Towards the Semantic Web and Web Services

XML Finland 2002

Slide Presentations

Eero Hyvönen and Mika Klemettinen (editors)

October 21-22, 2002
Marina Congress Center
Helsinki, Finland

XML Finland Association

XML Finland Association is a non-profit association with the purpose of providing knowledge about XML and related standards.

For more information about the association and for joining as a member, see http://www.xml-finland.org/.

XML Finland 2002 - Towards the Semantic Web and Web Services

October 21-22, 2002
Marina Congress Center, Helsinki

XML Finland 2002 - Exhibitors

INDEX INFORMATION TECHNOLOGIES
X-HIVE

INNOXMLp
XML Parsing to the Limits

Microsoft®

republica

XML Finland 2002 - Partners

DONE
www.donesolutions.com

eCraft
Management Solutions

genio innovations

ieturi

XML Finland 2002 - Supporting Patrons

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

OntoWeb

TIEKE

HELSINGIN YLIOPISTO

Sun microsystems

Foreword

This volume contains the slide and other presentations not included into the research papers proceedings of the XML Finland 2002 conference in Helsinki, Finland, October 21-22, 2002 – the seventh annual conference organized by the XML Finland Association. The themes of the year 2002, the Semantic Web and Web Services, attracted some 200 participants.

The Semantic Web is a vision of the W3C consortium about the next generation Web, which is used not only by the humans but also by the machines. The vision is brought alive with XML-based "semantic" standards and interfaces that make web contents understandable to the machines. This enables the creation of more intelligent Web Services than before.

After the previous XML Finland conference in 2001, the feedback from the XML Finland Association members showed us the clear interest in these topics. At that time, Finnish universities, research centers, and companies were initiating research and development projects in this area. The XML Finland Association together with the patrons and sponsors of the conference are glad to have the opportunity to bring you the latest international and domestic news and research results concerning the Semantic Web and Web Services.

We were honored to have as the keynote speakers two leading experts: Prof. Dieter Fensel from the Leopold-Franzens Universität, Innsbruck, and the European OntoWeb network, and Research Fellow Ora Lassila from the Nokia Research Center, Boston. Their presentations enlightened the current state of the art in the Semantic Web and Web Services as well as envisioned the future developments. In addition to the keynote presentations, several invited talks, hands-on tutorials, and technical papers were presented. To complement the conference, major vendors presented their products and solutions in an exhibition.

One of the most important goals of the XML Finland conferences is to bring together XML researchers, professional users, and interested people with less experience of using the technologies. We hope that the XML Finland 2002 conference enabled its participants to make contacts and share experiences and ideas. This proceedings – together with a separate volume containing the research papers – makes the ideas and results available to a larger audience.

We would like to take the opportunity to thank all the members of the XML Finland 2002 Organizing Committee and Program Committee, authors, presenters, exhibitors, and the participants of the conference. We would also like to thank our supporters OntoWeb, Sun Microsystems, Microsoft, TIEKE Finnish Information Society Development Centre, Helsinki Institute for Information Technology (HIIT), the University of Helsinki, Nokia, Done Information, eCraft Management Solutions, Genio, Tieturi, Citec, Dataclub, Index IT, Innovative Ideas, Ontopia, Republica, Profium, and TietoEnator.

Helsinki, October 15, 2002

Kaisa Kostiainen
Done Information
Chair of the XML Finland 2002 Organizing Committee

Mika Klemettinen
Nokia Research Center
Chair of the XML Finland 2002 Program Committee

Eero Hyvönen
University of Helsinki and Helsinki Institute for Information Technology
Co-Chair of the XML Finland 2002 Program Committee

Conference Organization

Program Committee

Chair

Mika Klemettinen, Nokia Research Center

Co-Chair

Eero Hyvönen, University of Helsinki and HIIT

Members

Leyla Akgez, Nokia
Tomi Kauppinen, Genio Innovations
Kaisa Kostiainen, Done Information
Mikko Lounela, Research Institute for the Languages of Finland
Mikko Paananen, Republica
Maija Pennanen, Geological Survey of Finland
Petri Pusa, SysOpen
Jörgen Westerling, eCraft Management Solutions

Organizing Committee

Chair

Kaisa Kostiainen, Done Information

Members

Leyla Akgez, Nokia
Eero Hyvönen, University of Helsinki and HIIT
Tomi Kauppinen, Genio Innovations
Mika Klemettinen, Nokia Research Center
Mikko Lounela, Research Institute for the Languages of Finland
Mikko Paananen, Republica
Maija Pennanen, Geological Survey of Finland
Petri Pusa, SysOpen
Jörgen Westerling, eCraft Management Solutions

Table of Contents

Invited Talk

Tutorials

Presentations

Presentations only on the Web

Key Note

Towards the Semantic Web
Ora Lassila
Nokia Research Center

Invited Talks

.Net and Web Services
Kimmo Bergius
Microsoft

SunONE and Web Services
Hans Appel
Sun Microsystems

Tutorial

Enabling the Semantic Web with Web Services
Jens-Jacob Andersen
The Danish Post

Presentations

FIPA: Agents Meet the Semantic Web
Heimo Laamanen
Sonera and FIPA

ebXML
Pekka Rautiainen
EP-Logistics Ltd.

UDDI, WSDL, and SOAP
Nicklas Andersson
eCraft Management Solutions

Web Services for Citizens
Jouko Salonen
Republica

Network Publishing - Publishing to Web, Print, and Mobile Using XML, SVG, and RDF
Mikael Ahlavuo
Adobe

Structured VAT Information at the National Board of Taxes – A Case About Improved Searchability and Multichannel Publishing
Mikael Hällström
National Board of Taxes

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

Diffuse: Tracking Standards Development for the Information Society

Martin Bryan
Technical Manager, Diffuse Project
mtbryan@diffuse.org

The Diffuse Project

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

Background

- Open Information Interchange (OII) initiative under INFO2000 (1994-2000)
  - Grew from <100 standards to over 2000 in 20 sections
  - Printed list distributed to mailing list
  - Web site started in 1995
- IST Diffuse Project (2000-2002)
  - Over 3000 standards referenced in 30 sections
  - 30 Business Guides
  - Over 320 IST RTD projects using/developing standards
  - Monthly/quarterly reports on standards and RTD project developments
  - Over 4000 users from up to 100 countries each week
  - Japanese version of Diffuse.org maintained by Japanese Standards Association

The Diffuse Project

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

IST Action Lines

- Key Action Line I: Systems and Services for the Citizen
- Key Action Line II: New Methods of Work and Electronic Commerce
- Key Action Line III: Multimedia Content and Tools
- Key Action Line IV: Essential Technologies and Infrastructure
- Cross Programme Themes and Future and Emerging Technologies

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

The Diffuse Team

- Project Manager
  - Aatto J. Repo, TIEKE (Finnish Information Society Development Centre)
- Editorial Manger
  - Man-Sze Li, IC Focus (UK)
- Technical Manager
  - Martin Bryan, The SGML Centre (UK)
- RTD Projects Rapporteur
  - Marja-Leena Vespäläinen, TIEKE (Finland)

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

Who is Diffuse used by?

- RTD project teams looking for relevant standards
- Business managers looking for guidance on how best to apply standards within their organization
- Members of standards organizations needing to track what is going on elsewhere
- Public administrators needing to determine how best to direct policies

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

What Diffuse Offers

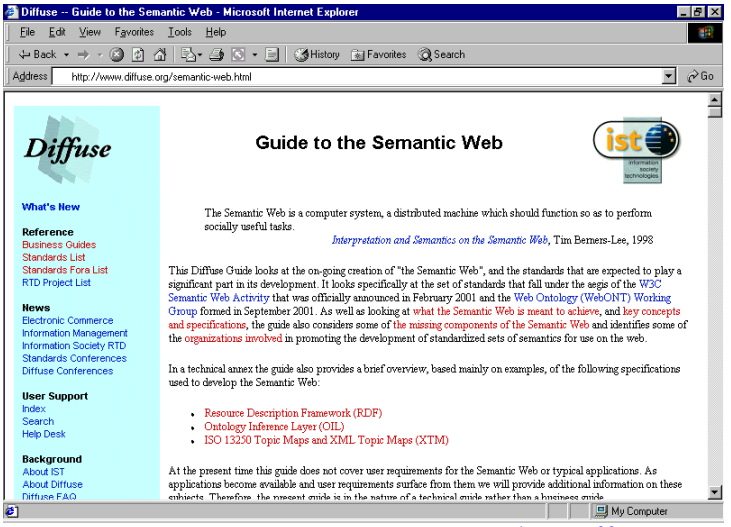
- Reference Material
  - Business Guides
  - Standards List
  - Standards Fora List
  - RTD Project Lists
- News
  - Electronic Commerce Interoperability
  - Information Management Standardization
  - IST RTD Standards Impact
  - Standards Conferences
- User Support
  - Index and Searching
  - Help Desk

Diffuse Business Guides

- Provide high-level, business-oriented, overviews of sets of related standards
- Cover specific business issues:
  - Electronic Payment
  - Open Source Licensing
  - Web Services
  - Knowledge Management
- Or specific technical issues
  - XML and related standards
  - Internationalization/Localization
  - Semantic Web
- 30 Guides published to date

The Diffuse Project

Example Business Guide



The Diffuse Project

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

Diffuse Standards List

- Provides a single source for information about standards generated by a wide range of standards fora
- Summarizes objectives of each standard and identifies uses to which it is put
- Where possible provides link to on-line sources of data (over 10,000 external links)
- Identifies IST RTD projects that have helped in the development of the standard, or are evaluating its use

The Diffuse Project

Diffuse

What's New

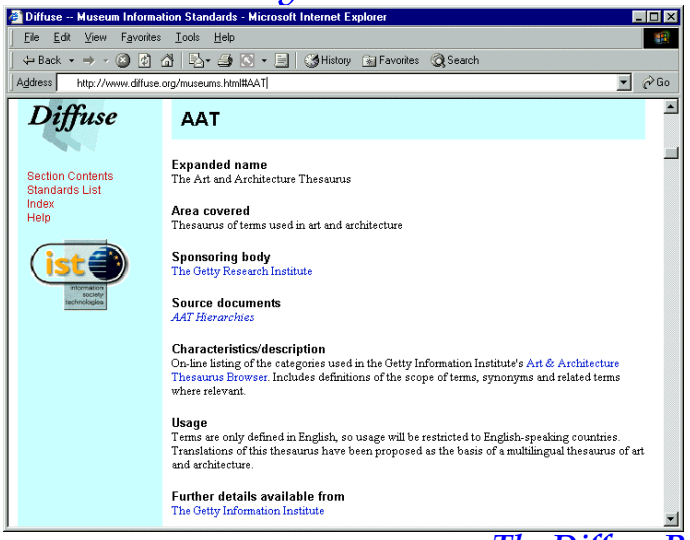
Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

What we say about standards



The Diffuse Project

Standard categories

APPLICATION SPECIFIC

Electronic Commerce
- Architectures
- Business semantics
- Electronic data interchange (EDI)
- Information security G
- Payment
- Product data G

Sectorial Data Interchange
- Geographic information
- Medical informatics
- Museum information
- Scientific information

GENERAL PURPOSE

Information Management
- Data classification
- Metadata interchange G
- Directories
- Archiving G
- Library information

Data Representation
- Character sets
- Text-based documents
- Multimedia/hypermedia
- Audio
- Video
- Raster graphics G
- Vector graphics G
- Colour information

Communications G
- File transfer
- Electronic mail and newsgroups
- Electronic conferencing
- Mobile data communication G
- Webcasting
- Digital television
- Application program interfaces G

G See also Business Guide to subject
✚ Entries updated this month

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

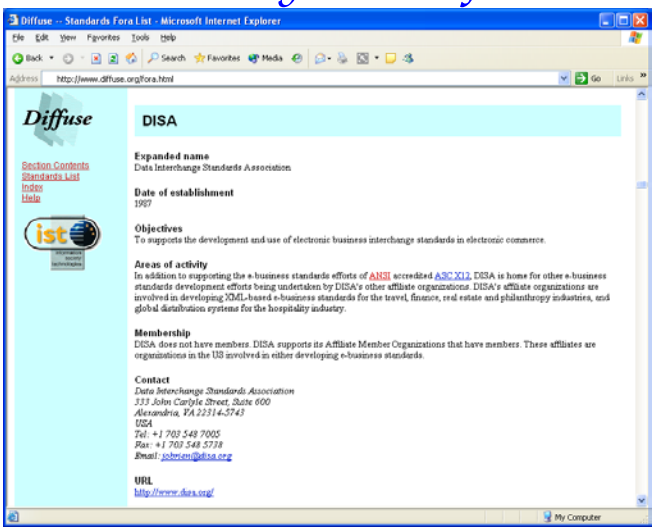
Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

Diffuse Standards Fora List

- Summarizes the objectives of standards body and stable industry consortia
- Provides details of the main areas of activity within each standards forum
  - ✦ For bodies with many technical committees links are provided to each relevant committee
- Summarizes the membership criteria
- Provides a contact address and a URL to the relevant web site

The Diffuse Project

What we say about fora



Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

Diffuse RTD Project List

- Identifies relationship between IT standards and IST RTD projects
- Summarizes details of projects
  - Name and acronym
  - Participants
  - Contacts
  - Website URL
- Summarizes project objectives
- Explains links to relevant standardization initiatives and referenced standards

Diffuse

What's New

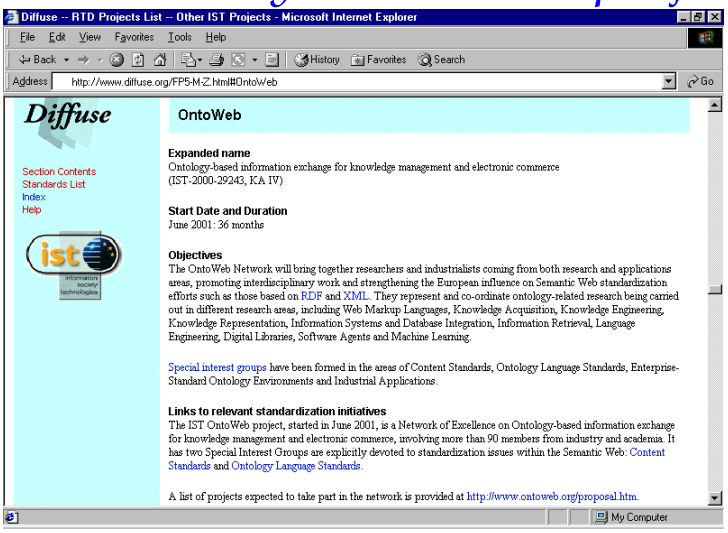
Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

What we say about RTD projects



The Diffuse Project

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

Keeping you up-to-date

- Monthly mailing to user community summarizing what is new on the site
- Monthly report of standards used for electronic commerce (KAII activities)
- Monthly report on standards impacting electronic content management (KAIII activities)
- Quarterly reports on impact of standards within IST RTD projects
- Reports on what happens at key European conferences related to IT standardization

The Diffuse Project

Diffuse Conferences

- 1st Diffuse Conference, 7th March 2001
  Convergence to Consolidation: What's Next in the Information Market?
- 2nd Diffuse Conference, 6th February 2002
  Will Web Services Revolutionize e-Commerce?
- Final Diffuse Conference, 12th December 2002
  Convergence of Web Services, Grid Services and the Semantic Web for delivering e-Services?
  Will also look at the role of Open Source Licensing for the distribution of Web Services

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

About Web Services

- Diffuse Guide to the Web Services
  - Web Services: What are they?
  - Web Services "standards" and their implementation
  - What next?
  - Emerging Web Services specifications
  - Overview of Grid Services
- Diffuse Conferences
- Reports from Standards Conferences
  - Global Grid Forum 5
  - First International Semantic Web Conference
- RTD Projects

Web Services Specifications

- From W3C
  - Web Services Description Language (WSDL)
  - Web Services Conversation Language (WSCL)
  - Web Services Description Requirements
  - Simple Object Access Protocol (SOAP)
- Other proposed specifications
  - Universal Description, Discovery and Integration (UDDI)
  - LDAP Schema for UDDI
  - Web Services for Interactive Applications (WSIA)
  - Web Services for Remote Portals (WSRP)
  - Business Process Execution Language for Web Services (BPE4WS)

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

About the Semantic Web

- Diffuse Guide to the Semantic Web
- Information Management Standards
  - Data classification
  - Metadata interchange
  - Directories
  - Archiving
  - Library information
- Reports from Standards Conferences
  - First International Semantic Web Conference
  - OntoWeb project meeting
  - IST Semantic Web Technologies Workshop
- RTD Projects

Semantic Web Standards

- From W3C
  - RDF and RDF Vocabulary (RDFS)
  - OWL Ontology Web Language
  - XLink and HLink
- From ISO
  - ISO 13250 Topic Maps
  - ISO 15938 Multimedia Content Description Interface (MPEG-7)
  - ISO 21000 Multimedia Framework (MPEG-21)
  - ISO 11179 Specification and Standardization of Data Elements
- and DARPA Agent Markup Language (DAML)

Diffuse

What's New

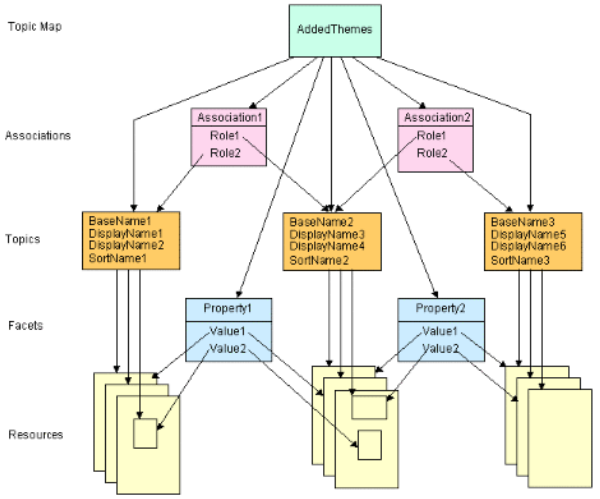
Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

What are Topic Maps?



