xlinkit: A Consistency Checking and Smart Link Generation Service

Anthony Finkelstein & Christian Nentwich
University College London
Software Systems Engineering Group
Department of Computer Science
{a.finkelstein|c.nentwich}@cs.ucl.ac.uk
http://www.cs.ucl.ac.uk/staff/{A.Finkelstein|C.Nentwich}
Also: http://www.xlinkit.com

UCL

With a Little Help from My Friends

… principally
Wolfgang Emmerich

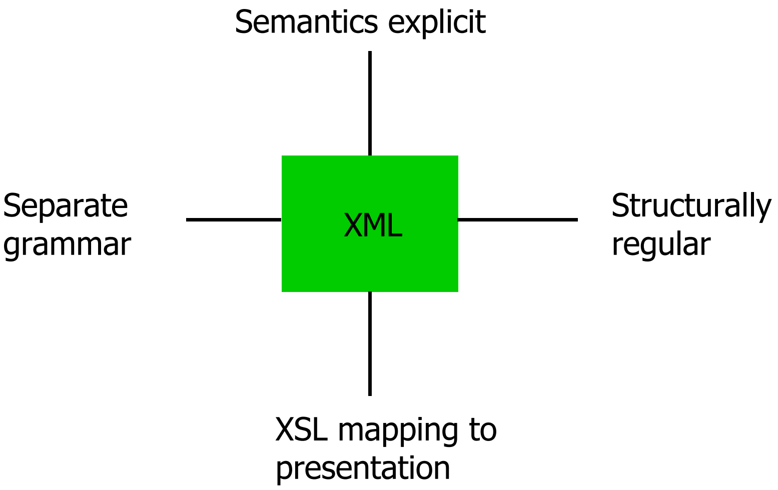
… and also
Licia Capra
Ernst Ellmer
Torbjorn Revheim
Danila Smolko
Andrea Zisman
Giulio Carlone

Outline

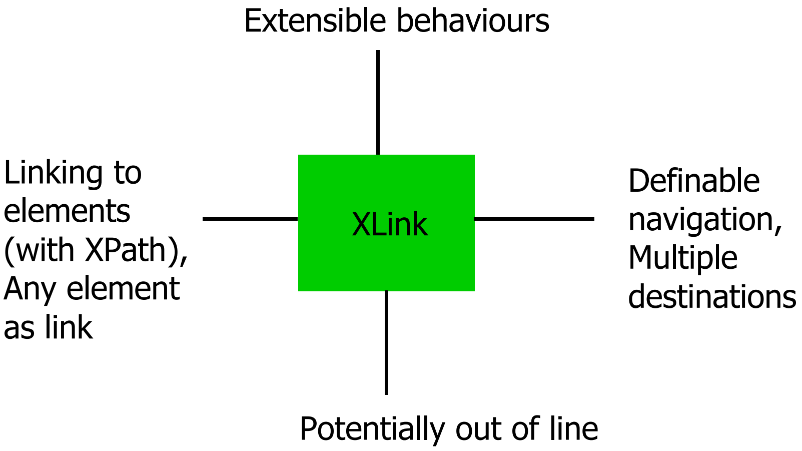
- Context & Background
- Example
- Rule Language
- Link Generation
- Content Management
- Architecture
- Applications
- Evaluation
- Scaleability
- Future Work

Mostly a demo!

