981771" head="LID">
     <feat class="onbep" subset="lwtype"/>
     <feat class="stan" subset="naamval"/>
     <feat class="agr" subset="npagr"/>
    </pos>
    <morphology>
     <morpheme>
      <t offset="0">een</t>
     </morpheme>
    </morphology>
    <lemma class="een"/>
   </w>
   ....
```

GROUP

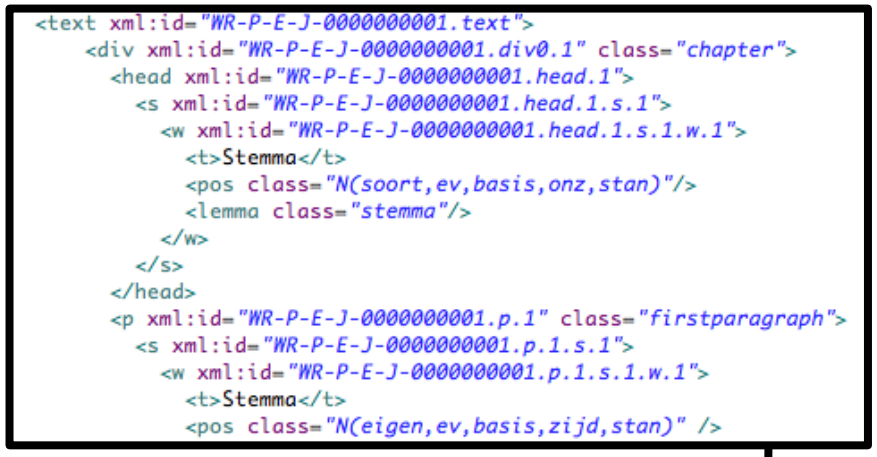
WORD

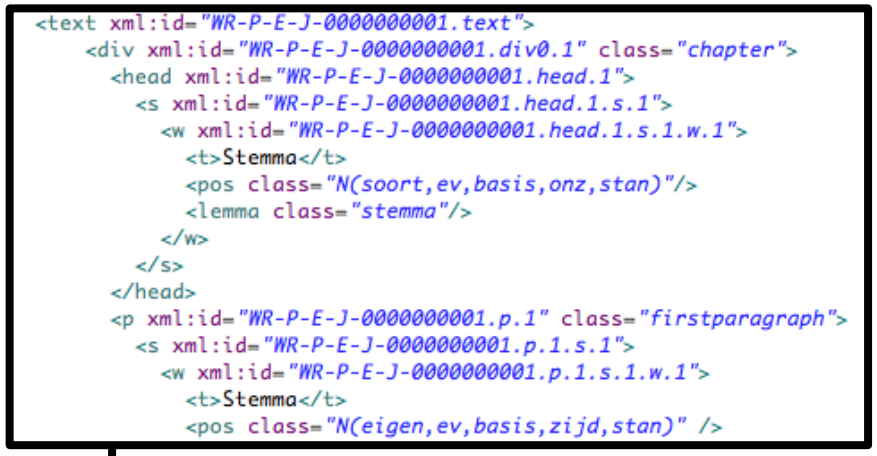
WORDANNOTATION

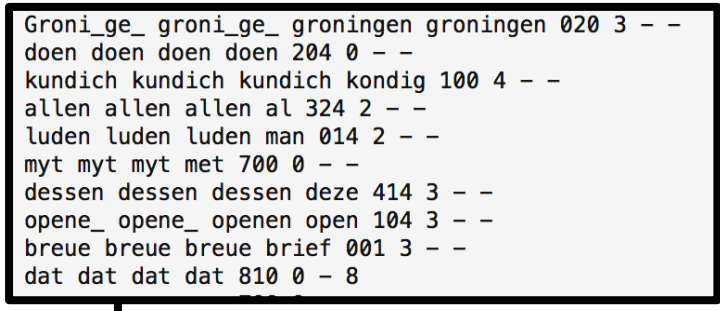

