

Validating XML in

```
<SPEECH>
<SPEAKER>HAMLET</SPEAKER>
<LINE>Rest, rest, perturbed spirit!</LINE>
<STAGEDIR>They enter</STAGEDIR>
<LINE>So, gentlemen,
<LINE>I will do open war with you.
<LINE>And let so vile a man
<LINE>to his love and to you,
<LINE>I shall not look to be together;
<LINE>I will not part with you.
<LINE>The time is out of joint: O cursed spite,
<LINE>That ever I was born to set it right!
<LINE>Nay, come, let's go together.
</SPEECH>
```

exist

Dannes Wessels
“DiZzzZz”





Overview

- Currently implemented
- How does Xerces validate?
- Two resolve routes
- The goal
- Catalog
- The plan
- To think of



Currently implemented

- Validation while inserting documents.
 - webDav
 - REST
 - XMLRPC
 - Interactive client
- Validation of stored documents
 - xQuery function `validation:validate()`

“Insert”

Two steps:

- Register grammar files in **WEB-INF/catalog.xml**
- Store grammar files in **WEB-INF/entities/**

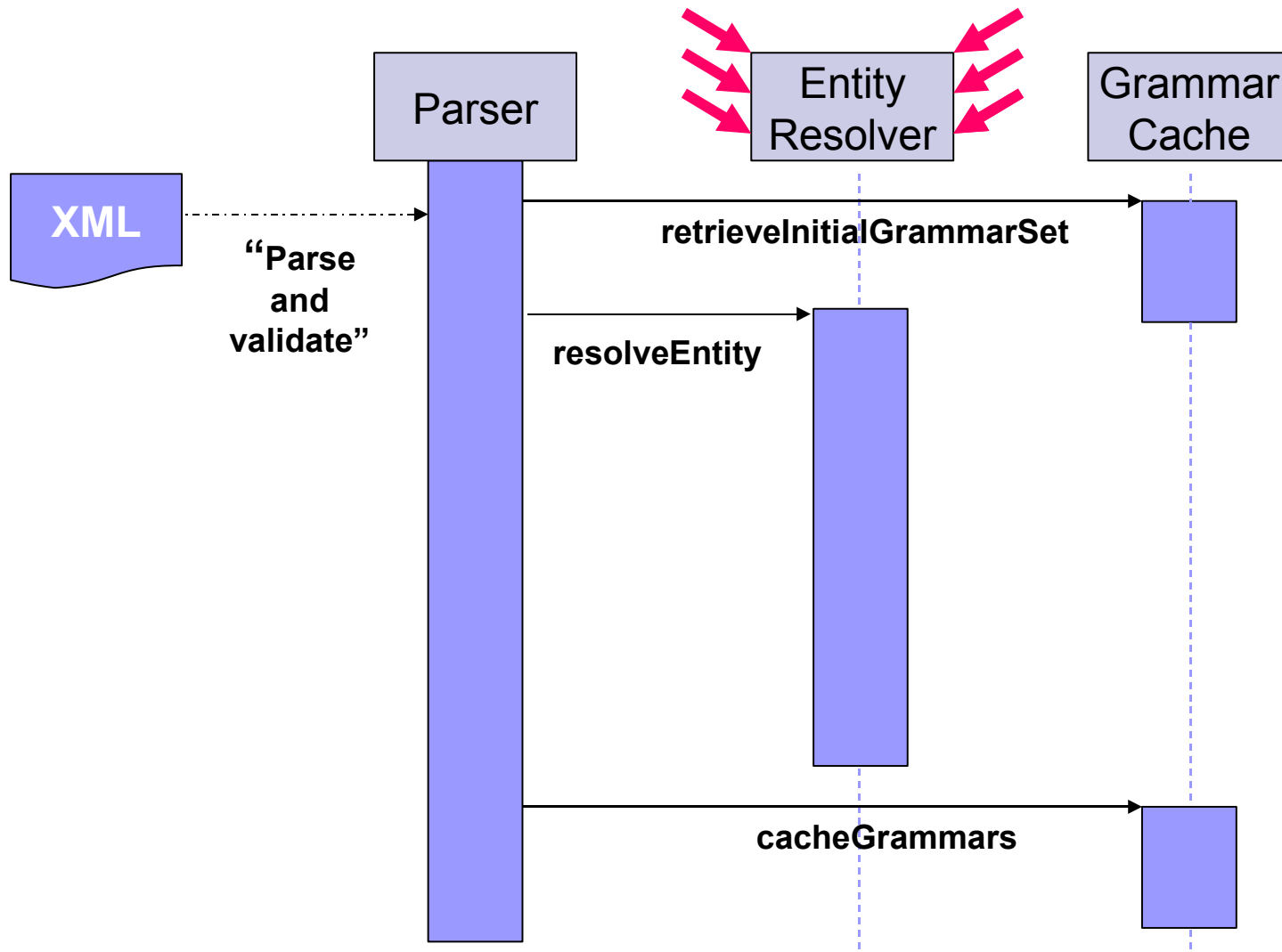
But...

- Grammar files and catalog file cannot be stored in database

“Function”

- InteractiveClient : ‘validate’ command
 - validate *<xmlresource>*
- xQuery : *http://exist-db.org/xquery/validation*
 - validate(*xml* , *grammar*)
 - validate(*xml* , *catalog.xml*)
 - validate(*xml* , *collection*) / validate (*xml*) (*search*)

How does Xerces validate?





Two entity resolve routes....

