

GeoNetwork's CSW 2 Implementation

Andrea Carboni

FAO

Summary

- Used documents
- Current status of GeoNetwork's CSW implementation
- Implementation details
- Issues
- Implementation proposal

Used documents and current status

Used Documents

- OGC catalogue services specification 2.0.1 (04-021R3)
- ISO19115/19119 Application profile 0.9.3 (04-038R2)
- Filtering encoding specification 1.1.0 (04-095)
- OWS common implem. specification 0.3.0 (04-016R3)
- ISO19139 XML schema updated to March 2006 from <http://www.isotc211.org/schemas/2005>

Current status

- Implemented operations: GetCapabilities, DescribeRecord, GetRecordById, GetRecords
- Both filter and CQL expressions are handled (the CQL implementation is trivial)
- Almost all queriables fields are handled (from both the core specification and the ISO profile)

CQL Status

- Uses a simple free library from <http://zing.z3950.org/cql/java/>
- The library is pretty old and not completed (last version is 0.7 and the last update was on 4 Sep 2003)
- Supports only AND and OR clauses (even nested) and the following operators: equal, not equal, <, <=, >, >= (no spatial operators)
- Strings containing blanks cannot be searched (like 'basic africa')

What Is Missing

- DescribeRecord is not complete because it is not clear what it should return
- Advanced spatial operators on search (operators not handled by lucene)
- Distributed search
- A stylesheet to convert filter expressions to lucene expressions (work in progress)

Implementation details

Implementation Details

- Modularization
 - Server side : the real implementation
 - Client side : a GUI application for testing and debugging
 - Common library created for interoperability