```
{"mtasId":0, "prefix":"t", "value":"Een", "position": 0, "parentMtasId":68},
{"mtasId":1, "prefix":"t_lc", "value":"een", "position": 0},
{"mtasId":2, "prefix":"feat.lwtype", "value":"onbep", "position": 0},
{"mtasId":3, "prefix":"feat.naamval", "value":"stan", "position": 0},
{"mtasId":4, "prefix":"feat.npagr", "value":"agr", "position": 0},
{"mtasId":5, "prefix":"pos", "value":"LID", "position": 0},
{"mtasId":6, "prefix":"morpheme", "value":"een", "position": 0},
{"mtasId":7, "prefix":"lemma", "value":"een", "position": 0},
{"mtasId":47, "prefix":"chunk", "value":"NP", "position": 0},
{"mtasId":54, "prefix":"dependency.dep", "value":"", "position": 0},
{"mtasId":55, "prefix":"dependency", "value":"det", "positions":"[0, 2]"},
{"mtasId":68, "prefix":"s",  "positions":  "0-5", "parentMtasId":69},
{"mtasId":69, "prefix":"p",  "positions":  "0-5"}
```


KWIC search for [#0]

everything on the first position







Configuration

Configuration

Configuration

FoLiA Parser

CRM Parser

QUERY

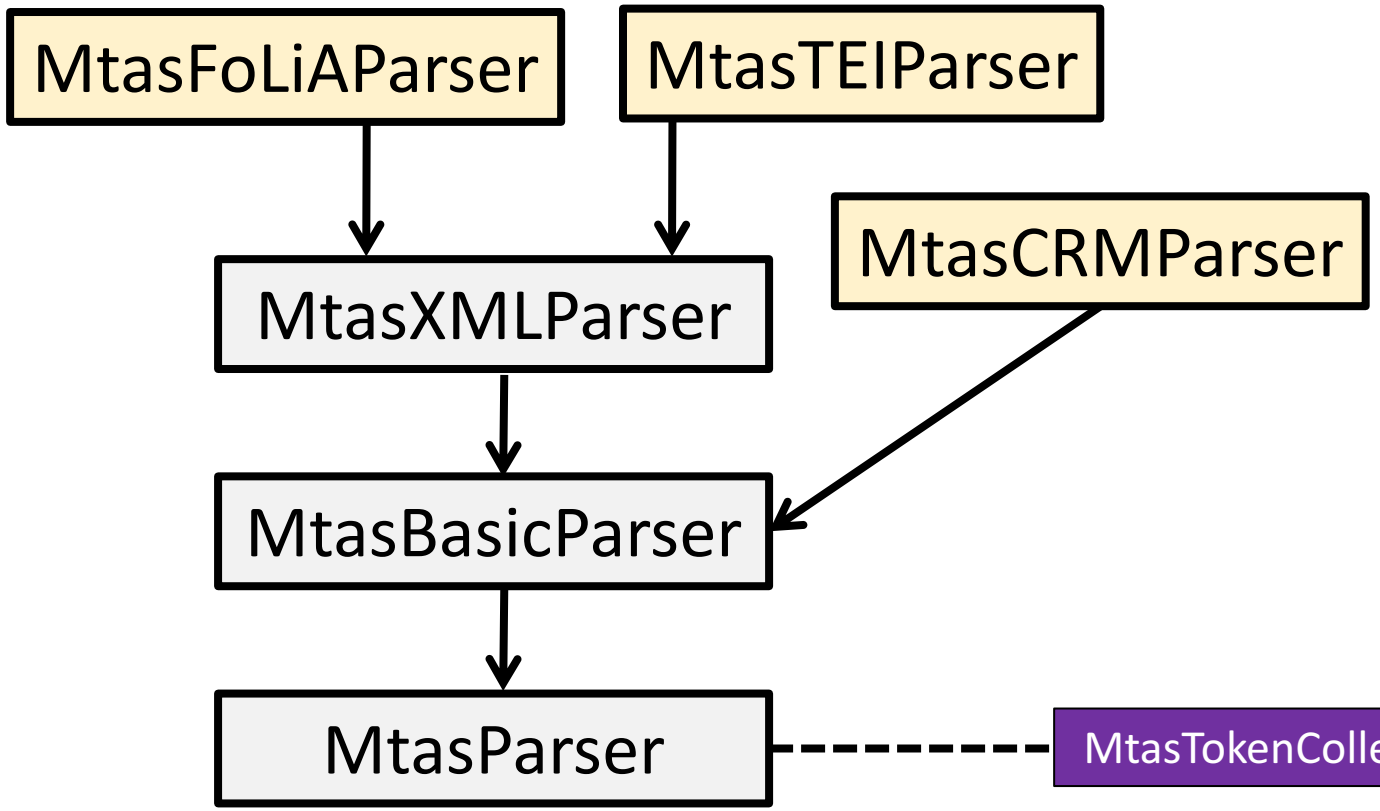
[pos="ADJ"] [lemma="Amsterdam"]

Solr

MTAS

RESPONSE

PARSERS




