



Processing dialect data using an XML database



The Data

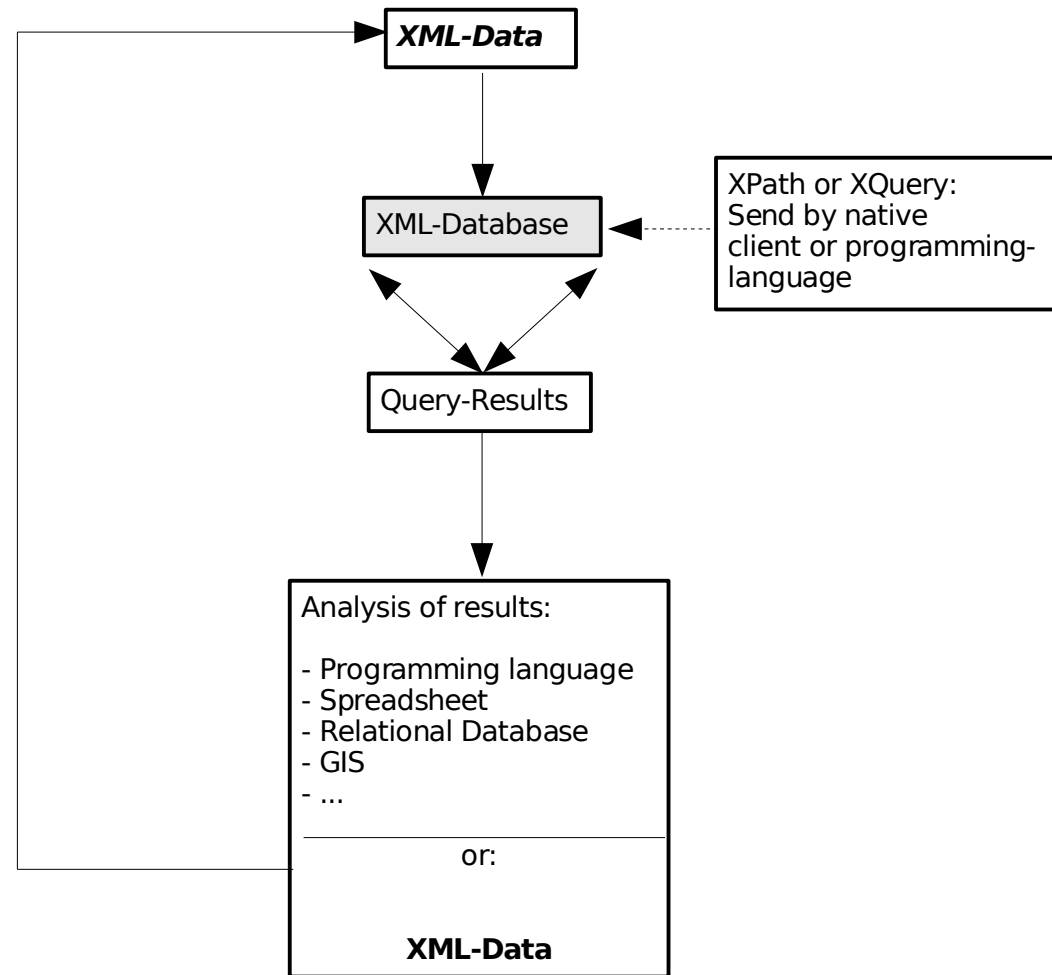
- ◆ **Ultimately, we will have 200 sites**
- ◆ **At the moment: 66 + 14 without coordinates**
- ◆ **Annotated in XML, using the CLaRK-system**
- ◆ **Ca. 155 words per site**

In Tübingen:

- ◆ **Integration into the XML database eXist**
- ◆ **Generation of wordlists in L04-format via XSLT**
- ◆ **Querying the database with the native client and from Java**



The XML Workflow





The eXist-database

- ◆ **OpenSource, written in Java**
- ◆ **Runs on Windows, Linux, ...**
- ◆ **A lot of interfaces: XML-RPC, SOAP, HTTP, ...**
- ◆ **Programing languages: Java, Perl, PHP, ...**
- ◆ **Integrated webserver**
- ◆ **Native client**