What do Topic Maps tell us

- Topics can have multiple names
  - Different languages and different communities use different names to identify subjects
  - The same name may occur in more than one topic
- References to topics have roles
- Associations can have more than two ends
  - Each end plays a specific role
- You need to be able to merge Topic Maps generated by different communities
  - You cannot guarantee the use of a specific name by just one community

Diffuse

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

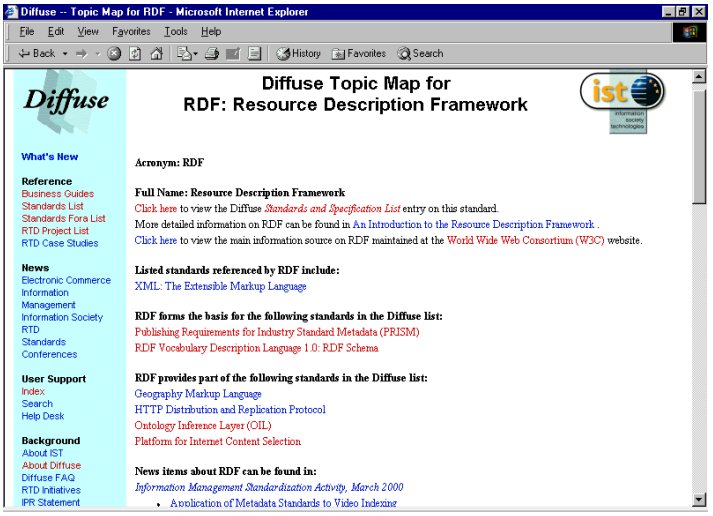
User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

How Diffuse uses Topic Maps

- As an extension to the alphabetical index
- As an extension to the list of Other references to monthly reports for complex subjects such as XML and RDF
- As a way of identifying RTD projects using a standard when it is widely adopted
- Using an XML schema based on stereotypes that map directly to ISO 13250
- Using XSLT to convert the XML files to HTML for dispay on a browser

Example of Diffuse Topic Map



What's missing?

- Multilingual semantics
  - Multiple meanings for same string within and between languages
  - Significant overlaps between multilingual terms
- Accurate specification of meaning
  - URL fragment identifiers based on Xquery
  - Scoping of meanings
  - Role definition
- Semantic Web navigation
- Semantics for business processes
- Workflow and fallback for integrated services

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

Meeting your needs

- The Diffuse Help Desk is available to help answer questions you may have about the relevance of standards within your organization
- If the standard you want to know about is not listed, let us know so that we can add it to one of the 31 categories of standards listed
- Let us know how standards have impacted your research work so that we can share this knowledge through our scheduled reports
- Join the Diffuse mailing list to get regular updates

The Diffuse Project

What's New

Reference
Business Guides
Standards List
Standards Fora List
RTD Project List

News
Electronic Commerce
Information Management
Information Society RTD
Standards Conferences

User Support
Index
Search
Help Desk

Background
About IST
About Diffuse
Diffuse FAQ
RTD Initiatives
IPR Statement
Disclaimer

Where to find us

- http://www.diffuse.org
- diffuse@tieke.org
- mtbryan@diffuse.org
- TIEKE, Salomonkatu 17A, 10th Floor, Helsinki
- Details of Final Diffuse Conference: http://www.diffuse.org/event3.html

The Diffuse Project

profium

Resource Description Framework (RDF) Tutorial

Janne Saarela
janne.saarela@profium.com

XML Finland 2002
Oct. 21st, 2002.

profium

Overview

- RDF background and raison d'etre
- Introduction to Resource Description Framework (RDF)
  - data model
  - syntax
  - model operations
  - schemas
- Link to
  - XML and XML Schemas
- RDF applications

profium

RDF background

- Back in 1997 W3C has developed vocabularies that were metadata i.e. information about information.
- In W3C case the information that was described was something on the Web i.e. something that had an identity in the form of a URI.
- E.g. Platform for Internet Content Selection (PICS) was a metadata standard addressing parental control over browsing.
  - technically PICS was based on S-expressions i.e. (for (good (old-lisp fans)))

profium

RDF background (cont'd)

- W3C got other proposals to work on things like
  - micropayments
  - IPR issues
- Instead of creating vocabularies W3C decided to develop an infrastructure where users ***can develop*** the vocabularies and ***can mix*** these vocabularies in a non-ambiguous way for resource description.
- Establishment of a dedicated working group back in late 1997.

profium

RDF background (cont'd)

- In late 1997 XML was getting ready but it did not address issues such as
  - description of URI based objects
  - non-ambiguous interpretation of semantics within XML documents
  - join operations between XML data models
- (BTW, XML still does not do these things for you)

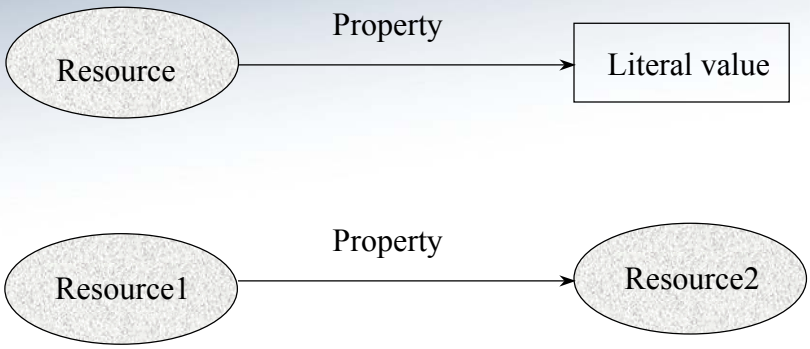
profium

Exercise 1

- How many alternative XML markups can you create for expressing the fact that 'the language on www.profium.com Web page is English')

profium

RDF data model



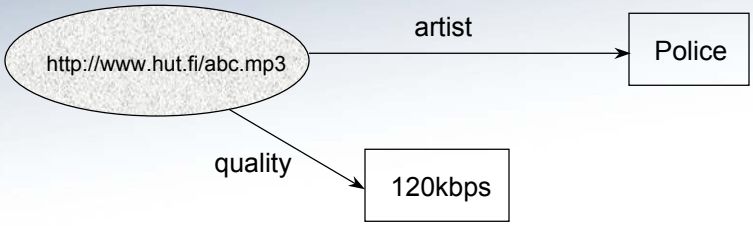
profium

RDF data model (cont'd)

- Resources are URI addressable objects
  - http://www.hut.fi/abc.mp3
  - mailto:janne.saarela@profium.com
  - irc://foobar.org/janne,isnick
- "A literal value can be of any XML basic data type"
  - How many XML basic data types are there in XML V1.0?
  - forward reference to XML schemas that are supposed to provide basic data types

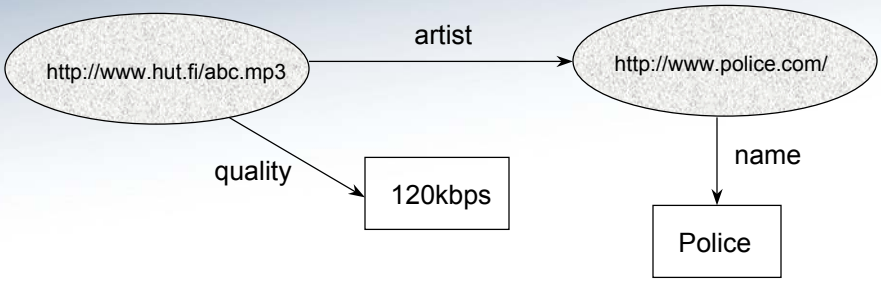
profium

RDF example 1A



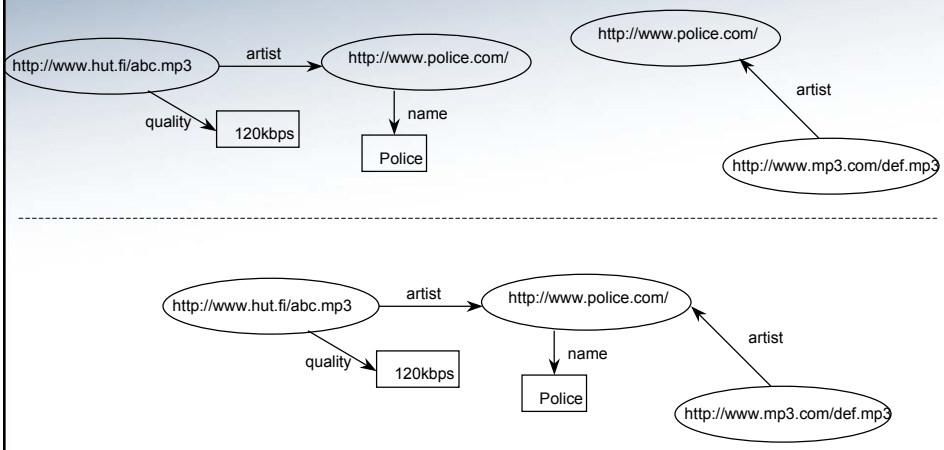
profium

RDF example 1B



profium

RDF interleaving



profium

RDF data model (cont'd)

- A labelled directed graph can also be expressed as 3-tuples or triples
  - (subject, predicate, object)
  - these 3-tuples are also called ***statements***
- those familiar with logic programming might like
  - predicate (subject, object) or
  - triple(subject,predicate,object).

profium

Exercise 2: mapping between graphs and triples

- Express the graph in the bottom of page 7 as a set of 3-tuples. Write the triples in this table

Subject	Predicate	Object

profium

RDF syntax

- RDF uses XML as the transfer encoding.
- Note: native XML data model is ***hierarchical***, native RDF data model is a ***directed graph***!
- We will now study how RDF constructs enable the graph encoding using XML

profium

RDF syntax: Step 1

- RDF V1.0 has a namespace reserved for its elements and attributes.
  - http://www.w3.org/1999/02/22-rdf-syntax-ns#
- The XML document has RDF as the root element. This is where RDF compatible applications notice the XML is indeed RDF.
- This type of XML documents are typically referred to as RDF/XML documents.
  - And their MIME type is application/rdf+xml

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
…
</rdf:RDF>
```

profium

RDF syntax: Step 2

- Inside the RDF root element we can have descriptions about resources (R and D of RDF).
- The `Description` element identifies the subject of a 3-tuple or a statement in the `about` attribute

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description about="http://www.hut.fi/abc.mp3">
     …
  </rdf:Description>
</rdf:RDF>
```

profium

RDF syntax: Step 3

- Any element nested inside the Description element is a predicate i.e. a property.
- A literal value of a property is encoded as element content.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description about="http://www.hut.fi/abc.mp3">
      <property>literalvalue</property>
  </rdf:Description>
</rdf:RDF>
```

- (http://www.hut.fi/abc.mp3, property, "literalvalue")

profium

RDF syntax: Step 3 (cont'd)

- However, the property needs to be globally unique for non-ambiguous processing. Therefore we need to associate a namespace URI with the property. This URI is the "home" of the RDF schema where the property is defined.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description about="http://www.hut.fi/abc.mp3">
      <dc:language>en</dc:language>
  </rdf:Description>
</rdf:RDF>
```

- (http://www.hut.fi/abc.mp3, http://purl.org/dc/elements/1.1/language, "en")

profium

RDF syntax: Step 4

- What if the value of a property is not literal but another resource?
- The value of the property is then identified using `resource` attribute from the rdf namespace.
  - Please note that the property element has become an empty element below.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:my="http://profium.com/arts/">
  <rdf:Description about="http://www.hut.fi/abc.mp3">
      <my:artist rdf:resource="http://www.police.com" />
  </rdf:Description>
</rdf:RDF>
```

- (http://www.hut.fi/abc.mp3, http://profium.com/arts/artist, http://www.police.com)

profium

RDF syntax: Step 5

- The previous examples have encoded single 3-tuples i.e. statements.
- One can have multiple statements in one RDF/XML document.
- Option 1: the statements that share the subject but have different properties can become children of the `Description` element.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:my="http://profium.com/arts/">
  <rdf:Description about="http://www.hut.fi/abc.mp3">
    <my:property1>value1</my:property1>
    <my:property2>value2</my:property2>
    <my:property3>value3</my:property3>
  </rdf:Description>
</rdf:RDF>
```

profium

RDF syntax: Step 5 (cont'd)

- Option 2: the statements that do not share the subjects force us to use multiple `Description` elements.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:my="http://profium.com/arts/">
  <rdf:Description about="http://www.hut.fi/abc.mp3">
    <my:property1>value1</my:property1>
  </rdf:Description>
  <rdf:Description about="http://www.police.com">
    <my:name>Police</my:name>
  </rdf:Description>
</rdf:RDF>
```

profium

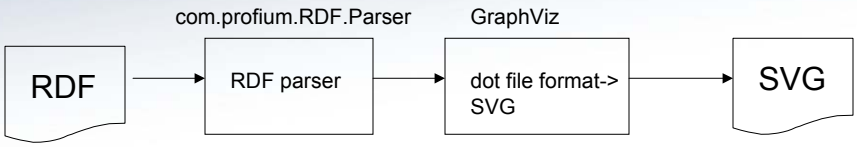
Exercise 3: data model operations in RDF

- Enter ex3 directory
- Edit the rdf/xml document `model.rdf` in this directory to encoding the graph on the bottom of page X of this tutorial.
- Parse and visualize the data model of this document by running `'parse model1.rdf'` on the command line
- Visualize the resulting `result.svg` file with your IE browser (It should look similar to the graph where we started from)

- We use Profium Java RDF Parser V1.6 to run this exercise. You can use this parser freely for non-commercial use.

profium

Background information to the exercise



profium

Containers in RDF

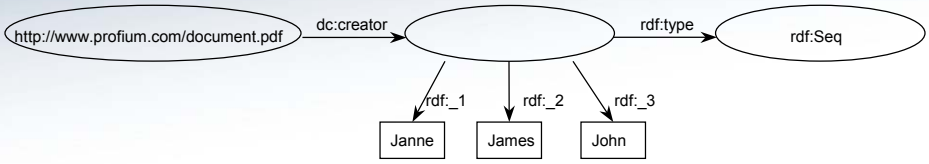
- RDF has support for primitive containers that can group values of a property
  - A Bag is an unordered collection of values
  - A Seq is an ordered collection of values
  - An Alt is a mutually exclusive set of values
- Example:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1">
  <rdf:Description about="http://www.profium.com/document.pdf">
    <dc:creator>
      <rdf:Seq>
        <rdf:li>Janne</rdf:li>
        <rdf:li>James</rdf:li>
        <rdf:li>John</rdf:li>
      </rdf:Seq>
    <dc:creator>
  </rdf:Description>
</rdf:RDF>
```

profium

Containers in RDF (cont'd)