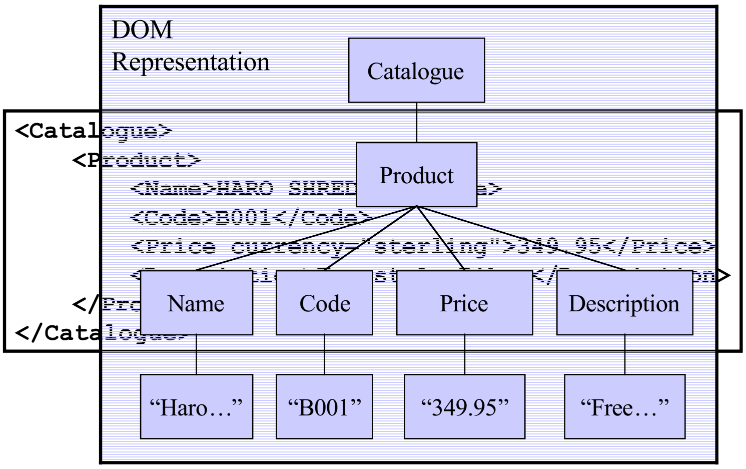
Markup



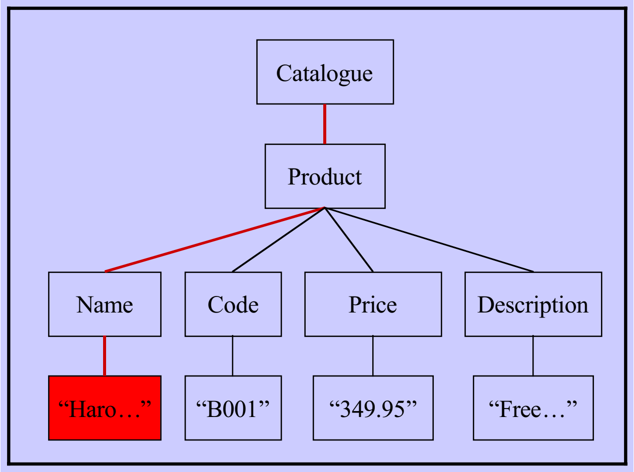
Linking


Extensible behaviours
Linking to elements (with XPath), Any element as link
XLink
Definable navigation, Multiple destinations
Potentially out of line

XML / DOM Representation



XPath



Catalogue

Product

Name

Code

Price

Description

“Haro…”

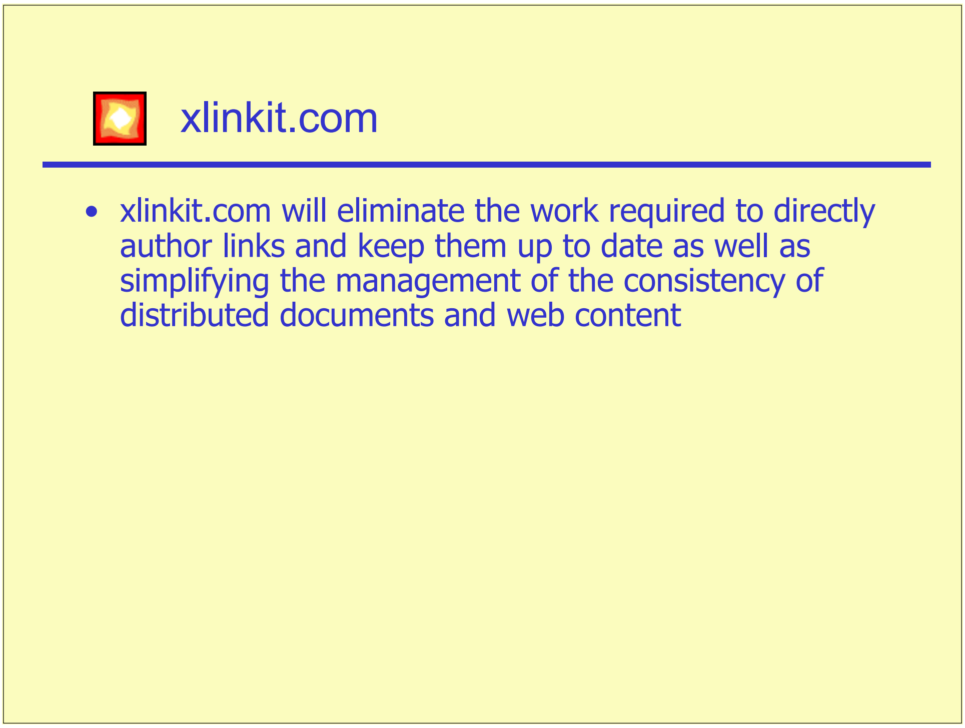
“B001”

“349.95”

“Free…”