- ◆ **XPath and XQuery for querying the database**
- ◆ **XSL transformations included**
- ◆ **XML validation against DTDs**

- ◆ **In Tübingen:**
- ◆ **Most programs are written in Java**
- ◆ **XML-RPC is used as interface from within the Java programs**
- ◆ **Based on Java 5, developed in NetBeans IDE**



XP / XQUERY

- ◆ **XPath: Getting all sampa/variant-tags which contain a blank**

```
collection("/db/nbd")//sampa/variant[contains(., " ")]
```

- ◆ **XQuery: Building a list of all entries which have an english-tag "lamb"**

```
let $nbd := collection("/db/nbd")
```

```
for $entry in $nbd//entry where  
    string($entry/english)="lamb"
```

```
return <site>{$entry/../name}{$entry/../num}{$entry}</site>
```



The eXist database

The screenshot displays the eXist Admin Client interface. It includes a file browser window showing a list of files and folders, a query input window with a query history, and a terminal window showing the execution of a query.

Permissions	Owner	Group	Resource	Date
rwur-ur-u	admin	dba	PODiaD.dtd	Tue Dec 20 ...
rwur-ur-u	admin	dba	bell.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	beryaha.dat...	Tue Dec 20 ...
rwur-ur-u	admin	dba	byala.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	bychva.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	chasha.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	cheresha.dat...	Tue Dec 20 ...
rwur-ur-u	admin	dba	dadoh.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	den.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	djob.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	dobyr.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	doshyl.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	dva.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	ezik.dat.xml	Tue Dec 20 ...
rwur-ur-u	admin	dba	greshka.dat	Tue Dec 20 ...

```
type help or ? for help.
exist:/db> cd "buldialect"
exist:/db/buldialect> cd "data_petya"
exist:/db/buldialect/data_petya> cd "54-words"
exist:/db/buldialect/data_petya/54-words>
```

Found 19 items. Compilation: 3ms, Execution: 2503ms



The EXIST DATABASE

XQuery Sandbox - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost:8080/exist/sandbox/sandbox.xql

PT Google praeganz.de GIS wortlisten Project A1: Courses Trinity Rescue Kit... worldKit :: easy we...

Toms eMail-Dienst XQuery Sandbox

XQuery Sandbox Home Download Wiki Demo

Paste saved query: Maximize

document("/db/nbd/bansko.txt.xml")//site

Send Clear Check Display: 20 More Options

Found 1 in 0.027 seconds. Showing Items 1 to 1

```
1 <site>bansko, mzl
  <num>4062</num>
  <name>bansko</name>
  <entry>
    <key>arne</key>
    <english>bamb</english>
    <clom ana="Ncnst">jarne</clom>
    <nlom>agne</nlom>
    <variant ana="Ncnst">jagne</variant>
    <sampa>
      <nlom>agne</nlom>
      <variant ana="Ncnst">jagne</variant>
    </sampa>
  </entry>
  <key>aa</key>
  <english>lc</english>
  <clom ana="Ppe-ost">aa</clom>
  <nlom>az</nlom>
  <variant ana="Ppe-ost">ja</variant>
  <sampa>
    <nlom>az</nlom>
    <variant ana="Ppe-ost">ja</variant>
  </sampa>
  </entry>
  <key>Gene</key>
  <english>white plural</english>
  <clom ana="Apt">Gene</clom>
  <nlom>belic</nlom>
  <variant ana="Apt">b3lic</variant>
  <sampa>
    <nlom>belic</nlom>
    <variant ana="Apt">bElic</variant>
  </sampa>
```

Done



References

- ◆ **eXist XML database:** <http://exist.sourceforge.net>
- ◆ **XQuery standard:** <http://www.w3.org/XML/Query>
- ◆ **XQuery tutorial:** <http://www.w3schools.com/xquery/default.asp>