- The previous RDF/XML expresses this data model

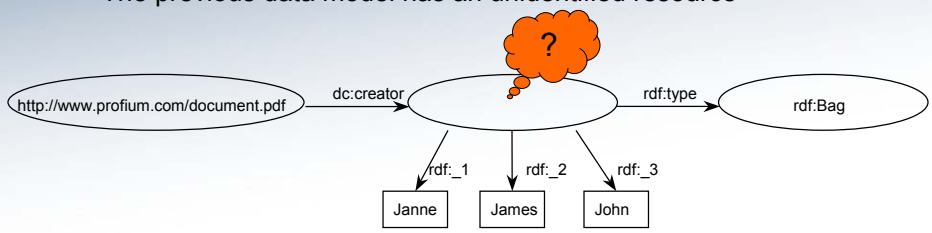


- Note: the <rdf:li> properties have become rdf:_1, rdf:_2, ... properties

profium

Blank node identifiers or bNodes

- The previous data model has an unidentified resource



- This node can either be manually labelled or the RDF parser creates a blank node identifier automatically

profium

Model operations in RDF

- A graph is a set of 3-tuples
- These sets can be operated with ***union/merge*** operation
- A merged graph from two separate graphs contains unique statements only once

Subject	Predicate	Object
A	p1	v1
A	p2	v2

Subject	Predicate	Object
A	p1	v1
v2	p3	v3

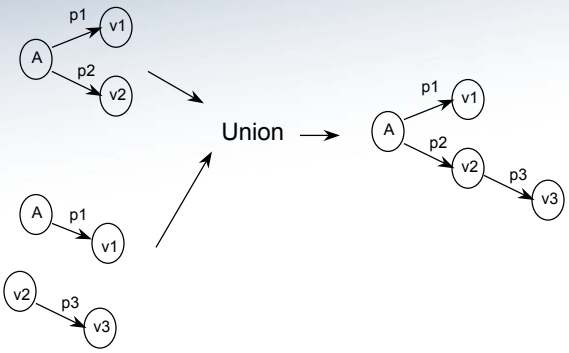
Union →

Subject	Predicate	Object
A	p1	v1
A	p2	v2
v2	p3	v3

profium

Model operations in RDF (cont'd)

- if you rather view the graphs...



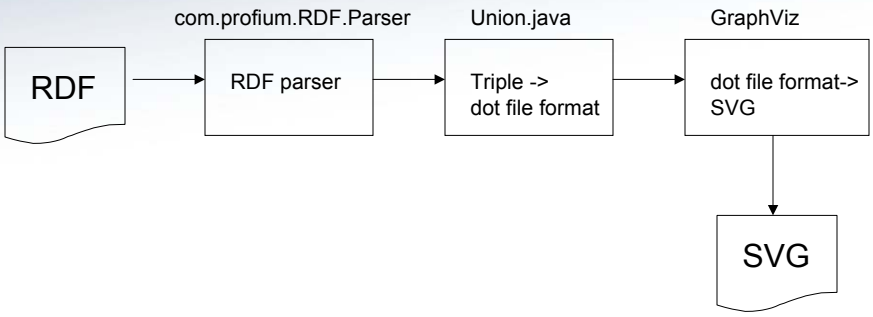
profium

Exercise 4: data model operations in RDF

- Enter ex4 directory
- Examine the following files
  - `model1.rdf`, `model2.rdf` and `model3.rdf` RDF descriptions
  - parse and visualize each of these separately by running `'parse model1.rdf'`
  - You can visualize the resulting `result.svg` file with your IE browser.
  - Finally, pass all the data models as a parameter and see the union of all of them.
    `'parse model1.rdf model2.rdf model3.rdf'`
- We use Profium Java RDF Parser V1.6 to run this exercise. You can use this parser freely for non-commercial use.

profium

Background information to the exercise



profium

RDF schemas

How to construct RDF vocabularies?

profium

RDF schemas

- RDF Schemas establish a type system
- While doing so, they define
  - properties one can use
  - validation constraints
- You can think of RDF Schemas as XML DTDs or XML Schemas
  - you can validate RDF data models against RDF schemas
- Note: RDF Schemas have been W3C Candidate recommendations for long
  - RDF Core Working Group has the goal to finish them by the end of 2002
  - whatever is presented next may change...

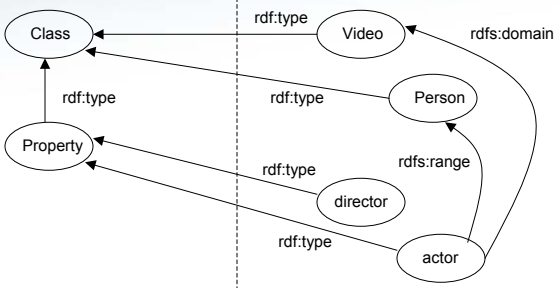
profium

Example RDF schema

RDF model & syntax
(rdf: prefix)

RDF schemas
(rdfs: prefix)

Video library schema
(video: prefix)



profium

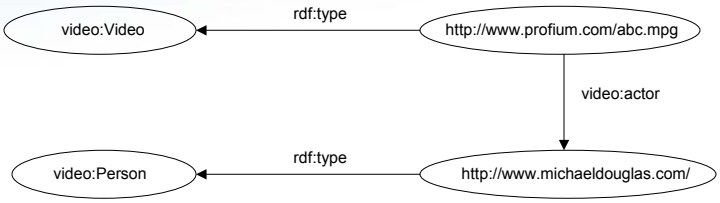
RDFS constraints

- (video:actor, 'rdfs:range', 'video:Person')
  - Please note: RDFS uses RDF data model
  - ***range*** defines what type actor property must have
- (video:actor, 'rdfs:domain', 'video:Video')
  - ***domain*** defines what type the subject where actor property is applied must have

profium

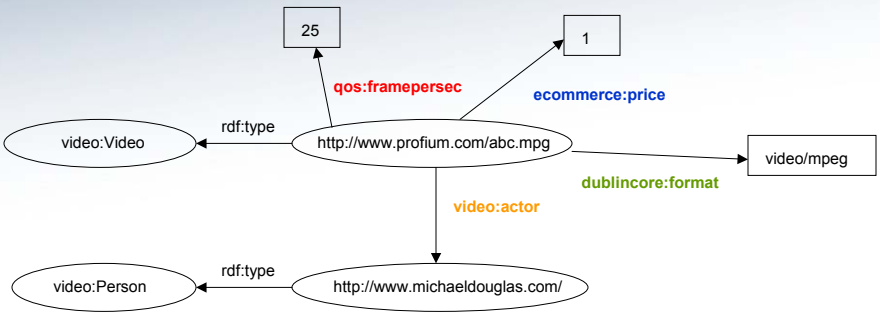
RDF data model to be validated

- This data model is valid w.r.t. RDF schema



profium

RDF: mix-and-match architecture



profium

Exercise 5: model an RDF schema

- Enter ex5 directory
- Use Oiled to create an RDF schema that encodes the video library schema.
  - start Oiled by double-clicking on oiled.bat
- Create two classes and two properties
- Export the design into as simple RDFS file in the ex5 directory
- Parse and visualize the schema by running `'parse schema.rdf'`
  - You can visualize the resulting `result.svg` file with your IE browser.

- Oiled is (C)opyright 2001 University of Manchester and can be downloaded from http://oiled.man.ac.uk/. Oiled is under GPL.

profium

Relationship between RDF and XML

What were the differences again?

profium

RDF and XML

- RDF uses XML transfer syntax. RDF introduces special elements that enable the encoding of directed graphs using XML.
- RDF literal nodes use XML basic data types.
  - Until RDF Core WG decides the typing issues, the only supported literal node type is Unicode string or XML markup.
  - XML basic data types are available through XML Schemas: Data types specification at W3C. Link between RDF and XML Schemas is work-in-progress.

profium

RDF applications

A quick look on the marketplace

profium

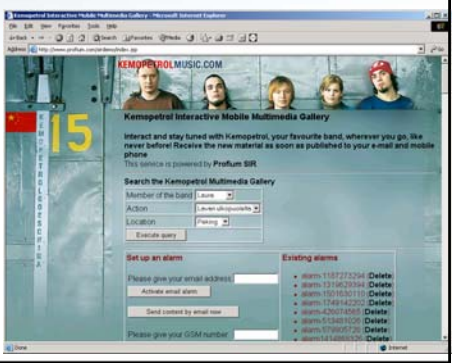
RDF applications

- General RDF based vocabularies exist
  - http://www.dublincore.org
  - http://www.prismstandard.org
- General RDF parsing components exist for a variety of programming languages
  - http://www.ilrt.bris.ac.uk/discovery/rdf/resources/

profium

RDF applications (cont'd)

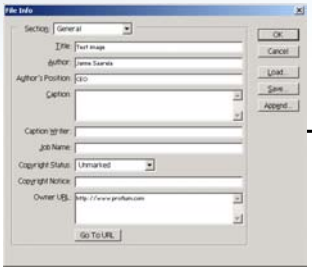
- RDF middleware offerings
  - Profium SIR, Ontoprise, IA Jitzu, Sesame, ...
  - e.g. multimedia gallery at
  - http://www.kemopetrolmusic.com



profium

RDF applications (cont'd)

- RDF compatible authoring tool offering
  - Adobe Extensible Metadata Platform (XMP) within Acrobat 5.0, Illustrator 10, InDesign 2, Photoshop 7




```
<?xpacket begin='' id='W5M0MpCehiHzreSzNTczkc9d'
bytes='1336'?>

<rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-
rdf-syntax-ns#'
 xmlns:iX='http://ns.adobe.com/iX/1.0/'>

 <rdf:Description about=''
  ...
```

profium

RDF applications (cont'd)

- RDF online services
  - RSS: Rich Site Summaries
  - e.g. www.newsnow.co.uk

profium

Untouched areas of this tutorial

What remains as you a home exercise?

profium

RDF data model special features

- ***reification*** – the ability to make statements about statements i.e. higher-level statements
- ***abbreviated syntax*** – the ability to write RDF/XML documents with shorted syntactic structures
- RDF/XML use of XML facilities such as XML Base

profium

References

- RDF Model and Syntax specification
  - http://www.w3.org/TR/REC-rdf-syntax
- RDF Schemas specification
  - http://www.w3.org/TR/rdf-schema/
- RDF model theory
  - http://www.w3.org/TR/rdf-mt/
- RDF Interest Group
  - http://www.w3.org/RDF/Interest/
- Link to RDF resources
  - http://www.w3.org/RDF

profium

Summary

- RDF background and raison d'etre
- Introduction to Resource Description Framework (RDF)
  - data model
  - syntax
  - model operations
  - schemas
- Link to
  - XML and XML Schemas
- RDF applications

profium

Future directions with RDF

- RDF Query languages is still unspecified.
  - A declarative query language needs to be developed
- RDF processing APIs are currently ad-hoc standards
  - A procedural API à la DOM is required for RDF
- Expressivity limits between RDFS technology and Web Ontology Language (OWL) technology need to be developed

profium

Contact information

Janne Saarela
Profium Oy
Lars Sonckin kaari 12
02600 Espoo
Finland
Tel. +358 (0)9 855 98 000
Fax. +358 (0)9 855 98 002
Email. janne.saarela@profium.com
Internet. www.profium.com

XML-based Natural Language Generation

Graham Wilcock
Docent of Language Technology
University of Helsinki
graham.wilcock@helsinki.fi

Acknowledgements

- Part 1 (What is NLG?)
  is largely based on Stephan Busemann's
  Natural Language Generation: An Overview
  (http://www.dfki.de/~busemann/VL-SS99/)
- Parts 2 (XML-based NLG) and 3 (demos)
  are based on work in the USIX-Interact project.
  (http://www.mlab.uiah.fi/interact/)

Requirements

- For this tutorial, you need:
  - a general knowledge of XML
  - a basic idea of XSLT transformations
- You don't need:
  - any previous knowledge of natural language generation

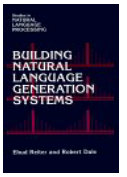
Tutorial Outline

- Part 1: What is Natural Language Generation?
  - "Shallow" and "deep" NLG
- Part 2: XML-based NLG
  - Templates + trees + pipelines
  - A step-by-step example
- Part 3:Demos
  - Web NLG with Cocoon servlets
  - Voice NLG with XML + Java speech

Part 1: What is NLG?

Natural language generation is the process of deliberately constructing a natural language text in order to meet specified communicative goals. [McDonald 1992]

- Recommended textbook**:**
  Ehud Reiter & Robert Dale,
  Building Natural Language Generation Systems,
  Cambridge University Press, 2000.



NLG inputs and outputs

- NLG input
  - a communicative goal, including
  - non-linguistic representation of information
- NLG output
  - a text, plain or formatted (HTML, JSML)
  - may be combined with graphics, tables etc.
- Knowledge sources required
  - domain-specific knowledge
  - language-specific knowledge
  - knowledge about human communication

Template- v. plan-based NLG

- "Template-based" NLG
  - Canned texts with variable slots
  - String manipulation (often in Perl)
  - Single-shot processing
- "Plan-based" NLG
  - Text planning and sentence planning
  - Tree structure transforms (often in Java)
  - Multi-stage, multi-level processing

Tasks of plan-based NLG

- Content determination
- Discourse planning
- Lexicalization
- Referring expression generation
- Sentence aggregation
- Surface realization

NLG: Content Determination

- Deciding what to say
  - Construct a set of MESSAGES from the data source
  - A message may become a word, phrase or sentence
  - Messages are based on domain entities (concepts)

IDENTITY(NEXTSHIP, MS-LILLY) → *The next ship is the MS-LILLY.*

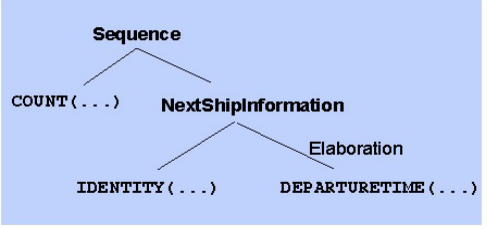
DEPARTURETIME(MS-LILLY, 1000) → *The MS-LILLY departs at 10am.*

COUNT(SHIP, SOURCE(HAMBURG), DESTINATION(COPENHAGEN), 5, PERDAY)
→ *There are 5 ships daily from Hamburg to Copenhagen.*

NLG: Discourse Planning

- Organize messages in a coherent text plan
  - A text is structured, not random sentences
  - Conceptual grouping, rhetorical relations

There are five ships daily from Hamburg to Copenhagen. The next ship is the MS-LILLY. It departs at 10am.