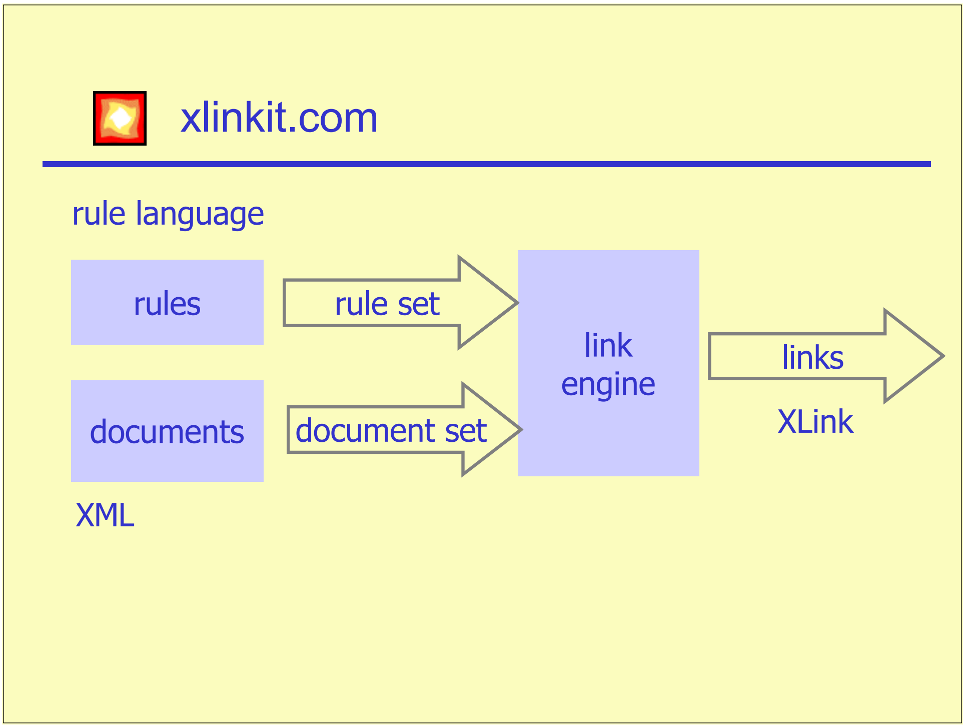
xlinkit.com

- xlinkit.com is a lightweight application service which provides rule-based link generation and checks the consistency of distributed documents and web content
- You tell xlinkit.com the information you want to link and rules that relate the information. xlinkit.com will generate the links that you can then use for navigation. It will also diagnose inconsistent information and, if you want, provide you links directly to the inconsistent items of information



xlinkit.com

- xlinkit.com will eliminate the work required to directly author links and keep them up to date as well as simplifying the management of the consistency of distributed documents and web content



xlinkit.com

rule language

rules

rule set

documents

document set

link engine

links

XLink

XML

xlinkit.com

And now a demo …

Rule language

- Our rules are given in a simple first order logic language with the following restrictions:
  - the sets we are working on are sets of DOM nodes and are always finite,
  - the only predicates allowed are equality and inequality, and no functions are allowed.

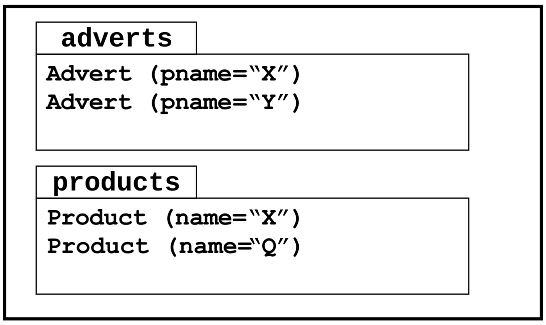
```
∀ a in adverts
  (∃ p in products (
    $a/AdvertName/text()=
    $p/CatalogueName/text()))
```

