



**X****M****L**

# Data Management

dott. Chris Mair  
2009/2010 @unibz

I – Apache Xerces

last modified 2009-10-12

# Apache Xerces

- Xerces is a library for parsing, validating and manipulating XML documents
- it is available in C++ or Java under the Apache Software License (Free Software & Open Source)
- it has handy command line sample apps
- get the Java version from  
<http://xerces.apache.org/xerces2-j>

# Xerces command line

- use a shell script `xerces` to run the xerces sample apps:

```
#!/bin/sh
ARGS=$@
INSTALL_DIR=/opt/xerces/xerces-2_9_1/
CLASSPATH=.
for JAR in $INSTALL_DIR/*.jar; do
    CLASSPATH=$CLASSPATH:$JAR
done
export CLASSPATH
java $ARGS
```

# Run Xerces

- run the `dom.Counter` to check for well-formedness:

```
$ xerces dom.Counter telegram.xml  
[Fatal Error] telegram.xml:5:26: The element type  
"frm" must be terminated by the matching end-tag "</frm>".
```

- add the `-v` switch to validate against a DTD (we didn't formally see them yet):

```
$ xerces dom.Counter -v telegram.xml  
[Error] telegram.xml:5:8: Element type "frm" must be  
declared.  
[Fatal Error] telegram.xml:5:26: The element type  
"frm" must be terminated by the matching end-tag "</frm>".
```

# Experiments

- see “Learning XML - Chapter 2” and lecture notes I
- create your own XML documents
- check well-formedness using Xerces
- things to try: empty and container elements; entity declarations and references; comments and more ;)