NLG: Lexicalization

- Mapping from domain concepts to lexemes
- Determines the content words to be used
  - discourse focus - *buy vs sell*
  - collocations - *exert influence, administer punishment*
  - lexical semantics - *male unmarried adult vs bachelor*
  - basic level categories - *dog vs poodle*
  - attitude - *house vs home, father vs dad*
  - idioms - *kick the bucket*
- Partly determines the syntactic structure

NLG: Referring Expressions

- Allow the hearer to identify discourse objects
- Kinds of referring expressions
  - Proper names - *Hamburg, The United States*
  - Definite descriptions - *the ship that leaves at 10*
  - Proforms - *it, later, there*
- Initial reference
  - use a full name - *the MS-LILLY*
  - relate to a salient object - *the ship's snack bar*
  - specify physical location - *the ship at pier 12*

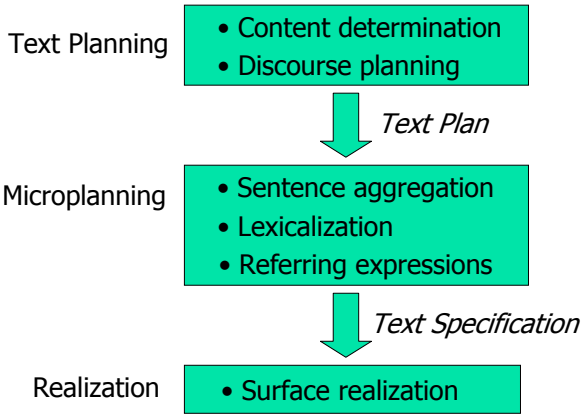
NLG: Sentence Aggregation

- Mapping from messages to sentences
  - One-to-one mapping results in poor text
  - Need to combine messages in complex sentences
- Without aggregation
  - *The next ship is the MS-LILLY. It leaves Hamburg at 10am. It has a restaurant. It has a snack bar*.
- With aggregation
  - *The next ship, which leaves Hamburg at 10am, is the MS-LILLY. It has a snack bar and a restaurant.*

NLG: Surface Realization

- Converts text specifications into output text
- Generates grammatically correct text
  - insert function words - *he wants to book a ticket*
  - word inflection – *like* + *ed* → *liked*
  - grammatical word order
- Techniques for grammatical knowledge
  - declarative bidirectional grammars
  - grammars designed for generation
  - templates, easy and fast to implement

NLG Pipeline Architecture



Deep v. Shallow Generation

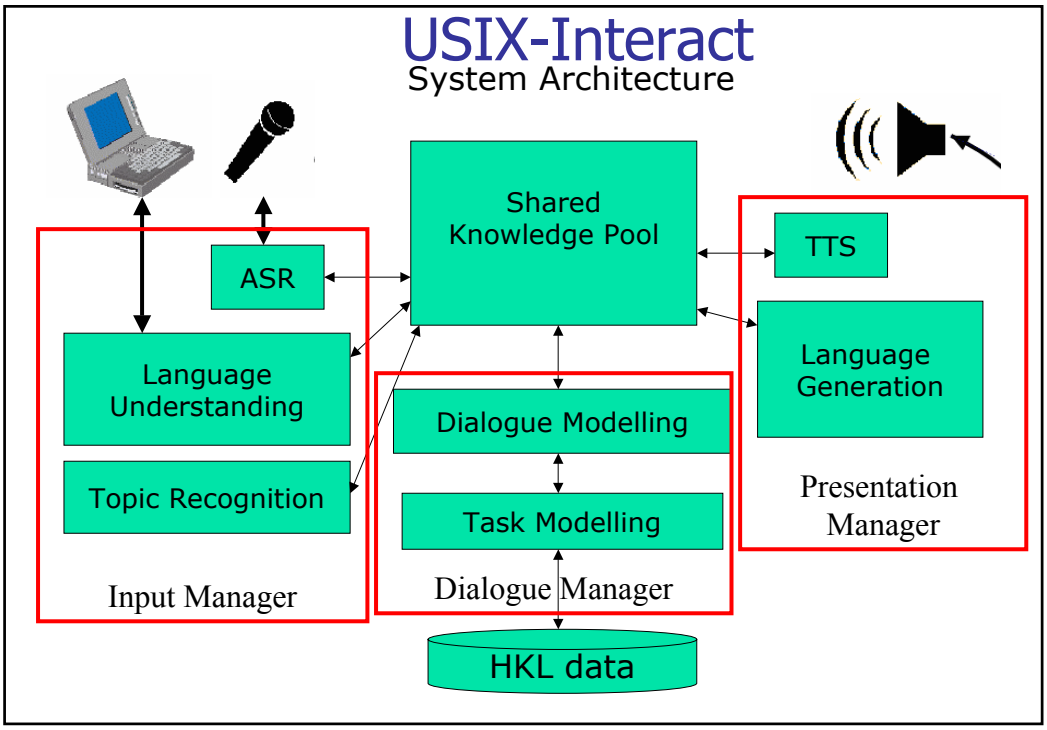
- Deep generation tries to model everything
  - Research paradigm in NLG
  - Aims at general solutions, re-usable and scalable
  - Works for small domains, limited linguistic coverage
- Shallow generation aims at real applications
  - Commit to domain and task-specific choices
  - No general solutions attempted
  - Coverage based on domain corpus
  - Allow different methods within the same system
- Shallow generation: Stephan Busemann (DFKI)

Part 2: XML-based NLG

- Templates
  - Implement NLG template-based generation using predefined XSLT templates
- Trees
  - Implement NLG tree structures (text plans, text specifications) using XML tree structures
- Pipelines
  - Implement NLG pipeline architecture using piped sequences of XML-to-XML transformations

A Step-by-Step Example

- Simplified (rapid prototyping)
  - Quickly-developed "first prototype" generator for USIX-Interact spoken dialogue project
  - Dialogue response planning and generation
- XML + XSLT implementation
  - Agenda of domain concepts
  - Template-based text plans ("aggregations")
  - Pipeline of XSLT transformations



Response planning

- New Information (NewInfo)
  - Response planning starts from NewInfo
  - Response always includes NewInfo
- Topic
  - NewInfo may be *wrapped* by Topic link
  - Polite response: Topic and NewInfo
  - Elliptical response: NewInfo only

Generation from NewInfo (Ex.1)

- *Which bus goes to Malmi?*
  - Topic: transportation: bus
  - Topic: destination: Malmi
  - NewInfo: bus number: 74
- *Number 74.*

Input: Agenda in XML

- Unordered set of domain concepts
  - Marked as "Topic" or "NewInfo"
- Specified by Dialogue Manager
- Starting point for Generator
  - Shared XML representation

Input: Agenda (Ex.1)

```
<agenda id="1">
  <concept info="Topic">
    <type>transportation</type>
    <value>bus</value>
  </concept>
  <concept info="Topic">
    <type>destination</type>
    <value>malmi</value>
  </concept>
  <concept info="Topic">
    <type>bus</type>
    <value>exists</value>
  </concept>
  <concept info="NewInfo">
    <type>busnumber</type>
    <value>74</value>
  </concept>
</agenda>
```

- *Number 74.*

Template-based NLG in XSLT

- Create predefined XSLT named templates
  - Decide a set of predefined Text Plan tree structures (here, the structures are called "aggregations")
  - Define one named template per Text Plan aggregation
- XSLT processing
  - Test the concept values given in Agenda
  - Select a Text Plan based on concept values
  - The named template for the Text Plan copies Agenda concepts into its slots using <xsl:copy-of>

Text Planning: Select the Plan

```
<xsl:template match="agenda">
<xsl:choose>
<xsl:when test="concept[@info='NewInfo']/type='transportation'">
  <xsl:call-template name="BY-TRANSPORT"/>
</xsl:when>
<xsl:when test="concept[@info='NewInfo']/type='bus'">
  <xsl:choose>
  <xsl:when test="concept[@info='NewInfo']/type='busnumber'">
    <xsl:call-template name="NUM-DEST-TIME"/>
  </xsl:when>
  ...
  </xsl:choose>
</xsl:when>
<xsl:when test="concept[@info='NewInfo']/type='busnumber'">
  <xsl:call-template name="NUMBER-ONLY"/>
</xsl:when>
...
</xsl:choose>
</xsl:template>
```


Text Planning: Insert Concepts

```
<xsl:template name="NUM-DEST-TIME">
<aggregation>
  <subject cat="NP">
    <xsl:copy-of select="./concept[type='busnumber']"/>
  </subject>
  <predicate cat="VP">
    <xsl:copy-of select="./concept[type='bus']"/>
  </predicate>
  <complement cat="PP">
    <xsl:copy-of select="./concept[type='destination']"/>
  </complement>
  <adjunct cat="PP">
    <xsl:copy-of select="./concept[type='bustime']"/>
  </adjunct>
</aggregation>
</xsl:template>
```

Text Planning: Text Plan (Ex.1)

```
<agenda id="1">
  <concept info="Topic">
    <type>transportation</type>
    <value>bus</value>
  </concept>
  <concept info="Topic">
    <type>destination</type>
    <value>malmi</value>
  </concept>
  <concept info="Topic">
    <type>bus</type>
    <value>exists</value>
  </concept>
  <concept info="NewInfo">
    <type>busnumber</type>
    <value>74</value>
  </concept>
</agenda>
```

- Text Plan

```
<aggregation id="1">
  <subject cat="NP">
    <concept info="NewInfo">
      <type>busnumber</type>
      <value>74</value>
    </concept>
  </subject>
</aggregation>
```


NLG Pipeline in XML

- Pipeline architecture
  - XML input (agenda), XML output (JSML)
  - Sequence of XML-to-XML transformations
  - agenda → aggregation → speech markup
- Tree transformations in DOM or XSLT
  - Modify tree structures and nodes in them
  - For example (in Referring Expressions stage): replace domain concepts in aggregation tree by linguistic referring expressions

Microplanning: NewInfo

```
<!-- REFERRING EXPRESSIONS: DESCRIPTIONS -->
<xsl:template match="concept[@info='NewInfo']">
  <xsl:choose>
  <xsl:when test="type='busnumber'">
    <word>number</word>
    <word><xsl:value-of select="value/text()"/></word>
  </xsl:when>
  <xsl:when test="type='destination'">
    <word>to</word>
    <word><xsl:value-of select="value/text()"/></word>
  </xsl:when>
  <xsl:when test="type='bustime'">
    <word>at</word>
    <word><xsl:value-of select="value/text()"/></word>
  </xsl:when>
  </xsl:choose>
</xsl:template>
```


Microplanning: Pronouns

```
<!-- REFERRING EXPRESSIONS: PRONOUNS -->
<xsl:template match="concept[@info='Topic']">
  <xsl:choose>
  <xsl:when test="type='busnumber'">
    <word>it</word>
  </xsl:when>
  <xsl:when test="type='destination'">
    <word>there</word>
  </xsl:when>
  <xsl:when test="type='bustime'">
    <word>then</word>
  </xsl:when>
  </xsl:choose>
</xsl:template>
```

Microplanning: Text Spec (Ex.1)

- Text Plan

```
<aggregation id="1">
  <subject cat="NP">
    <concept info="NewInfo">
      <type>busnumber</type>
      <value>74</value>
    </concept>
  </subject>
</aggregation>
```

- Text Specification

```
<aggregation id="1">
  <subject cat="NP">
    <word>number</word>
    <word>74</word>
  </subject>
</aggregation>
```


Realization: Output (Ex.1)

- Text Specification

```
<aggregation id="1">
  <subject cat="NP">
    <word>number</word>
    <word>74</word>
  </subject>
</aggregation>
```

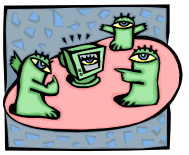
- Output

```
<jsml lang="en">
  <div type="sent">
    number 74
  </div>
</jsml>
```

(Java Speech Markup Language)

Another Example (Ex.2)

- *How do I get to Malmi?*
  - Topic: destination: Malmi
  - NewInfo: transportation: bus
  - NewInfo: bus number: 74
- *By bus – number 74 goes there.*



Input: Agenda (Ex.2)

```
<agenda id="2">
  <concept info="NewInfo">
    <type>transportation</type>
    <value>bus</value>
  </concept>
  <concept info="Topic">
    <type>destination</type>
    <value>malmi</value>
  </concept>
  <concept info="NewInfo">
    <type>bus</type>
    <value>exists</value>
  </concept>
  <concept info="NewInfo">
    <type>busnumber</type>
    <value>74</value>
  </concept>
</agenda>
```

Text Planning: Text Plan (Ex.2)

```
<agenda id="2">
  <concept info="NewInfo">
    <type>transportation</type>
    <value>bus</value>
  </concept>
  <concept info="Topic">
    <type>destination</type>
    <value>malmi</value>
  </concept>
  <concept info="NewInfo">
    <type>bus</type>
    <value>exists</value>
  </concept>
  <concept info="NewInfo">
    <type>busnumber</type>
    <value>74</value>
  </concept>
</agenda>
```

```
<aggregation id="2">
  <adjunct cat="PP">
    <concept info="NewInfo">
      <type>transportation</type>
      <value>bus</value>
    </concept>
  </adjunct>
  <prosody cat="pause"/>
  <subject cat="NP">
    <concept info="NewInfo">
      <type>busnumber</type>
      <value>74</value>
    </concept>
  </subject>
  <predicate cat="VP">
    <concept info="NewInfo">
      <type>bus</type>
      <value>exists</value>
    </concept>
  </predicate>
  <complement cat="PP">
    <concept info="Topic">
      <type>destination</type>
      <value>Malmi</value>
    </concept>
  </complement>
</aggregation>
```


Microplanning: Text Spec (Ex.2)

```
<aggregation id="2">
  <adjunct cat="PP">
    <concept info="NewInfo">
      <type>transportation</type>
      <value>bus</value>
    </concept>
  </adjunct>
  <prosody cat="pause"/>
  <subject cat="NP">
    <concept info="NewInfo">
      <type>busnumber</type>
      <value>74</value>
    </concept>
  </subject>
  <predicate cat="VP">
    <concept info="NewInfo">
      <type>bus</type>
      <value>exists</value>
    </concept>
  </predicate>
  <complement cat="PP">
    <concept info="Topic">
      <type>destination</type>
      <value>Malmi</value>
    </concept>
  </complement>
</aggregation>
```

```
<aggregation id="2">
  <adjunct cat="PP">
    <word>by</word>
    <word>bus</word>
  </adjunct>
  <prosody cat="pause"/>
  <subject cat="NP">
    <word>number</word>
    <word>74</word>
  </subject>
  <predicate cat="VP">
    <word cat="V">
      <lexeme>go</lexeme>
      <features>3sg</features>
    </word>
  </predicate>
  <complement cat="PP">
    <word>there</word>
  </complement>
</aggregation>
```

Realization: Output (Ex.2)

```
<aggregation id="2">
  <adjunct cat="PP">
    <word>by</word>
    <word>bus</word>
  </adjunct>
  <prosody cat="pause"/>
  <subject cat="NP">
    <word>number</word>
    <word>74</word>
  </subject>
  <predicate cat="VP">
    <word cat="V">
      <lexeme>go</lexeme>
      <features>3sg</features>
    </word>
  </predicate>
  <complement cat="PP">
    <word>there</word>
  </complement>
</aggregation>
```

- Output

```
<jsml lang="en">
  <div type="sent">
    by bus
    <break size="large"/>
    number 74 goes there
  </div>
</jsml>
```


Part 3: Demos

- Voice NLG
  - Bilingual Finnish/English generator
  - FreeTTS Java speech synthesizer
  - Currently has only English voice
- Web NLG
  - Bilingual Finnish/English generator
  - Cocoon XML server implementation
  - Towards the Semantic Web

Demo: Voice NLG

- Implemented in Java
  - Java API for XML (JAXP) to execute XSLT
  - Java Speech API to execute speech synthesis
- XML-based NLG
  - Input: Annotation Graph in XML
  - Output: bilingual Finnish and English in JSML
- Java speech synthesis
  - Input: (currently) English in JSML
  - Output: speech

Annotation Graphs

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<AGSet id="AGSet" version="1.0"
   xmlns="http://www.ldc.upenn.edu/atlas/ag/"
   xmlns:xlink="http://www.w3.org/1999/xlink"
   xmlns:dc="http://purl.org/DC/documents/rec-dces-19990702.htm">

<AG id="AGSet_AG1" type="transcription" timeline="AGSet_Timeline1">
<Anchor id="AGSet_AG1_anchor1" offset="1" unit="index"/>
<Anchor id="AGSet_AG1_anchor2" offset="2" unit="index"/>