XML Encoding

```
<consistencyruleset>
   <globalset id="adverts" xpath="/Advert"/>
   <globalset id="products" xpath="/Catalogue/Product"/>
   <consistencyrule id="r1">
       <description>
              The product name of an advertised product
              must be in the catalogue
       </description>
       <forall var="a" in="$adverts">
              <exists var="p" in="$products">
                     <equal op1="$a/ProductName/text()"
                            op2="$p/Name/text()"/>
              </exists>
       </forall>
   </consistencyrule>
</consistencyruleset>
```

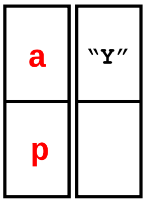
Example

- Identify sets to check
  - set of *Advert* elements, set of *Product* elements
- Let `adverts` be the set */Advert*
- Let `products` be the set */Catalogue/Product*



How it Works ...

```
<forall var="a" in="$adverts">
        <exists var="p" in="$products">
                <equal op1="$a/ProductName/text()"
                        op2="$p/Name/text()"/>
        </exists>
</forall>
```



a

"Y"

p

"Y" * {} =>

{(Consistent, {"X","X"}),(Inconsistent, {"Y"})}

Content Management

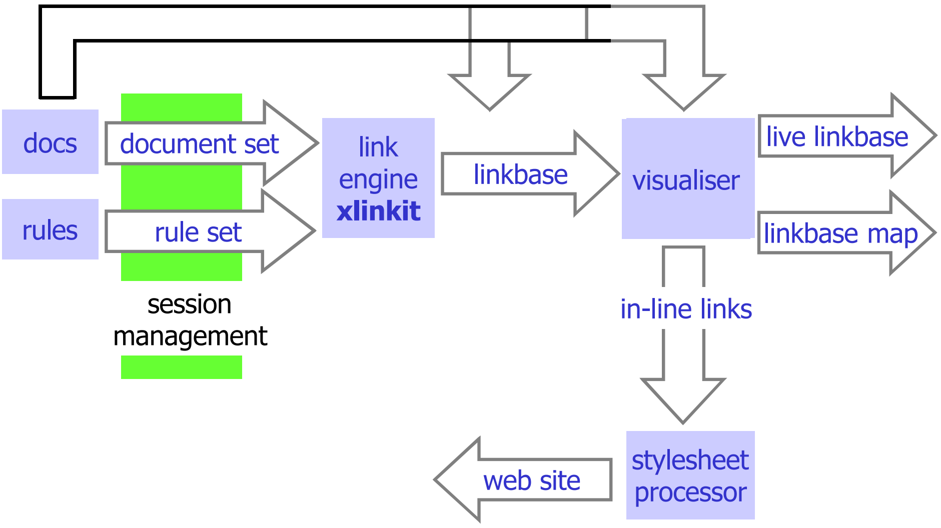
- Rule Sets
- Document Sets

```
<DocumentSet name="BikeDoc">
  <Description>Bike shop documents</Description>
  <DocFile href="catalogue.xml"/>
  <Set href="Adverts.xml"/>
  <Set href="Customers.xml"/>
  <Set href="Services.xml"/>
</DocumentSet>
```

- Retrieval of documents from databases

```
<DocFile fetcher="JDBCFetcher"
href="jdbc:mysql://www.xlinkit.com/wilburs
#select * from report"/>
```

