

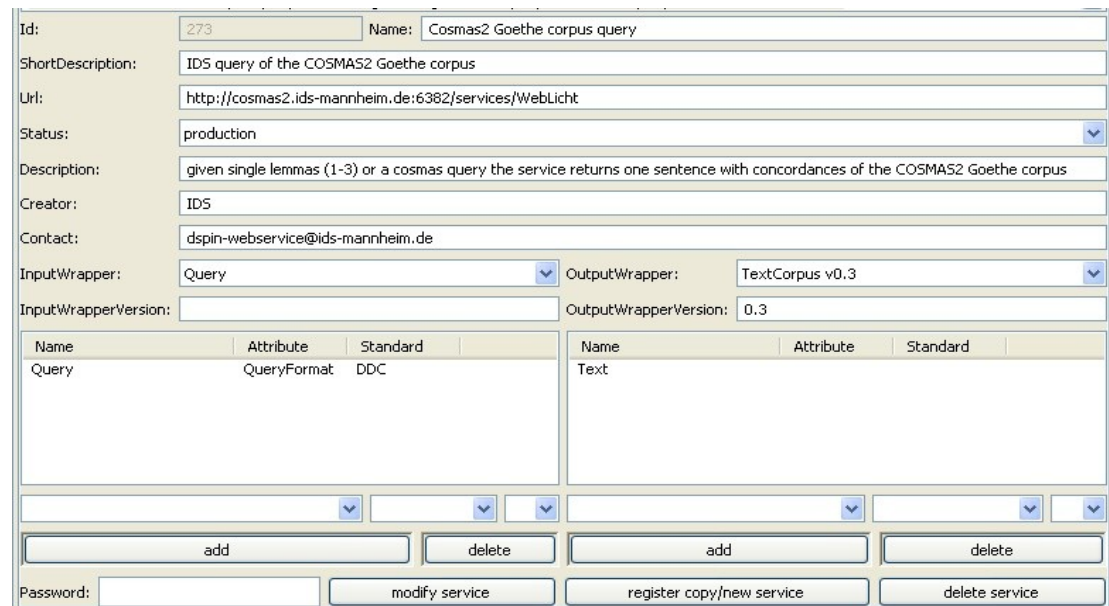
Integration of IDS Web Services

Jens Rinne and Oliver Schonefeld
Institute for the German Language (IDS), Mannheim

Web Services Workshop, Freudenstadt

Overview

1. COSMAS II
2. Technical aspects
3. Example Query
4. Issues



The screenshot displays a web service management interface. At the top, it shows the service ID (273) and name (Cosmas2 Goethe corpus query). Below this, there are fields for ShortDescription, Url, Status, Description, Creator, and Contact. The InputWrapper is set to 'Query' and the OutputWrapper is 'TextCorpus v0.3'. Below these fields are two tables. The first table has columns for Name, Attribute, and Standard, with one row: Name: Query, Attribute: QueryFormat, Standard: DDC. The second table has columns for Name, Attribute, and Standard, with one row: Name: Text. At the bottom, there are buttons for 'add', 'delete', 'modify service', 'register copy/new service', and 'delete service'. A password field is also visible at the bottom left.

Name	Attribute	Standard
Query	QueryFormat	DDC

Name	Attribute	Standard
Text		

COSMAS II

Service to connect to the COSMAS II of the IDS

- Corpus Search, Management and Analysis System
- COSMAS II enables its users to compose and analyze so-called virtual corpora
- COSMAS II provides complex search options (including lemmatization, proximity operators, search across sentence boundaries, logical operators, etc.)

Technical Aspects

Outside: POST Web Service

Inside: POST WS wrapped by POST WS

- Convert query to internal syntax
- Use a part of the COSMAS II query power

WebLicht  IDS

POST

<http://cosmas2.ids-mannheim.de:6382/services/WebLicht>

RequestBody: lesen /s0 schreiben

(RequestBody contains single lemma or a valid c2query)

IDS WS Intern Example Query

```
<?xml version="1.0" encoding="UTF-8"?>
<D-Spin xmlns="http://www.cosmas.ids-mannheim.de:6382/data" version="0.3">
  <Query xmlns:ids="http://www.cosmas.ids-mannheim.de:6382/data/querydata">
    <ids:archiv>W - Archiv der geschriebenen Sprache</ids:archiv>
    <ids:corpus>goe - Goethes Werke</ids:corpus>
    <ids:querystring>lesen</ids:querystring>
    <ids:makeKwic>true</ids:makeKwic>
    <ids:kwicpre>6</ids:kwicpre>
    <ids:kwicpost>6</ids:kwicpost>
    <ids:kwicmode>33024</ids:kwicmode>
    <ids:makeBeleg>true</ids:makeBeleg>
    <ids:belegpre>0</ids:belegpre>
    <ids:belegpost>0</ids:belegpost>
    <ids:belegmode>74289</ids:belegmode>
    <ids:endIndex>1</ids:endIndex>
    <ids:startIndex>0</ids:startIndex>
    <ids:cutValue>0</ids:cutValue>
    <ids:wordCase>2048</ids:wordCase>
    <ids:opts>3145984</ids:opts>
    <ids:listSortOrder>1</ids:listSortOrder>
    <ids:makeWordForms>true</ids:makeWordForms>
    <ids:makeRandom>true</ids:makeRandom>
    <>...</>
  </Query>
</D-Spin>
```

Issues 1

- Use the query power of COSMAS II
- Enable user to specify request parameter
 - Individual query tags

```
<Query>  
  <ids:archiv>DEREKO</ids:archiv>  
  <ids:corpus>Own Korpus</ids:corpus>  
  <ids:optionRandomSort>>true</ids:optionRandomSort>  
  <>...</>  
</Query>
```

Issues 2

- Authentication (and Authorization)
 - Currently, only public parts of DeReKo are available via Web Service
 - Due to license contracts users have to sign an EULA to gain access to more resources
 - Need for Authentication
- Currently (no production-ready) solution for Shibboleth/(REST-)Web Services
 - Possible approaches: Security Token Service, Short-Lived-Credentials