<Annotation type="DAct" id="AGSet1_AG1_annotation1"
   start="AGSet1_AG1_anchor2" end="AGSet1_AG1_anchor2">
<Feature name="score">1.0</Feature>
<Feature name="name">ALOITUS</Feature>
</Annotation>
```

FreeTTS

- A speech synthesizer in Java
  - Free open source from Sun
  - Based on Flite and Festival (Edinburgh)
  - Portable (needs Java 1.4)
  - JSAPI interface
- Current limitations (FreeTTS 1.1)
  - Supplied voices: English
  - JSML accepted but not applied

Demo: Web NLG

- NLG web server
  - XML-based NLG
- Cocoon implementation
  - Cocoon web publishing framework (Java)
  - Runs in a servlet engine (Tomcat: Java)
  - Uses XSLT transformers (Xalan: Java)
  - Configurable XSLT pipelines

What is cocoon ?

"Apache Cocoon is an XML publishing framework that raises the usage of XML and XSLT technologies for server applications to a new level. Designed for performance and scalability around pipelined SAX processing, Cocoon offers a flexible environment based on a separation of concerns between content, logic, and style. To top this all off, Cocoon's centralized configuration system and sophisticated caching help you to create, deploy, and maintain rock-solid XML server applications."

- Apache Cocoon website

Configurable pipelines

- Pipelines specified in Cocoon configuration file