How to Use Linkbases



docs

rules

document set

rule set

session management

link engine **xlinkit**

linkbase

visualiser

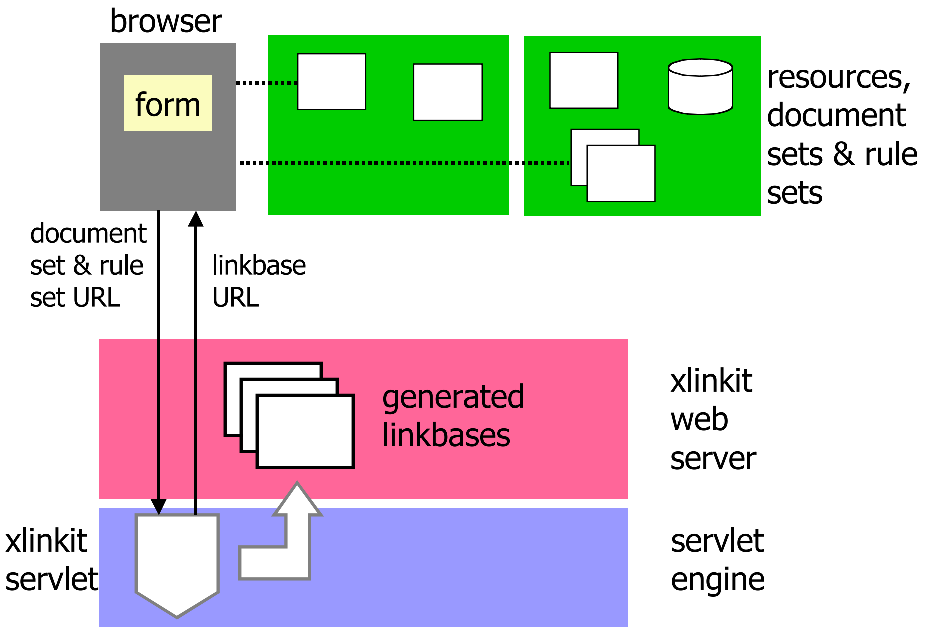
live linkbase

linkbase map

in-line links

stylesheet processor

web site



browser

form

resources, document sets & rule sets

document set & rule set URL

linkbase URL

generated linkbases

xlinkit web server

xlinkit servlet

servlet engine

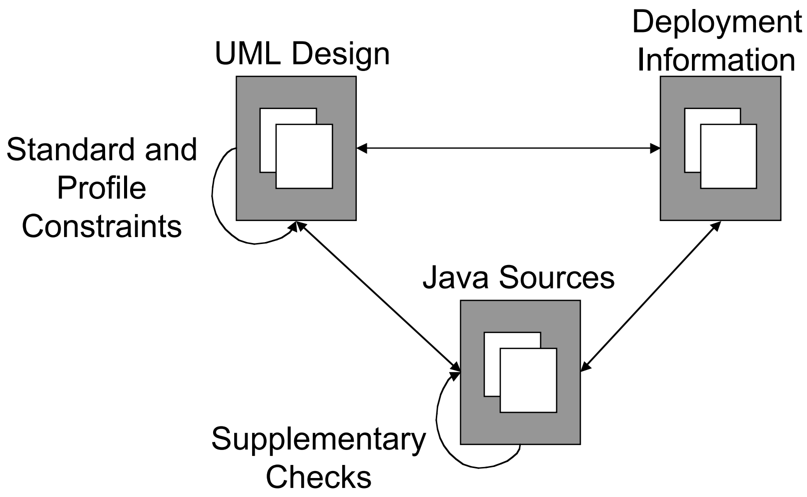
Applications

- Enterprise data integration
- Lightweight portals
- Value added content aggregation

Finance, Pharmaceuticals, Engineering Product Data Management, Customer Relationship Management, Network Policy …

And of course Software Engineering!

A ‘Cool’ Example



UML Design

Deployment Information

Standard and Profile Constraints

Java Sources

Supplementary Checks

Evaluation

- UML models in XMI, small model with 93 model elements, medium sized model with 610 model elements. 19 related industrial models ranging from 64 to 2834 elements - size from 100 kilobytes to 6 megabytes.
- Checked against the UML Core Constraints (34 rules)
- Total check time = 2.6 minutes (>20 megabytes data) on 750 Mhz Pentium and using IBM JDK 1.2
- No check longer than 2.38 minutes, most of the time on evaluating XPath expressions. 8101 inconsistent links!

Extensions

- User-defined predicates
  - Examples: "fuzzy" matches; sub-tree matching; thesaurus-based matching etc.