```java
package mtas.analysis.parser;

import java.io.Reader;

public class MtasCustomParser extends MtasParser {

  @Override
  public MtasTokenCollection createTokenCollection(Reader
      throws MtasParserException, MtasConfigException {
    // TODO Auto-generated method stub
    return null;
  }

  @Override
  public String printConfig() {
    // TODO Auto-generated method stub
    return null;
  }
```

Document	Field	TokenId	Prefix	Postfix	Position	ParentId	Payload
1	text	0	s	-	0 - 10	-	-
1	text	1	t	lk	0	0	-
1	text	2	pos	VNW	0	0	-
1	text	3	t	ben	1	0	-
…	…	…	…	…	…	…	…

FoLiA-server

f5ec0e6e-210e-4588-bc9f-b6d80d95aafe

NEDERLAB

Metadata

phpMyAdmin

SCRIPT

php

HPC Cloud


```
{
 "NLProfile_name": "nederlabTitleProfile",
 "NLCore_NLIdentification_nederlabID": "f5ec0e6e-210e-4588-bc9f-b6d80d95aafe",
 "NLCore_NLIdentification_versionID": "f5ec0e6e-210e-4588-bc9f-b6d80d95aafe",
 "NLCore_NLIdentification_sourceRef": ["WS-U-E-A-0000005860"],
 "NLCore_NLAdministrative_ingestTime": "2015-11-24T00:00:00Z",
 "NLCore_NLAdministrative_sourceCollection": "SoNaR",
 "NLCore_NLAdministrative_isThesaurusElement": true,
 "NLTitle_title": "koe",
 "NLTitle_genre": ["non-fictie"],
 "NLTitle_yearOfPublicationMin": 1999,
 ...
 "NLContent_mtas": "f5ec0e6e-210e-4588-bc9f-b6d80d95aafe"
}
```


MTAS

Solr

BROKER

Home **Search** Documentation Settings

Search

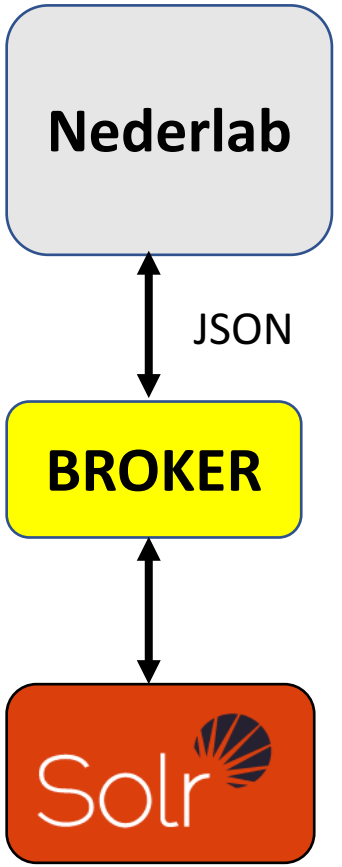

```
{
  "condition": {
    "type": "cql",
    "field": "NLContent_mtas",
    "value": "[lemma=\"ναῦς\"]"
  },
  "response": {
    "mtas": {
      "kwic": [
        {
          "field": "NLContent_mtas",
          "query": {
            "type": "cql",
            "value": "[lemma=\"ναῦς\"]"
          },
          "start": 0,
          "number": 2,
          "left": 2,
```


Reset | Align json | **Send request** | ☑ **Status** | Cache

Response

▼ Response


```
{
  "mtas": {
    "kwic": [
      {
        "key": "0",
        "list": [
          {
            "documentKey": "4e32b666-011c-11e4-b0ff-51bcbd7c379f",
            "documentTotal": 1,
            "documentMinPosition": 0,
            "documentMaxPosition": 3110,
            "list": [
              {
                "left": {
                  "2622": [
                    [
                      "t",
                      "ἡ"
                    ]
                  ],
                  "2623": [
                    [
                      "t",
                      "κοίλη"
                    ]
                  ]
                },
                "hit": {
                  "2624": [
                    [
                      "t",
                      "ναῦς"
                    ]
                  ]
```

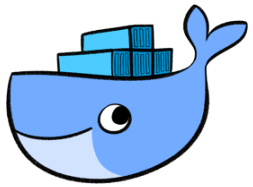



NEDERLAB

Source and documentation on GitHub

https://github.com/meertensinstituut/mtas

https://github.com/meertensinstituut/broker



Docker image with demo scenarios available

QUESTIONS ?