- “Insert”

- Use of apache.org resolver

- Old, not maintained
 - Ugly coded, spaghetti
 - URL/file horror

- “Function”

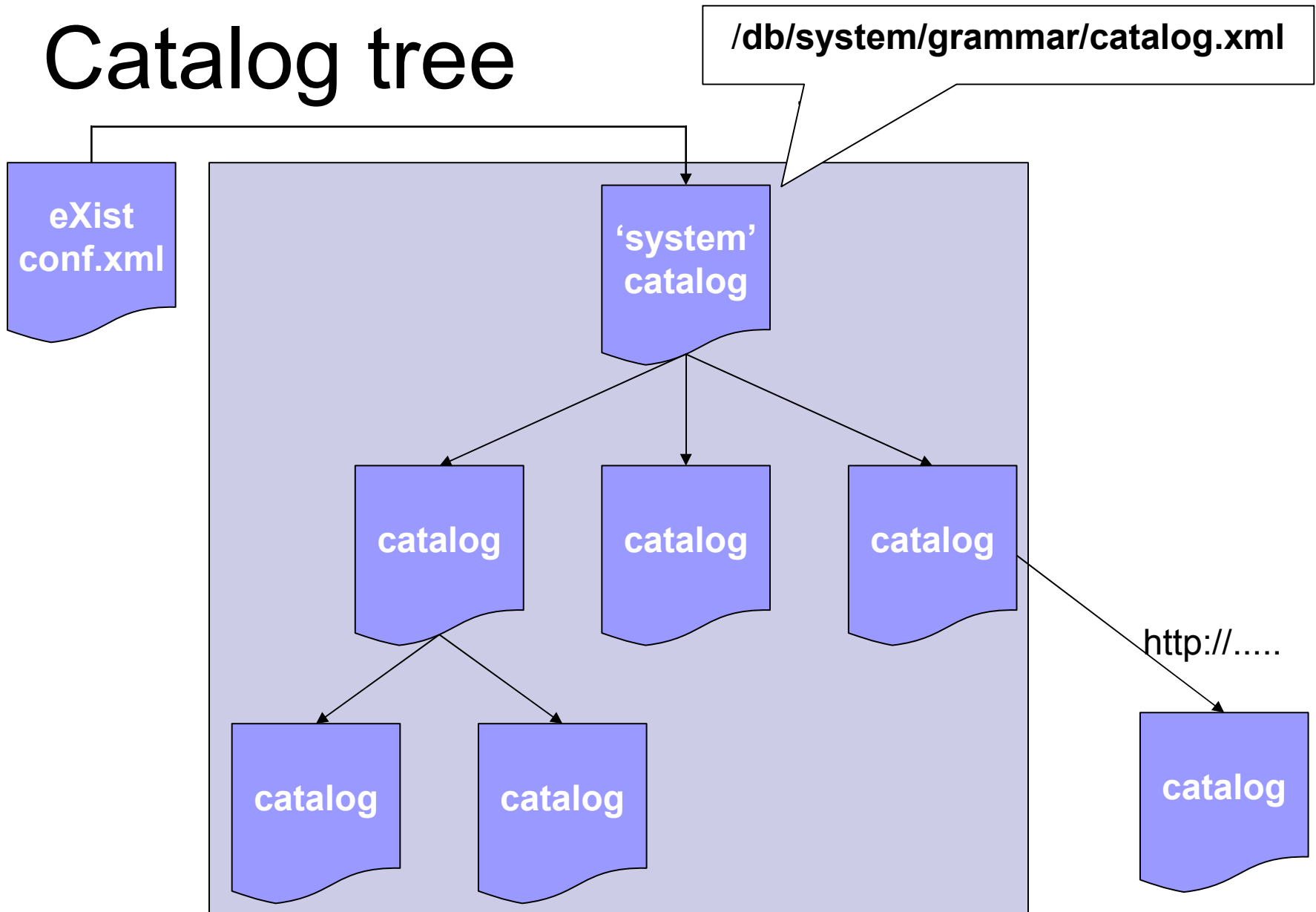
- Use of eXist specific entity resolver

- ‘Simple’ code
 - Access to resources in Database

The goal

- Bring routes together !
 - One entity resolver (eXist's ☺)
 - Automatic search in database!
- Ideas.....
 - Use features of OASIS catalog.xml specs to create 'tree of grammars'
 - One system 'grammar repository'
/db/system/grammar
 - users can store grammar 'everywhere' in database

Catalog tree



Catalog basics

```
<catalog xmlns="urn:oasis:names:tc:entity:xmlns:xml:catalog">
```

```
  <public publicId="-//OASIS//DTD XML DocBook V4.1.2//EN"  
          uri="/db/grammar/docbook.dtd"/>
```

```
  <uri name="http://www.oasis-open.org/committees/docbook/"  
       uri="/db/grammar/docbook.dtd /"/>
```

```
</catalog>
```

- DTD
 - public, system
- XSD
 - uri
- Catalog
 - nextCatalog



Catalog specials

- DTD

- delegatePublic, delegateSystem, rewriteSystem

- XSD

- delegateUri, rewriteUri



The plan...

- Finish the entity resolver code

The plan... (2)

- Improve Grammar caching strategy
 - See what is cached
 - Remove from cache
 - Per User cache of grammar set
- Generate catalog files by parsing DTDs
- Implement Schematron support



To think of...

- How to support w3c's 'validation' function?
- How use schema information in eXist to e.g. optimize Indexes?

To think of... too

- Apache's resolver is used by Cocoon too...
 - Eliminate validation function from cocoon?
 - Ignore?
 - Let cocoon use eXist's resolver?
 - **Need some investigation here...**