    Matching business day conventions in FpML
- Write your own predicate in JavaScript
- Define type signature and implementation URL for the predicates in an OperatorSet
- Reference the OperatorSet in the RuleSet

Scaleability

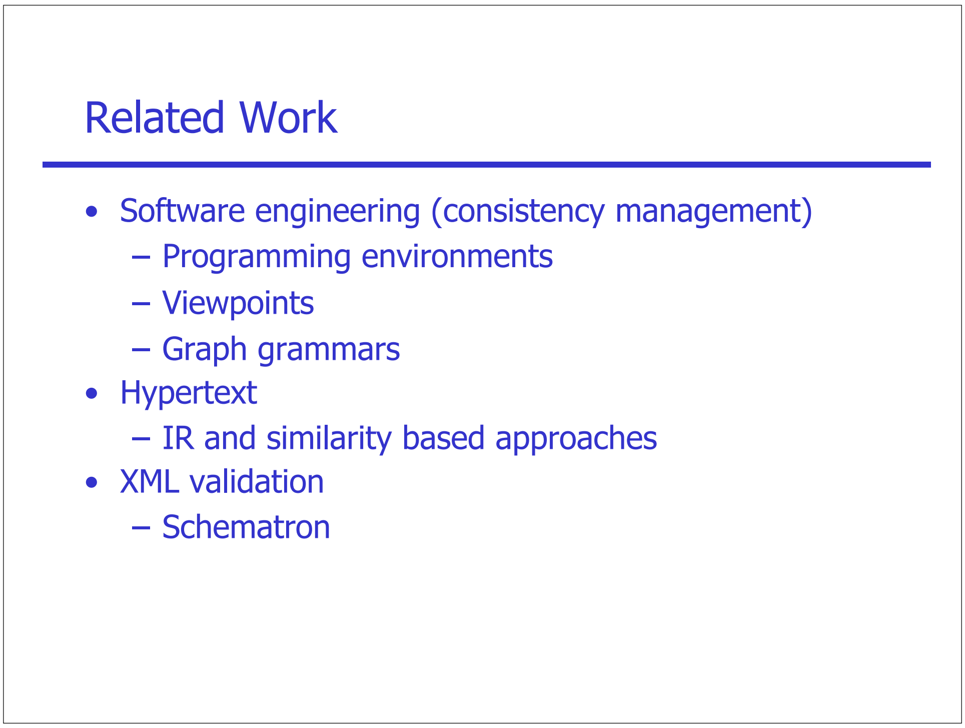
- Incremental checking
  - Assumption: documents change frequently, rules do not
  - Rules are defined for document types, relatively stable
  - Establish global consistency status once, minimise amount of re-computation as documents change

Identify changes to documents in document set
Determine which rules *intersect* with changes
Update consistency status

➡ static analysis - fast!