```
<map:sitemap xmlns:map="http://apache.org/cocoon/sitemap/1.0">
  <!-- English examples: Jokinen & Wilcock (SIGDIAL-2001) -->
  <!-- Get an agenda, apply XSLT transforms, serialize as HTML -->
  <map:pipeline>
   <map:match pattern="generate(agendas/*.xml)">
    <map:generate  src="agendas/{1}.xml"/>
    <map:transform src="transforms/SIGDIAL-2001/aggregation.xsl"/>
    <map:transform src="transforms/SIGDIAL-2001/lexicalization.xsl"/>
    <map:transform src="transforms/SIGDIAL-2001/realization.xsl"/>
    <map:serialize type="html"/>
   </map:match>
  </map:pipeline>
```

Adding a Finnish pipeline

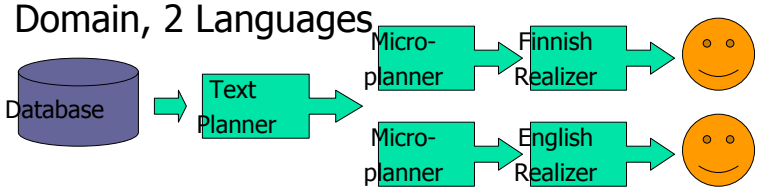
```
<!-- Finnish pipeline: Interact Demo 2002 -->
<!-- Get an A-Graph, apply XSLT transforms, serialize as HTML -->
<map:pipeline>
 <map:match pattern="suomeksi(agraphs/*.xml)">
  <map:generate src="agraphs/{1}.xml"/>
  <map:transform src="transforms/InteractDemo/AG-GetContent.xsl"/>
  <map:transform src="transforms/InteractDemo/AG-ResponsePlanner.xsl"/>
  <map:transform src="transforms/InteractDemo/FI-Lexicalization.xsl"/>
  <map:transform src="transforms/InteractDemo/FI-ReferringExps.xsl"/>
  <map:transform src="transforms/InteractDemo/FI-SurfaceRealizer.xsl"/>
  <map:transform src="transforms/displayverbatim.xsl"/>
  <map:serialize type="html"/>
 </map:match>
</map:pipeline>
```


Adding an English pipeline

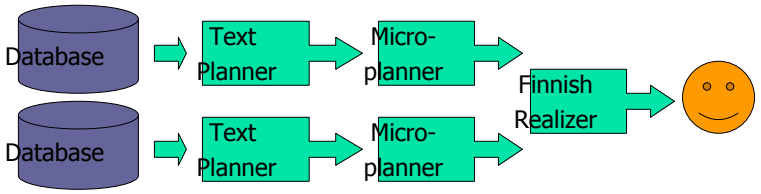
```
<!-- English pipeline: Interact Demo 2002 -->
<!-- Get an A-Graph, apply XSLT transforms, serialize as HTML -->
<map:pipeline>
 <map:match pattern="english(agraphs/*.xml)">
  <map:generate src="agraphs/{1}.xml"/>
  <map:transform src="transforms/InteractDemo/AG-GetContent.xsl"/>
  <map:transform src="transforms/InteractDemo/AG-ResponsePlanner.xsl"/>
  <map:transform src="transforms/InteractDemo/EN-Lexicalization.xsl"/>
  <map:transform src="transforms/InteractDemo/EN-ReferringExps.xsl"/>
  <map:transform src="transforms/InteractDemo/EN-SurfaceRealizer.xsl"/>
  <map:transform src="transforms/displayverbatim.xsl"/>
  <map:serialize type="html"/>
 </map:match>
</map:pipeline>
```

From Domains to Languages

- 1 Domain, 2 Languages

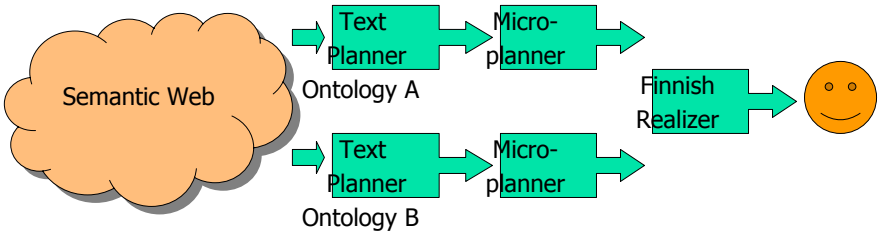


- 2 Domains, 1 Language



Towards the Semantic Web

- 2 Ontologies, 1 Language



-

XML Finland 2002

2002-10-22

Activities at the Finnish Information Society Development Centre, TIEKE:

Diffuse Project, ebXML, and eCommerce Development

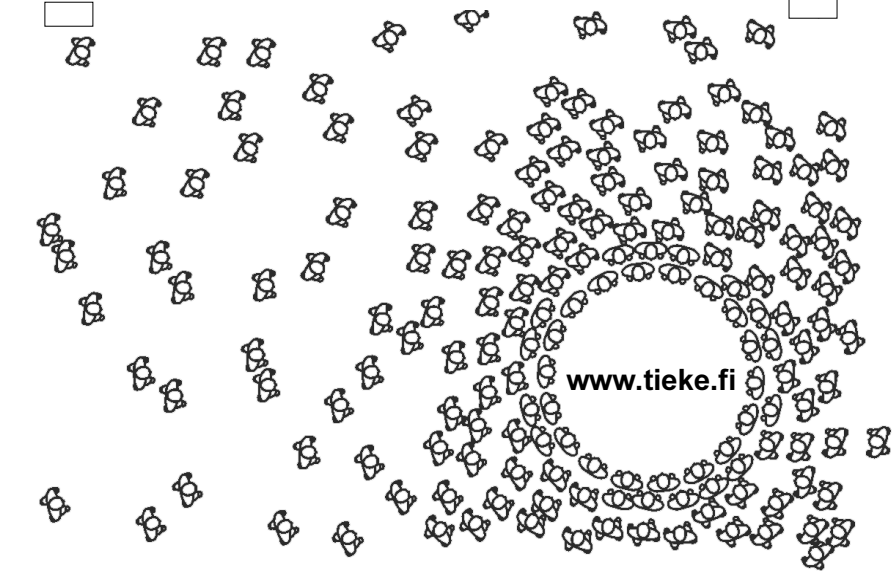
Jari Salo

Adviser

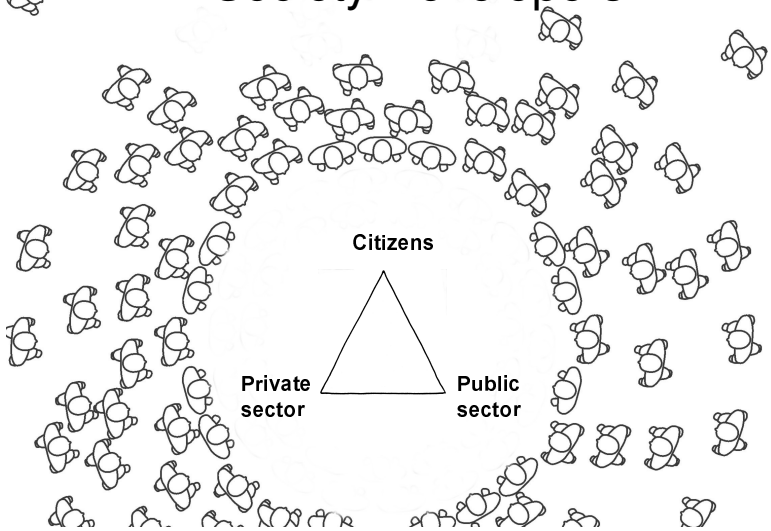
The Finnish Information Society Development Centre

jari.salo@tieke.fi

TIEKE Finnish Information Society Development Centre



Meeting Place for Information Society Developers



Our activities today

Interoperability

Focus:

- Electronic commerce
- Learning technologies
- Information security
  - *recommendations, guides*
  - *legal aspects, trade procedures, standards*

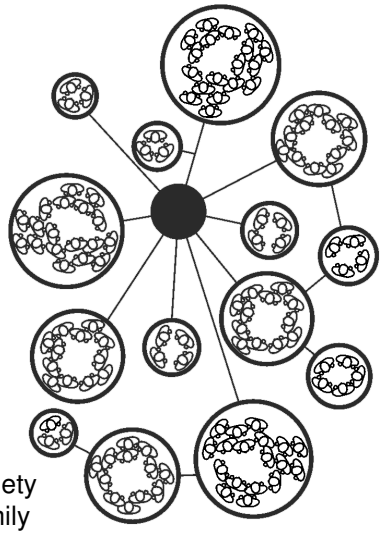
Networking

- *membership services*
- *ICT Cluster*
- *PR and information services*

Services for citizens

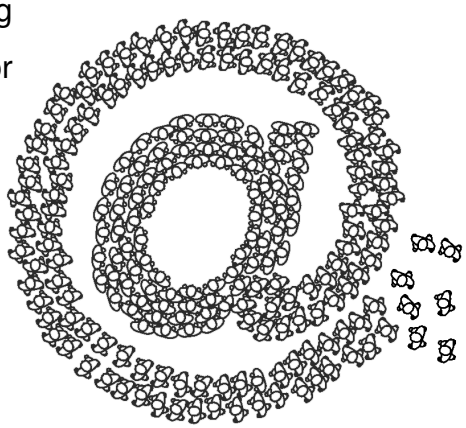
Focus:

- Qualifications for Information Society
  - Computer Driving Licence family
  - Adult education data base



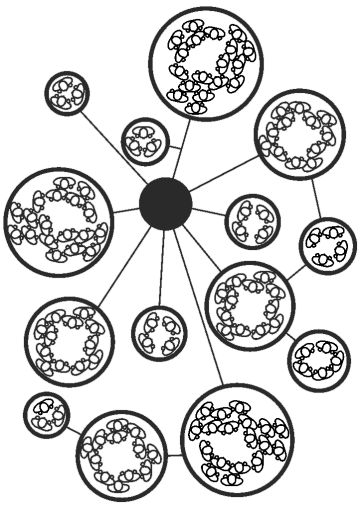
Our role in Information Society

- Act as a neutral player and an aggregate of know-how and networking
- Generate joint projects for implementation
- Promote awareness of national goals
- Increase international cooperation
- Initiate and perform development projects



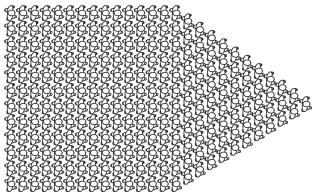
XML related Activities at TIEKE

- ebXML
  - Evolution-project
- Diffuse
- eLearning



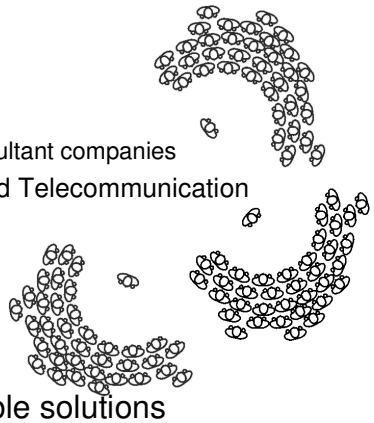
Our approach

Standards
Trade Procedures
UN/EDIFACT
XML/EDI
ebXML



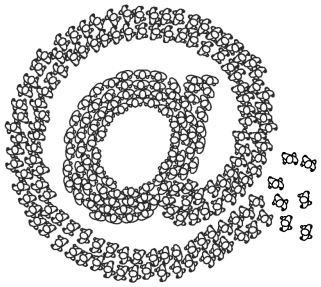
ebXML - Evolution project

- Basing on UN/CEFACT - OASIS ebXML-project
- Network Economy working group
- Evolution project
  - 12 companies from different sectors
    - Industry, trade, logistic services, consultant companies
  - Financed by Ministry for Transport and Telecommunication VALO-programme
- **Focus:** to clarify interoperability between EDI and XML solutions based on int'l standards and, disseminate information on available solutions



Evolution - why not Revolution?

- Several solutions: -Nets, X's, M's, L's
- Paradigm shift is very difficult to manage - need for step-by-step approach
- Interoperability between existing and new systems
- Business process integration with new partners
- Utilisation of existing investments



Diffuse

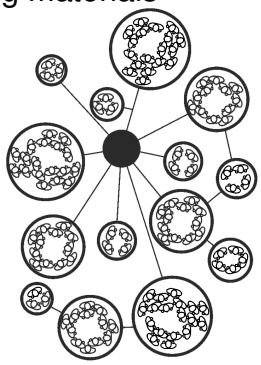
- European Union IST-programme
- Consortium: TIEKE, IC Focus and The SGML Centre

Diffuse Aims :

to provide a single, value-added, entry point to up-to- date reference and guidance information on available and emerging standards and specifications that facilitate the electronic exchange of information.

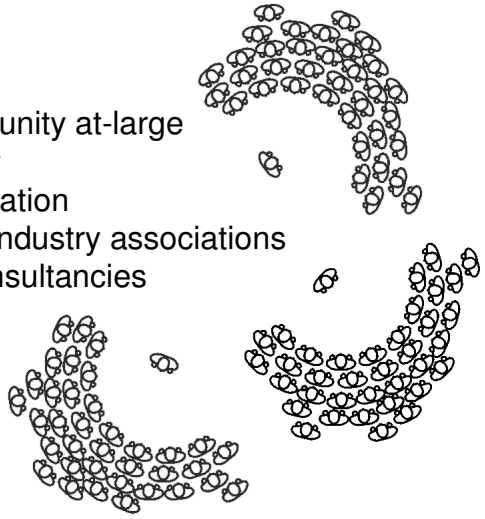
eLearning

- Harmonization of eLearning material structures on national level utilizing XML tecniques
- Description of the contents of eLearning materials utilizing structured meta data
- Interface definiton for exchange of student information between educational institutes, companies and public administration



TIEKE's interest groups

- Business community at-large
  - *ICT industry*
- Public administration
- Organizations, industry associations
- R & D units, consultancies



TIEKE in key figures

- Member organizations over 100
- Users of various services more than 1000
- Connected individuals more than 10 000
- Staff of 24 + networked contributors
- Budjet million 2 € in 2001

Thank You !

Tieke (saddleback) belongs to New Zealand´s unique wattlebird family (Callaeidae). It is a medium sized bird, and its main feature is a conspicuous chestnut-coloured saddle on its back, but it also has chestnut on the tip of its tail, a black bill, black legs, and orange "fleshy" wattles either side of its throat.

RosettaNet

Barbara Heikkinen
barbara.heikkinen@nokia.com
Nokia Research Center

XML Finland 2002
22nd of October

Contents

- What is RosettaNet?
- RosettaNet structure
- The Next Generation Architecture of RosettaNet
- Links: more information on XML and UML

- Contents of the slides are gathered from two sources
  - The External Communications set, Nokia, February 2002
  - The NextGen PIP Tutorial for Reviewers, RosettaNet, November 2001
    http://www.rosettanet.org/rosettanet/Rooms/DisplayPages/LayoutInitial?Container=com.webridge.entity.Entity%5BOID%5B7F2298D440526043A453A67B27BEF776%5D%5D

The RosettaNet Vision, Mission

Vision:
The Leader in **global** e-business **standards**

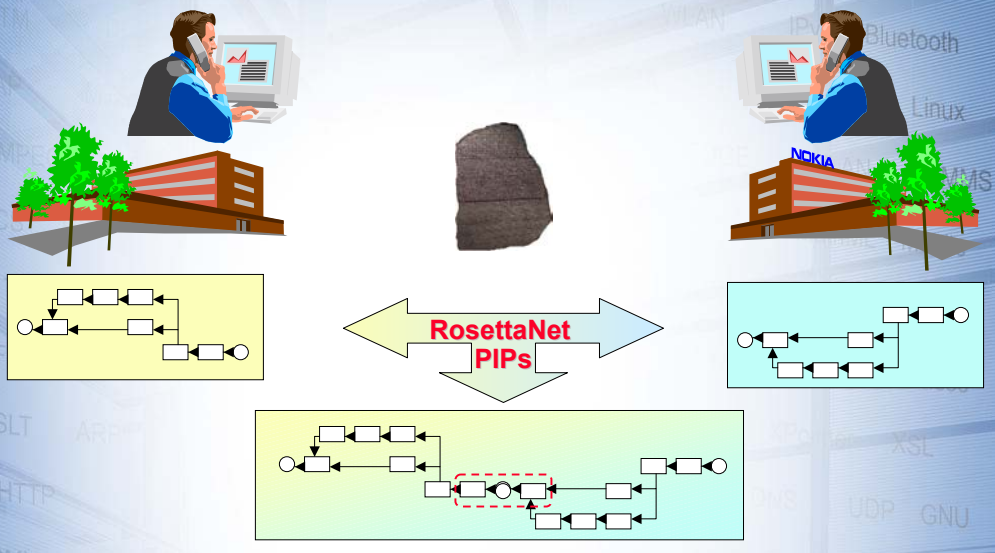
Mission:
RosettaNet drives collaborative development and **rapid** deployment of internet-based business standards,

creating a common language and **open** e-business processes that

provide measurable **benefits** and are vital to the evolution of the global, high-technology trading network.

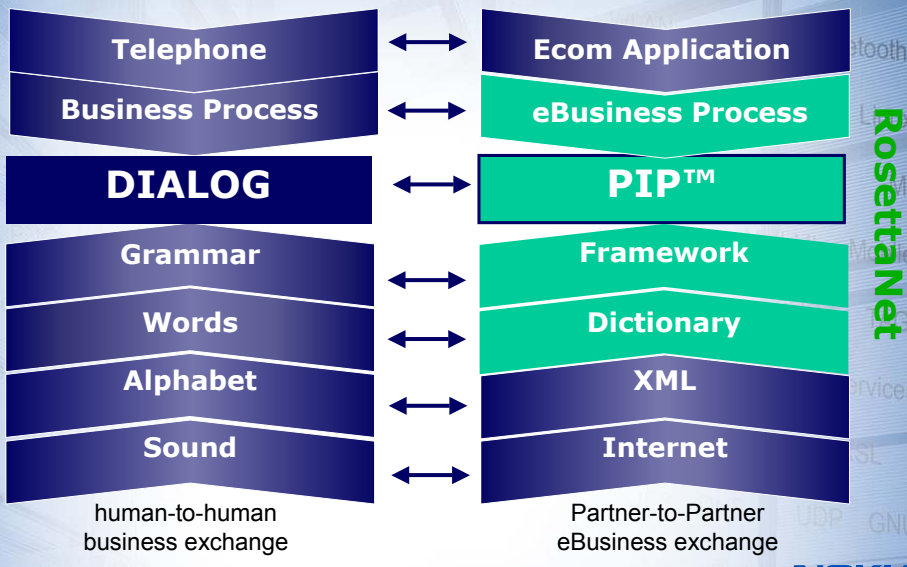
NOKIA

It's all about integrated business processes



NOKIA

E-Business Exchange



NOKIA

eBusiness Tools for *doing business with Nokia*

Exists

Electronic Data Interchange

Existing standard. Nokia recongises that EDI is needed by some partners,and will be supported accordingly

In production

Nokia Global Supply Web

NGSW is Nokia's web marketplace. Easy to set up and operate, but does not deliver a system to system connection

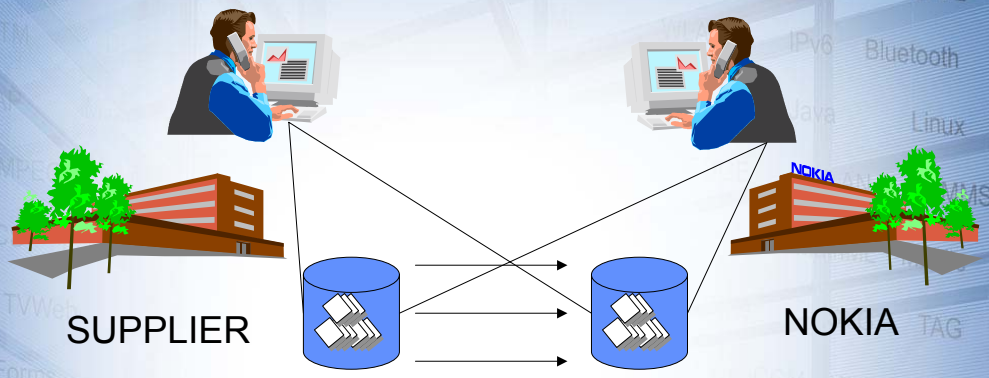
Under deployment

ROSETTANET
Lingua franca for eBusiness

Rosettanet is a non-proprietary XML protocol allowing system to system connections. Nokia is committed to integrate business process with Rosettanet standard.

NOKIA

Final Goal: Integrated Business Processes Conducted On-line



On-line Business processes between:

- Human-to-human
- System-to-human/human-to-system (WEB)
- **System-to-system (RosettaNet)**

Millions of components every hour

Purpose of RosettaNet work in Nokia

- Use RosettaNet standard for business process and system integration with external business partners
- Use RosettaNet processes and process interfaces as a guideline for Nokia internal process development and modular business platform development
- Actively participate to RosettaNet work and drive standard development to make sure it is supporting Nokia business requirements and needs
- Emphasize Nokia business partners to join RosettaNet and apply that as a standard for business process and system integration

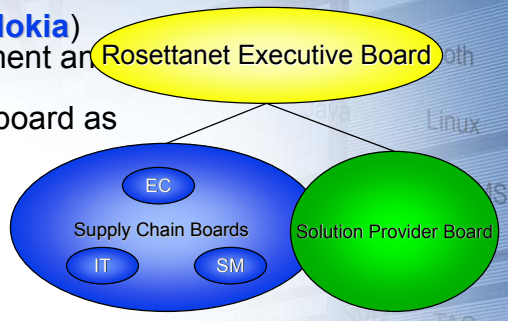
Press Release, February 26, 2002: Nokia continues to drive the implementation of RosettaNet standards

- Jean-Francois Baril, senior vice president, Sourcing & Procurement, Nokia Mobile Phones
  - In 2001, we concentrated on gaining experiences from RosettaNet with a selected number of business partners, ranging from component suppliers to customers and contact manufacturers.
  - We are convinced that RosettaNet is the right choice for our business integration, and **our target is to have 40% of Nokia's purchasing volume involved in RosettaNet interfaces already this year.**
- Currently, there are seven Partner Interfaces Processes (PIP) in use between Nokia and its business partners. They cover
  - collaborative forecasting,
  - order management,
  - transportation and distribution, and
  - finance.

NOKIA

Rosettanet structure

- Rosettanet Executive board (**Nokia**)
  - Rosettanet overall management an admin.
  - Nokia is represented in this board as well
- Solution Provider Board
  - System solution providers
  - i2, SAP, etc.
- Supply chain boards
  - EC, Electronic components (**Nokia**)
  - IT, Information Technology
  - SM, Semiconductor Manufacturing
- Role of supply chain boards
  - Drive and prioritize the standard development according to business requirements of this particular supply chain



NOKIA

The Key Elements of RosettaNet

- **PIP™s** – Partner Interface Processes – define business processes between supply-chain companies, providing the models and documents for the implementation of standards
- **Dictionaries** - provide a common set of properties for PIP™s & designates the properties used in basic business activities.
- **Product & Partner Codes** - Working in conjunction with RosettaNet dictionaries are product and partner codes, including the Data Universal Numbering System (D-U-N-S), Global Trade Item Number (GTIN) and United Nations/Standard Product and Services Code (UN/SPSC).
- **RNIF** - An open, common networked-application framework, the RosettaNet Implementation Framework (RNIF) provides common exchange protocols that enable the implementation of PIP™s.

NOKIA

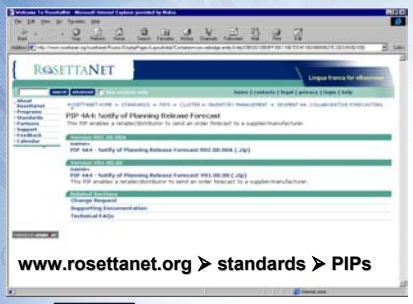
Partner Interface Process™ (PIP™)

- Depict activities, decisions and interactions that fulfill a business transaction
- Specify structure and format of business document payloads
- Organized by clusters and segments

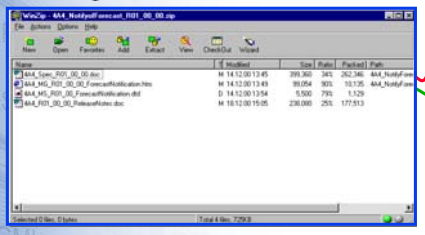
Cluster 0: RosettaNet Support	**Cluster 1:** Partner, Product and Service Review	**Cluster 2:** Product Information	**Cluster 3:** Order Management	**Cluster 4:** Inventory Management	**Cluster 5:** Marketing Information Management	**Cluster 6:** Service and Support	**Cluster 7:** Manufacturing
Segment 0A: Administrative	**Segment 1A:** Partner Review **Segment 1B:** Product and Service Review	**Segment 2A:** Preparation for Distribution **Segment 2B:** Product Change Notification **Segment 2C:** Product Design Information	**Segment 3A:** Quote and Order Entry **Segment 3B:** Transportation and Distribution **Segment 3C:** Returns and Finance **Segment 3D:** Product Configuration	**Segment 4A:** Collaborative Forecasting **Segment 4B:** Inventory Allocation **Segment 4C:** Inventory Reporting **Segment 4D:** Inventory Replenishment **Segment 4E:** Sales Reporting **Segment 4F:** Price Protection	**Segment 5A:** Lead Opportunity Management **Segment 5B:** Marketing Campaign Management **Segment 5C:** Design Win Management (EC) **Segment 5D:** Ship from Stock and Debit (EC)	**Segment 6A:** Provide and Administer Warranties, Service Packages and Contract Services **Segment 6B:** Provide and Administer Asset Management **Segment 6C:** Technical Support and Service Management	**Segment 7A:** Design Transfer **Segment 7B:** Manage Manufacturing WO & WIP **Segment 7C:** Distribute Manufacturing Information

NOKIA

ROSETTANET PIP DOCUMENTATION



PIP DOCUMENTATION TO BE DOWNLOADED



PIP SPECIFICATION

- word document
- one per PIP
- defines the process
  - players & roles
  - activities
  - messages

XML MESSAGE GUIDELINE (+DTD)

- HTML document
- one per each message
- defines the message
  - structure
  - data entities

Future: the Next Generation Architecture of RosettaNet

- Work in progress
  - First visible results probably during the next spring
- PIP® process specifications: choreography and message controls
  - To use Business Process Specification Schema (BPSS) from ebXML
- Re-architected business content
  - Standard and reusable PIP “components”
  - UML as a message design notation
- PIPs defined in machine-sensible XML Schema
  - Choreography and message controls in a BPSS document
  - Each business document defined by an XML schema

NOKIA

PIP® Business Document Specification in the NextGen Architecture

- Major change from current PIP specification
  - Design in terms of small cohesive “core objects”
  - Organize components into domain specific reusable packages
  - Build larger grained elements by composing them from core objects
  - Capture the unique requirements of business scenarios by specializing elements
- XML Schema can be generated from the UML class diagrams (e.g., via MOF and XMI)
- Motivations
  - Transition from DTDs to XML Schemas
  - Single XML Schema defines a document instead of *DTD + Message Guidelines*
  - Reparing of structural problems like inconsistency and complexity

➢ Towards applying software development methodologies for building and maintaining XML vocabularies

© Nokia

Why RosettaNet is called RosettaNet?