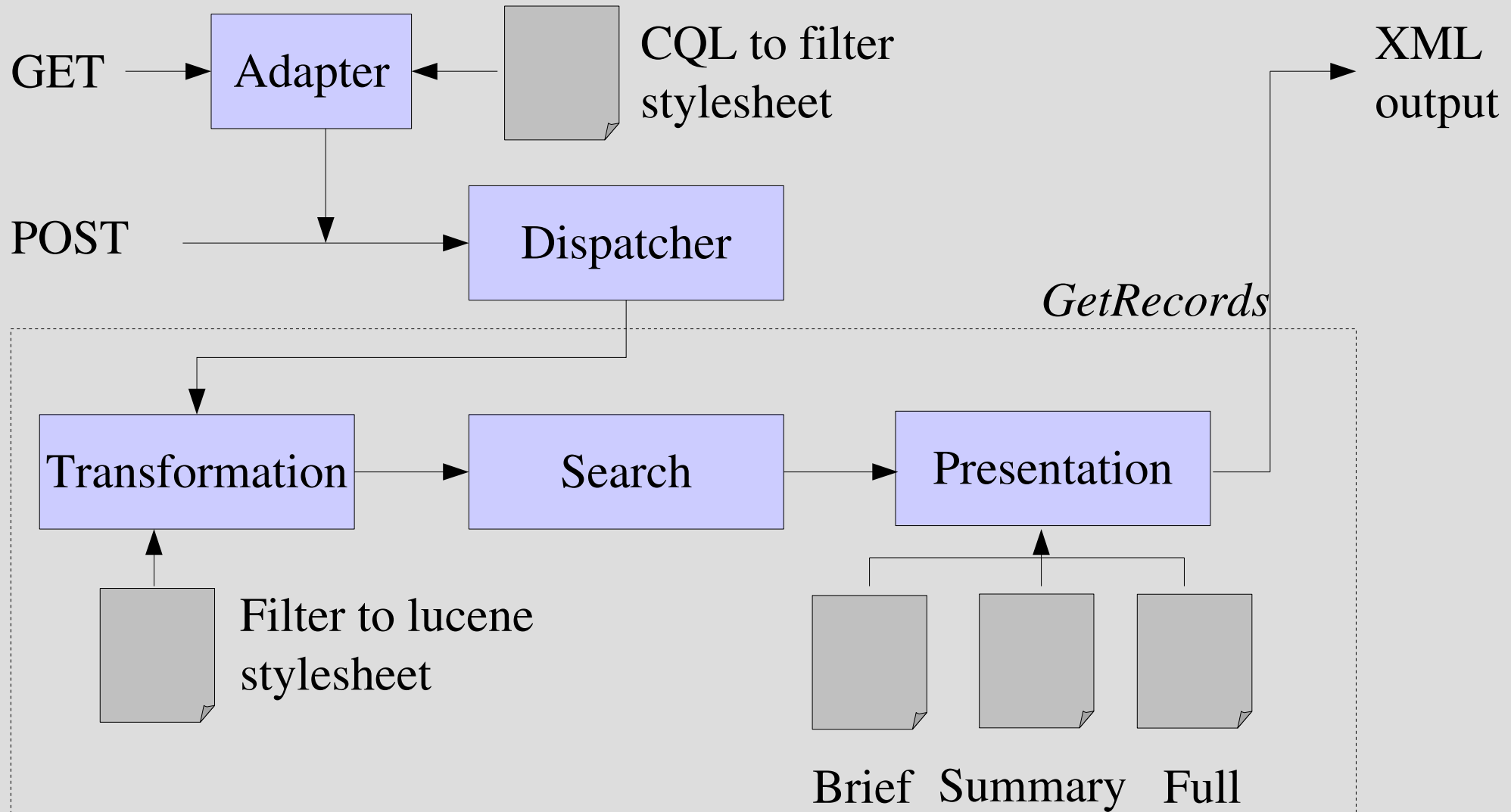
Server Side

- Uses XML over HTTP (no SOAP)
- Handles both GET and POST requests
- Handles both filter and CQL expressions
- Uses GeoNetwork's authentication

Search Engine

- Unified search management (a stylesheet converts CQL expressions into filter ones)
- A stylesheet convert filter expressions into lucene search trees
- Due to the use of Lucene, complex spatial queries cannot be performed.
- Handled spatial operators : Equal, Encloses, FullyEnclosedWithin, Overlaps, FullyOutsideOf

GetRecords Workflow



Common Library

- Simple and small communication layer provided for interoperability
- Exception tree mapped on CSW exception types
- Exception marshaling / unmarshaling between client and server (network transparency)
- XML over HTTP provided by communication classes
- High level classes to call server operations and get results

Client Side

- Standalone GUI application to test/debug services
- Authenticates on GeoNetwork
- Allows both GET and POST requests
- Allows to test : GetCapabilities, DescribeRecord, GetRecordById, GetRecords
- Almost all operation parameters can be specified
- GeoNetwork independent, can be used to test any CSW 2 server
- Debug facilities in case of error (full HTTP dump)

Application Screenshot

The screenshot displays the 'GeoNetwork CSW 2.0.1 test application' window. The interface is divided into several sections:

- Server:** Host: localhost, Port: 8080, CSW service: /geonetwork/srv/en/csw, Login service: /geonetwork/srv/en/xml.user.login
- Login:** Username: a, Password: aaaaaa
- Request:** Operation: GetRecords, Method: POST
- GetRecords parameters:** General tab selected. Result type: hits, Output schema: ogccore, Element set name: summary. Other fields (Output format, Start position, Max records, Sort by) are empty.
- Communication log:** Shows the request and response XML. The request is a POST to /geonetwork/srv/en/csw with headers including Host, Connection, Accept, Cache-Control, User-Agent, Cookie, Content-Type, and Content-Length. The response is an XML document with the following structure:

```
<?xml version="1.0" encoding="UTF-8"?>
<ok />

POST /geonetwork/srv/en/csw HTTP/1.1
Host: localhost:8080
Connection: close
Accept: application/xml
Accept-Charset: utf-8
Cache-Control: no-cache
User-Agent: GeoNetwork-csw-client/2.0.1
Cookie: JSESSIONID=171oh5xrqb91e;Path=/geonetwork
Content-Type: application/xml; charset=utf-8
Content-Length: 397

<?xml version="1.0" encoding="UTF-8"?>
<csw:GetRecords xmlns:csw="http://www.opengis.net/cat/csw" service="CSW"
  <csw:Query>
    <csw:ElementSetName>summary</csw:ElementSetName>
    <csw:Constraint version="1.1.0">
      <csw:CqlText>AnyText = 'africa%'</csw:CqlText>
    </csw:Constraint>
  </csw:Query>
</csw:GetRecords>

<?xml version="1.0" encoding="UTF-8"?>
<csw:GetRecordsResponse xmlns:csw="http://www.opengis.net/cat/csw">
  <csw:SearchStatus status="complete" />
  <csw:SearchResults numberOfRecordsMatched="1" numberOfRecordsReturned="1" />
</csw:GetRecordsResponse>
```

Issues and Implementation Proposal

Issues

- There are some anomalies because some drafts are not final yet. This leads to:
 - Communication problems with other implementations (due to the ISO profile)
 - Metadata inconsistencies (due to different ISO19139 schemas)
- Some standards are left out (FGDC, dublin core)

Implementation Proposal

- Use only one way to do things (HTTP or SOAP, GET or POST, CQL or filters but not both)
- HTTP POST with filters is better suited than HTTP GET with CQL
- Only synchronous communication (as proposed in the ISO profile). The only exception being harvesting.
- Provide metadata schema for some operations