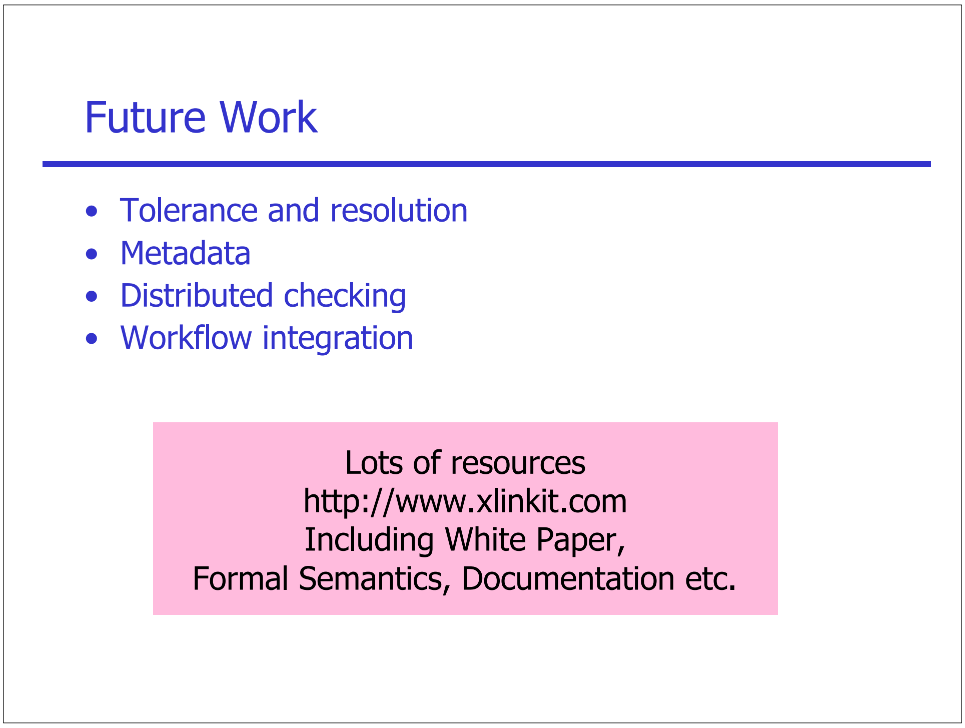
Scaleability

- Memory management
  - For very large collections of documents
  - Special-purpose Fetcher
  - Uses a Persistent DOM (Infonyte) with an XPath interface
  - Can now check documents which do not fit into memory



Related Work

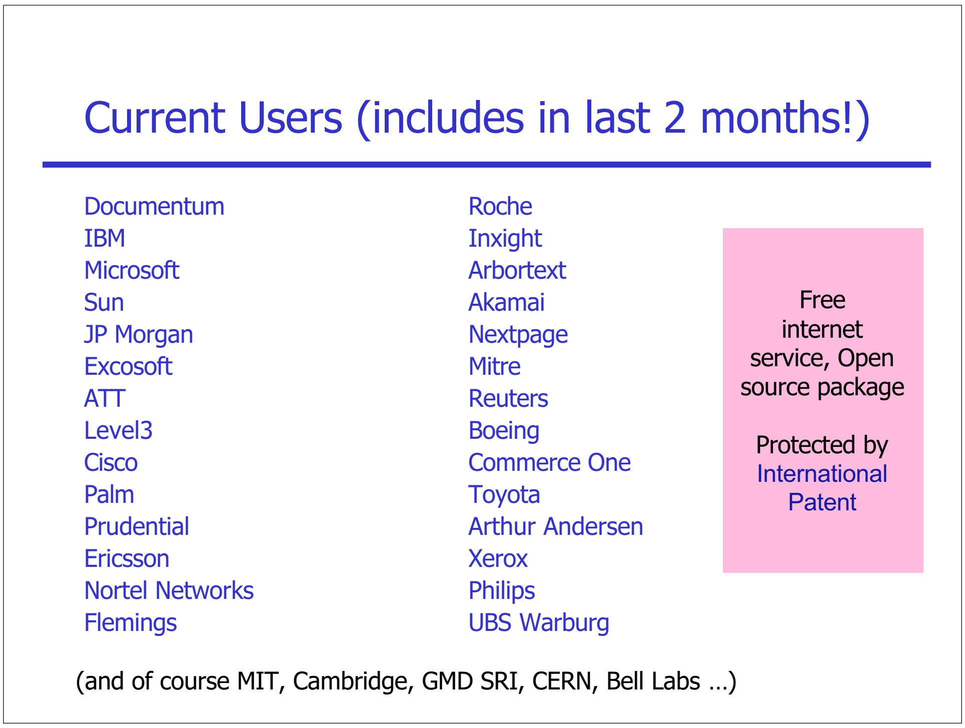
- Software engineering (consistency management)
  - Programming environments
  - Viewpoints
  - Graph grammars
- Hypertext
  - IR and similarity based approaches
- XML validation
  - Schematron



Future Work

- Tolerance and resolution
- Metadata
- Distributed checking
- Workflow integration

Lots of resources
http://www.xlinkit.com
Including White Paper,
Formal Semantics, Documentation etc.



Current Users (includes in last 2 months!)

Documentum
IBM
Microsoft
Sun
JP Morgan
Excosoft
ATT
Level3
Cisco
Palm
Prudential
Ericsson
Nortel Networks
Flemings

Roche
Inxight
Arbortext
Akamai
Nextpage
Mitre
Reuters
Boeing
Commerce One
Toyota
Arthur Andersen
Xerox
Philips
UBS Warburg

Free internet service, Open source package

Protected by International Patent

(and of course MIT, Cambridge, GMD SRI, CERN, Bell Labs …)