The name of RosettaNet originates from **The Rosetta Stone** (196 BC), which provided the key to solve the mystery of hieroglyphs. The Rosetta Stone is the only surviving fragment of a stella bearing a decree written in three languages: Greek, Demotic and Hieroglyphs.

The Rosetta Stone was probably originally erected at the temple of the city of Sais, later in the medieval period moved to Rosetta (aka el-Rashid), where it was discovered in July 1799 by Pierre François Xavier Bouchard.

The Rosetta stone is now located at British Museum in London.

NOKIA

Extensive representation from Electronics Industry

SEMICONDUCTOR SUPPLIERS
Altera, AMD, Hitachi, Intel, Lucent, Micron, Motorola, National, NEC, Philips, Samsung, ST, TI, Toshiba, Xilinx, Sony

Process integration takes two...

CUSTOMER
Agilent, IBM, CISCO, Solectron, Nokia

RosettaNet EC Managing Board

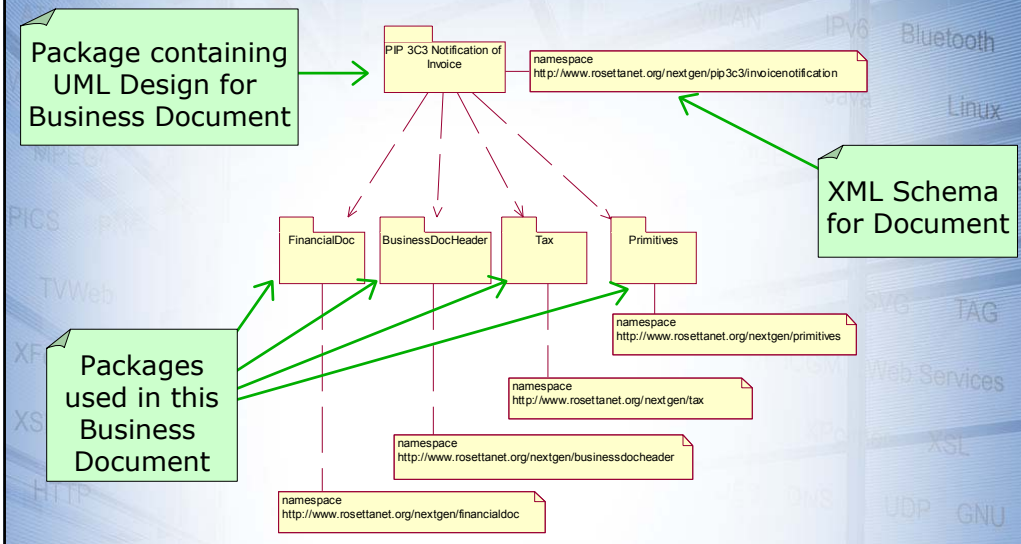
DISTRIBUTOR
Arrow, Avnet, Future, Pioneer, VEBA

CONNECTOR SUPPLIER
Tyco, FCI, Molex

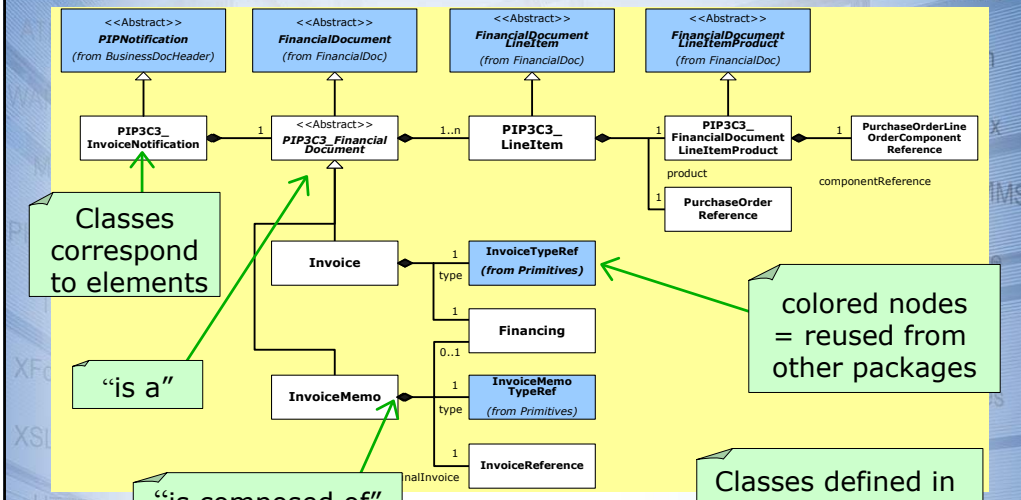
PASSIVE SUPPLIER
AVX, Bourns, Kemet

NOKIA

Example:
Reuse Structure of Business Document



Example:
Structure of Business Document



Links

- RosettaNet http://www.rosettanet.org
- ebXML, UN/CEFACT and OASIS http://www.ebxml.org

- UML (Unified Modeling Language), OMG http://www.omg.org/uml/
- UML Forum http://www.celigent.com/uml/
- Cetus Links on UML http://www.cetus-links.org/oo_uml.html

- XML Schema, W3C http://www.w3.org/XML/Schema
- XML Schemas, The XML Cover Pages http://xml.coverpages.org/schemas.html
- XML Schemas: Best Practices Homepage http://www.xfront.com/BestPracticesHomepage.html

NOKIA

More information on XML and UML

- XML Metadata Interchange (XMI), OMG http://www.omg.org/technology/documents/formal/xmi.htm
- Conceptual Modeling and Markup Languages, the XML Cover Pages http://xml.coverpages.org/conceptualModeling.html
- XML Modeling, David Carlson http://xmlmodeling.com/
  - Modeling XML Applications with UML (Book, Addison-Wesley) http://xml.coverpages.org/ni2001-04-19-a.html
  - Modeling XML Applications (Software Development Magazine) http://www.sdmagazine.com/documents/s=815/sdm0206e/
  - Modeling XML Vocabularies with UML (xml.com) http://www.xml.com/pub/a/2001/10/10/uml.html
  - Analysis and Design of XML Vocabularies with UML (Presentation) http://swradio.omg.org/workshop/proceedings/Carlson.ppt

  - Modeling the UDDI Schema with UML http://xmlmodeling.com/examples/uddi/
  - Modeling XHTML with UML http://xmlmodeling.com/examples/xhtml/

NOKIA

More information on XML and UML

- Migrating from XML DTD to XML Schema using UML, Rational
  http://www.rational.com/media/whitepapers/TP189draft.pdf
- XMI and UML combine to drive product development, IBM
  http://www-106.ibm.com/developerworks/xml/library/x-xmi/index.html
- UML and the Semantic Web, S. Cranefield, SWWS 2001
  http://www.semanticweb.org/SWWS/program/full/paper1.pdf
- Representing UML in RDF, S. Melnik
  http://www-db.stanford.edu/~melnik/rdf/uml/

- RosettaNet Next Generation Architecture
  http://www.rosettanet.org/rosettanet/Rooms/DisplayPages/LayoutInitial?Container=com.webridge.entity.Entity%5BOID%5B04E190566686B54EACA0A2E692A8A8B5%5D%5D
- ebXML Technical Architecture http://www.ebxml.org/specs/ebTA.pdf
- The eBusiness Transition Ad-Hoc Working Group (eBTWG), UN/CEFACT
  http://www.ebtwg.org/projects/
- OASIS Universal Business Language (UBL) TC
  http://www.oasis-open.org/committees/ubl/

Mikään ei ole muuttunut mutta kaikki on toisin

XML Finland 2002
Marina Congress Center
21.-22.10.2002
Martti Poutanen

© Index Information Technologies Oy

Index IT Oy

- XML/SGML järjestelmiin erikoistunut yritys
  - konsultointi ja järjestelmien rakentaminen
  - tuotekehitys-, tuki- ja koulutustoiminnot
    - Indox julkaisujärjestelmä
  - dokumenttien hallinta (www)
  - vankat kumppanit, X-Hive, Arbortext, i4i, OpenText
  - ratkaisut perustuvat standardeille
- kokemus
  - vuodesta 1994, yli 35 miestyövuoden kokemus vaativista rakenteisista järjestelmistä
  - useita referenssejä (> 30 tuotantojärjestelmää)
  - Talentum, SFS, WSOY,valtionhallinto
  - Finnair, Vaisala, Tamrock Stora Enso, Lääkelaitos

© Index Information Technologies Oy

XML on käytössä monessa

- viestinvälitys, metadata, Web services
  - SOAP, RDF, UDDI, WSDL..
- XHTML
  - vihdoinkin tolkkua HTML-murteisiin
- dokumentaatio
  - rakenteiset dokumentit
    - suuret tietomassat
    - kohdennettu tietosisältö
    - tuki älykkäille hakutoiminnoille
    - ulkoasu ja sisältö erotettu toisistaan
- vaarana hype-ilmiö
  - SGML lupasi samaa 10v sitten

© Index Information Technologies Oy

Standardin laajennukset

- SGML ei näissä oikein onnistunut
  - monimutkaisuus
    - ESIS,DSSSL, HyTime, FOSI..
  - kaupalliset välineet eivät tukeneet riittävän hyvin -> sovelluskohtaisia ratkaisuja
- XML näyttäisi onnistuvan paremmin
  - DOM
  - XSLT
  - Xlink, Xquery, Xupdate
  - XSL-FO

© Index Information Technologies Oy

Rakenteinen teksti

- SGML vs. XML
  - kieliopillinen ero melko pieni
    - sovellusten kannalta ero on suuri
  - well-formed vs. validoituva
    - DTD:n tarve
  - erikoismerkit
    - entiteetit vs. Unicode

© Index Information Technologies Oy

Dokumenttituotannon vaatimukset

- monimuotoinen jakelu
  - aikoinaan vain paperi
  - www, eri päätelaitteet, PDF, CD-ROM...
- kohdennettu jakelu
  - käyttäjän roolin perusteella
  - dokumentin roolin perusteella
  - sisällön luokituksen perusteella
    - “personoitu” tuotedokumentaatio
- eri sisällöntuotantotavat
  - editorit, selain-pohjainen tuotanto..
- versio/muutoshallinta

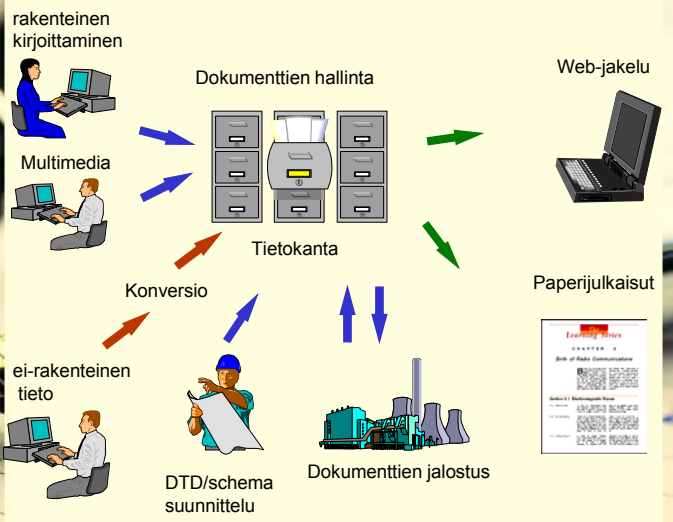
© Index Information Technologies Oy

Miksi rakenteinen dokumentaatio

- ulkoasusta riippumaton sisällön tuotanto
- dokumentaation looginen yhdenmukaisuus
  - taustalla DTD/schema
- loppudokumentaation koostaminen
  - tuoterakenteen mukaan
  - loppukäyttäjän mukaan
- monimuotoinen jakelu
  - HTML, PDF, PDA, A4....
- sisällön siirrettävyys sovellusten välillä
- älykkäät haut

© Index Information Technologies Oy

Kirjoita kerran, jakele monessa muodossa



© Index Information Technologies Oy

Rakenteinen editointi

- uusi teknologia tuo uusia mahdollisuuksia
  - selaimeen upotetut editorit
    - XML-Spy, XMetal
    - lomake-editorit
- perinteiset rakenteiset editorit
  - siirtyminen tukemaan SGML/XML-editointia
  - FrameMaker 7, Epic, SoftQuad
  - formatterit FOSI -> XSL-FO
- laajennukset tekstureihin
  - i4i S4 Text
- paljon uusia sovelluksia
  - mutta periaatteet eivät ole muuttuneet

© Index Information Technologies Oy

Puukäsittely vs. tapahtumakäsittely

- olennainen toiminto rakenteisessa dokumentaatiossa
  - tapahtumapohjainen (SAX)
  - puukäsittelijä (DOM)
- tapahtumakäsittelijä ei aina riitä
  - monimutkaiset dokumenttikoostajat
- XSLT ei vastaa “vanhoja” puukäsittelijöitä
  - esim. Balisen puukäsittelijä
  - suorituskyky
  - ei “todellinen” ohjelmointikieli
    - “ugly difficult language”
  - sopii hyvin esim. XML -> HTML konversioon

© Index Information Technologies Oy

Julkaiseminen ja haut

- suuret tietomassat vaativat hyvän hakumoottorin
  - webin hakukone-efekti
  - semanttinen haku
    - rakenteeseen sidottu
  - boolen haut
  - sana- ja lausehaku
  - jokerihaut
  - synonyymi- ja termihaku
  - taivutusmuodot (varsinkin suomen kielen)

© Index Information Technologies Oy

Esimerkki hakumoottorista

- Indox julkaisujärjestelmän hakumoottori
  - ytimenä Lucene
    - tehokas full-text indeksi
    - myös ei-XML datalle
  - synkronoitu jakelukannan muutoksiin
    - inkrementaalinen indeksointi
  - konfiguroitava indekseri
    - stop-listat
    - purettavat rakenteet
    - tyhjät elementit
  - haun tulokset Xpointer-listana tietokannan objekteihin

© Index Information Technologies Oy

Tietokannoista

- ”natiivi” XML-tietokannat
  - Xhive, Ipedo, DynaBase..
- relaatiotietokantavalmistajien XML-laajennukset
  - kyllä, mutta..
    - XML <-> taulurakenne
      - DTD muutokset jäykkiä
    - suorituskyky kuitenkin ok
- muutos 1. sukupolven tietokantoihin
  - ei välttämättä elementtitason kontrollia
  - API tehostunut
    - Xquery, Xupdate
  - huomattava nopeuden kasvu

© Index Information Technologies Oy

Case Indox

- dynaaminen tietokantajulkaiseminen
  - XSLT:n käyttö
    - sisällön formatointi
  - Xquery:n hyödyntäminen julkaisemisessa
  - Xlink:in käyttö
    - julkaisun koostamisessa
    - sisällön hallinnassa
  - hakumoottorin integrointi

© Index Information Technologies Oy

XSL-FO

- taittoteknisesti kattava (standardin kannalta)
  - saatavilla olevat sovellukset eri asia..
    - vajavainen tuki
    - prosessointi vaatii paljon tehoa ja muistia
- soveltuu erinomaisesti
  - raskaaseen automaattitaittoon
  - ihmisen on vaikea puuttua lopputulokseen
    - interaktiivinen XSL-FO taitto-ohjelma
  - dynaaminen PDF-tuotanto
    - www-sivu -> tulostusversio

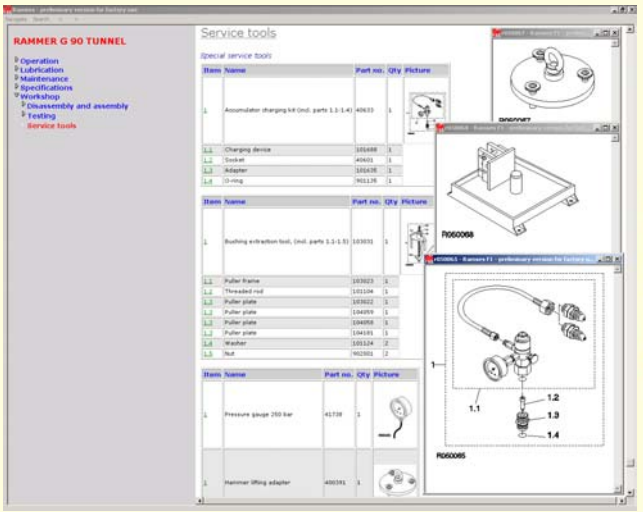
© Index Information Technologies Oy

Case Sandvik

- tekninen tuotedokumentaatio (Rammer)
  - hydr. vasara- ja murskaintuotteet
  - yli 400 erilaista loppudokumenttia
    - koostetaan automaattisesti
      - Balise puukäsittelijä
  - paperi ja CD-ROM jakelu
    - DynaText SGML-selain (1. sukupolvi)
    - IE-pohjainen (2. sukupolvi)
- “master”-aineiston hallinta
  - rakenteinen editointi (FrameMaker)

© Index Information Technologies Oy

Case Sandvik



© Index Information Technologies Oy

Case Vaisala

- ympäristön mittaustuotteiden dokumentaatio
  - monimutkainen tuoterakenne
    - päätuotteita, optiota
    - tarve erityyppisiin loppudokumentteihin
- alunperin SGML-pohjainen
- nykyinen XML-pohjainen
  - FrameMaker 7
  - tuotekonfiguraatiopohjainen GUI
    - Tekes/Tytti (Helsingin YO)
    - syöttää tiedot dokumenttikoostajalle

© Index Information Technologies Oy

Case Vaisala

- dokumenttikoostaja
  - Java/SAX/XSLT-pohjainen dokumenttikoostaja
    - SAX-käsittelijä (tapahtumapohjainen) yhdistelee ulkoisia dokumentteja XSLT-prosessoinnin kautta
    - rakenne- ja attribuuttipohjainen filtteri
  - yhdistelee ja ”lomittaa” kompleksisia tuotedokumentteja

© Index Information Technologies Oy

Case Vaisala

Kokemuksia XSLT:n käytöstä

- saa helposti aikaan hitaita sovelluksia
  - > 10 Mt:n XML tiedosto, n. 30h prosessointia..
  - oikein käytettynä kuitenkin varsin nopea
    - suuria eroja eri ohjelmien ja alustojen välillä
      - JVM Java toteutuksissa
- kielen rakenteen puutteet
  - monimutkaiset puukäsittelyt tehtävä perättäisajoina
    - tulospuuta ei voi käsitellä uudestaan
    - kömpelö muuttujakäsittely

© Index Information Technologies Oy

Kokemuksia XSLT:n käytöstä

- vaatii paljon muistia
  - varsinkin rekursiiviset käsittelyt
  - pahimmillaan satoja megoja
- ja paljon tekstiä (“pitkä kielioppi”)
- Java-laajennukset hyödyllisiä

© Index Information Technologies Oy

Native XML Database in Web Applications

Case: Done Xenus + Done Helpdesk

Jussi Volanen
Done Information, Inc

27/09/2002

Topics

- Theory
  - definition of Native XML database
- Xenus
  - What it is
  - Benefits
  - architecture
- Done Helpdesk
  - What it is
  - Application object mode
  - ER Model

What is native XML database?

Definition [http://www.xmldb.org/]

- Defines a (logical) model for an XML document and stores and retrieves documents according to that model.
- Has an XML document as its fundamental unit of (logical) storage.
- Is not required to have any particular underlying physical storage model.

What is Xenus?

- A native XML database using a relational database as storage system
- Text-based, XML stored as text in CLOB
- Database independent
  - SQL Server
  - Oracle
  - Access
- Combines benefits of both worlds
  - Possible to use XML and relational tables at the same time
  - Relations and constraints by RDBMS
  - ”Standardized” query language (SQL)
  - Loose coupling between application and database

Benefits of using Xenus

- Faster development
  - Generic Data Access layer implemented in Xenus
- Better code quality
  - Relations and constraints managed by database
  - Data Access layer tested
- Easier to make changes in data structure
  - Xenus database and application are loosely coupled
- Better performace
  - No mapping between the database and the application
  - Parsing the XML stream inside MSXML parser component

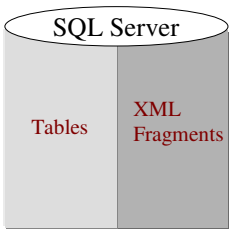
Why did we implement Xenus?

- It was fun and challenging
- Microsoft databases don't have support for native XML
- XML forms requires XML data sources
  - A relational database with XML mapping or a native XML database
- Different mapping techniques were found inefficient and hard to implement
- Applications with a server (SQL Server / Oracle) and a number of offline clients (MS Access) needed an abstraction layer to remove database differences
- Offline clients must be as cost-effective as possible

Why did we implement Xenus?

Experiences in previous projects

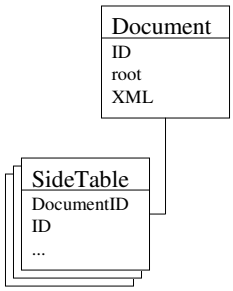
- A web application with XML forms with two approaches in database:
  - Table based mapping. Deep structure stored in tables
    - Not flexible
  - Tree of XML fragments in relational table
    - XML in one column
    - Impossible to use constraints
    - No search
    - Flexible, extendible



DONE

Xenus

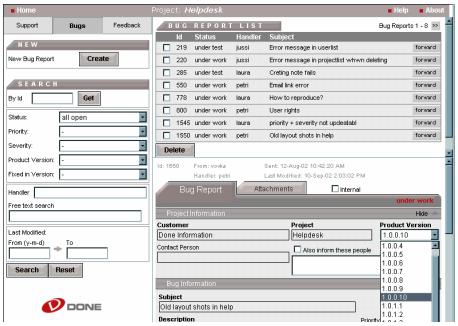
- XML in one column
- several indexing types
  - Schema, (XSLT, Oracle,...)
- Indexes stored in side tables
  - Side tables and relations between them are defined in schema
  - Updated when the document is changed
  - Partial Object relational mapping
- Combining application object model (UML), XML and relational database



DONE

Done Helpdesk

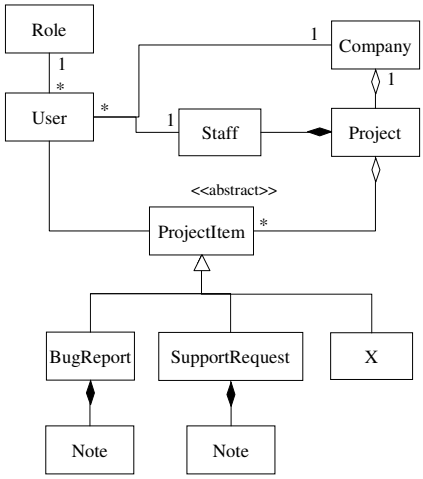
- Application for our support
  - Support requests
  - Bug reports
  - Feedback
  - FAQ
  - ...
- Offered as application service or installed for a customer



Done Helpdesk

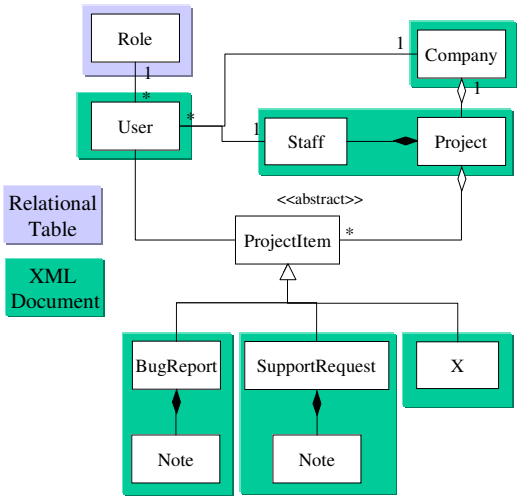
Application object model

- Company
  - Basic information about a company
- Project
  - Basic information
  - List of persons in the project
- Support Request
  - Notes
- Bug Report
  - Notes
- User
  - login, password, group, role



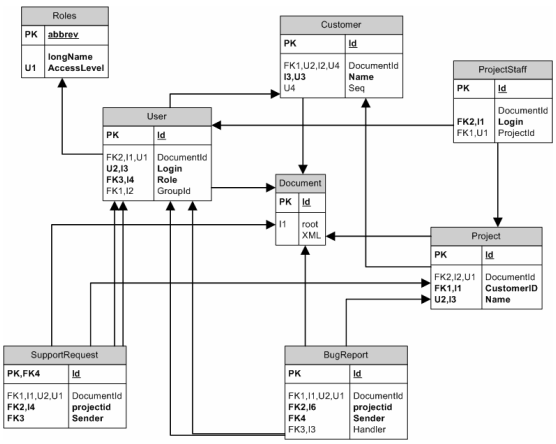
Application object model

```
<Xenus>
<Table name="User"/>
<Table name="Company"/>
<Table name="Project">
   <Table name="Staff">
      <Column name="userName">
         <Ref name="User.login"/>
      </Column>
   </Table>
<!– Note is not indexed -->
<Table name="BugReport"/>
<!– Note is not indexed -->
<Table name="SupportRequest"/>
</Xenus>
```



ER Model

- Document table for data
  - DocumentID FK in side tables
- No table for Note object



Summary

- Xenus
  - A Text-based Native XML database
  - Relational database as data storage
  - Combines relational and XML world
  - Makes it easier to implement web applications
- Done Helpdesk
  - Application for our support organization
  - Relations between XML documents
  - Relations between XML documents and relational tables

Questions?

Architecturing Component Based Systems with XML Technologies and Standards

Jens Jakob Andersen, The Danish Post

tsw@zorck.dk

Abstract.

This paper details how XML technologies and standards can be used to design and implement Enterprise IT systems, in such a way that these are easier to build, maintain and integrate. For many years there has been a nirvana of buzzwords in this arena, DCOM, COM, CORBA, J2EE, RPC etc., but with the emergence of especially web services as well as RDF, we now have some quite useful tools at our hands. The paper is built on experiences gained trough my work of many years of design and development of web architectures and internet systems.

Biography:

Jens Jakob Andersen, a leading Enterprise IT System Architect at The Danish Post, is an experienced Internet pioneer, with many years of design and development experience in Web architectures and Internet systems. He is also an elected member of the board of the Danish XML Forum, where he develops seminars, and participates in the work of evangelizing XML in Denmark.

Holy grail of (Enterprise) software system design – modules (analysis, design, maintain, integrate, document etc.) – realized as selfdeclarative loosely coupled services

Basically software is quite simple. At the core of all software are 3 elements, repeated again and again: fetch input data, perform calculation, store result. Enterprise software development is quite a bit harder. If only enterprises could run on a single piece of simple software it wouldn't be so. But they cannot. So we end up with 100+ systems build over the years with an extremely high level of complexity. And the added burden of business demanding reusability, stability, futureability, integrateability, securability, etc.... And all within an acceptable business budget and timeframe. While still keeping the legacy portfolio running.

We have seen many good intentions come, hype, linger and leave a trail of smudged IT-systems, just like a fire that has raged and finally burnt out, leaving a distinct smell in the air for days.

The hypothesis that I am chasing in this paper, is that XML technologies and standards are more than a fading hype, and being used with caution, can give us some actual longlifed benefits in designing and implementing Enterprise IT systems which are easier to build, maintain, integrate and maybe even also easier to outphase. This by enabling selfdecribed loosely coupled systems to interact, not just web-services, but also components, data and software.

I will take a pragmatic approach to the theory, since I do not want to disappear into the AI jungle, where many will be lost in darkness and not found until the next Henry Morton Stanley passes by with a new millennium version of "Dr. Livingstone, I presume?"

Disclaimer: This paper is not a technical tutorial, but hopefully great inspiration

Alphabet soup – DCOM, COM, CORBA, J2EE, RPC, COBOL Subprogram etc.

YAFLA's[1] has been the trademark of the IT industry for decades. Many of these have over the last decade been in the component area. Competing models from different vendors and consortiums (remember OSF?) have tried to claim the new software development arena. But none has succeeded yet, since none have really offered the right solution. COM is a great component model for sharing binary libraries between applications on the Windows platform, (as well as more or less the heartblood of Windows) but it doesn't commute over the net. EJB's are a great container model for sharing Java components in Java-land, but doesn't talk so well to e.g. Cobol sub-programs. RPC is great for communicating across the wire, so is JMI and even XML-RPC & SOAP can do the talking and walking for us.

But all these initiatives lag one big thing: They are nothing but tools to help developers code their applications, and distribute the application modules across more files and platforms. Component models are great as a tool to create reusable building

[11] Yet Another Few Letter Acronym

blocks, but they do not offer a framework where they can be used from a semantic and declarative viewpoint.

Xtremely siMple soLution – XML?

For me the most interesting single element of XML has nothing to do with standards, syntax, technology or workgroups. So far I have not met anyone outside of the marketing/sales force gang, who thought that XML was neither simple nor easy, but definitely intellectually stimulating.

But what I see that XML is spawning, is a new breed of projects, because marketing/sales guys using Xtreme Marketing Language trigger managers to think: Hey – XML is Internet, and the late nineties taught me that anything "Internet" is easy, cheap and fast, so lets get going on this SOAP/B2B/A2A/sem.web/ontology XML thingy. (Please don't enlighten them; it will take all the fun away from us engineers and architects for the next 3-5 years, if they find out already now what the harsh realities are.)

But jokes aside, XML brings a lightness to the table, which turns projects that earlier looked frightening into something which looks like it can be fixed in an afternoons work. And thus we get all these new projects attacking "heavy" problems.

The heaviest problem

I have a vision. If we could begin developing software in a way, so that:

- It would be very fast to develop new functionality
- It would perform fast and flawless
- It would be easy to extend and upgrade applications
- It would harvest the best ideas from the n-tier model and distributed computing, making it easy to scale up and out
- It would be easy to run by the operation guys
- It would be easy to integrate to on all levels

Then organizations could begin to save huge amounts of resources, as well as reaching business goals much faster.

Match this vision with the idea that some (and hopefully a lot of it over time) of all these saved resources will inevitable be channeled into projects that will:

- Do something positive about the environmental issues
- Do something positive about world poverty
- Do something positive about world peace and stability

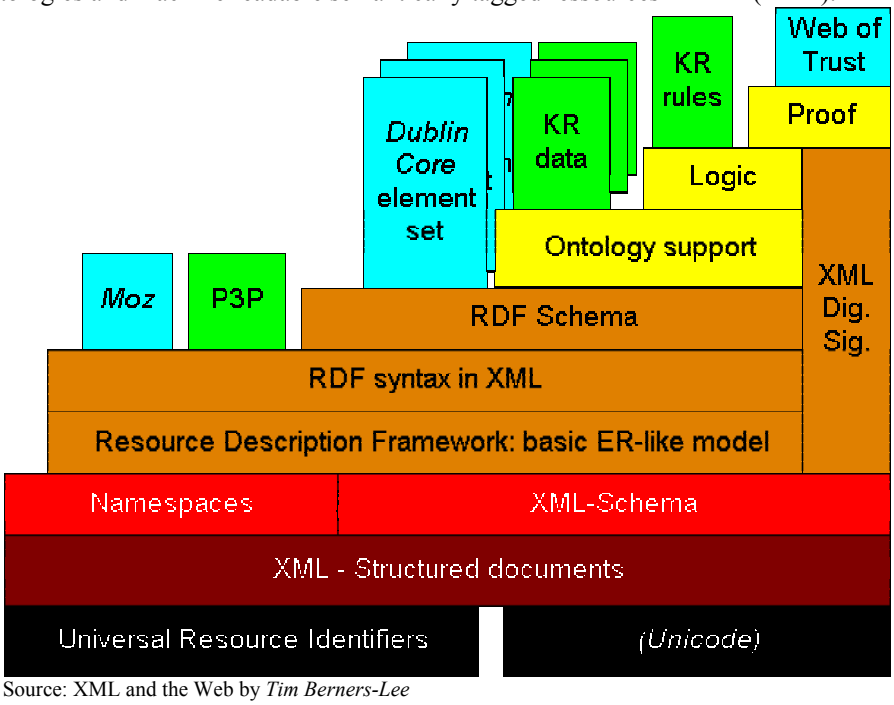
the vision suddenly becomes very exiting.

Web services – what is it and where does it fit in

Web-services are in my opinion a part of the new vision – if the Internet is the basic infrastructure, web-services envelopes and web-servers the mailboxes. I see web-services as the first step into a word of loosely coupled systems, talking across the Internet wire. On the other side, I do not think that we in any near future will see systems looking up new services via UDDI at runtime, and connect to these.

Web services – what are the missing links?

For a meaningful converstaion, we need meaningful content, in a way so the content is useable. Enter Semantic.Web theory. The model below is described in detail in the webcast found at *http://technetcast.ddj.com/tnc_play_stream.html?stream_id=459*. The short description of the model (the Semantic web "stack") is that sem.web eventually will deliver a web of Trust, based on proof, digital signatures, logical reasoning, ontologies and machine readable semantically tagged ressources in RDF (XML).



Source: XML and the Web by *Tim Berners-Lee*

In this paper the interesting thing about this sem-web roadmap, is that web-services are not even mentioned therein. Exit SOAP, and let us leave the web.services near the YAFLA playground.

RDF – Real meDicine Formula?

I see that RDF is a step on the way. It could also be ISO's Topic Maps that was the stepping stone. What RDF and Topic Maps are offering, is a model for describing relations (or associations) between properties and entities with values. Or you could view RDF as a Entity-Relation (ER) model for the web, with URI's being the tables and RDF's being the relations. Unfortunately, RDF is not enough in itself. It is great for adding descriptions to resources, but this is not creating value, only more data.

We can use RDF to describe our components and web-services, and if we mix in ontologies for adding correlative meaning to the RDF's, we have one of the very important building block of the solution- platform independent service descriptions, that are machine-readable.

There is a need for 2 levels of descriptions. One level where the individual IT-elements are describeb, such as field, table, database, funktion, parameters etc. This ontology can be shared across many different ortganizations.

The second level of description is needed for the specific organization. This will describe the specific elements suchs as customer, employee, process etc.

AI – Ages of Impires?

On top of RDF and ontologies, the sem.web needs a standard for describing logical rules, as well as some kind of AI (artificial intelligence) engine to make meaning out of it all. Maybe not a full-blown self-reasoning neural network, but at least some kind of logic engine that can process input according to well-defined rules, and generate documented decisions. I see that the sem.web vision has spurred a new wave of activity in the AI field, and maybe this time around AI is here to stay on a wide scale. What I still need to see is an AI engine which is very efficient, yet simple enough to configure for mere mortals to use.

The theory – A semantic component paradigm for software development

If we forget the Internet for a while, since the Internet is an uncontrolled jungle, and zoom into the Intranet level, something very interesting happens: we are in control of the environment.

This means that we can more easily define ontologies describing the data and services in our local business and computing environment, as well as the business-rules. This will enable a computing environment where we can build systems by weaving individual web-services together on basis of logical rules, when we get an AI (logical rule-engine) based middleware layer, and standards in this area.

After having moved over to logic-rule based coupling of web-services, the next step will be to zoom one step closer, and apply RDF and ontologies to the internals of the individual components (e.g. EJB's), and deliver the functionality inside the individual web-services by connecting the components with AI based execution platforms.

Logical missing link rules

This is an area where there still needs to emerge common standards. There are many initiatives, such as RuleML, WSFL, XLANG and the latest - BPEL4WS, Business Process Execution Language for Web Services, which is a joint-venture between IBM, BEA and Microsoft and is aimed at replacing WSFL and XLANG.

The initiatives can be divided into 2 classes:

- Markup languages for sharing "execution plans" between web-service orchestrators
- Markup languages for defining logical rules to be evaluated by AI-engines

A pragmatic approach for the near future

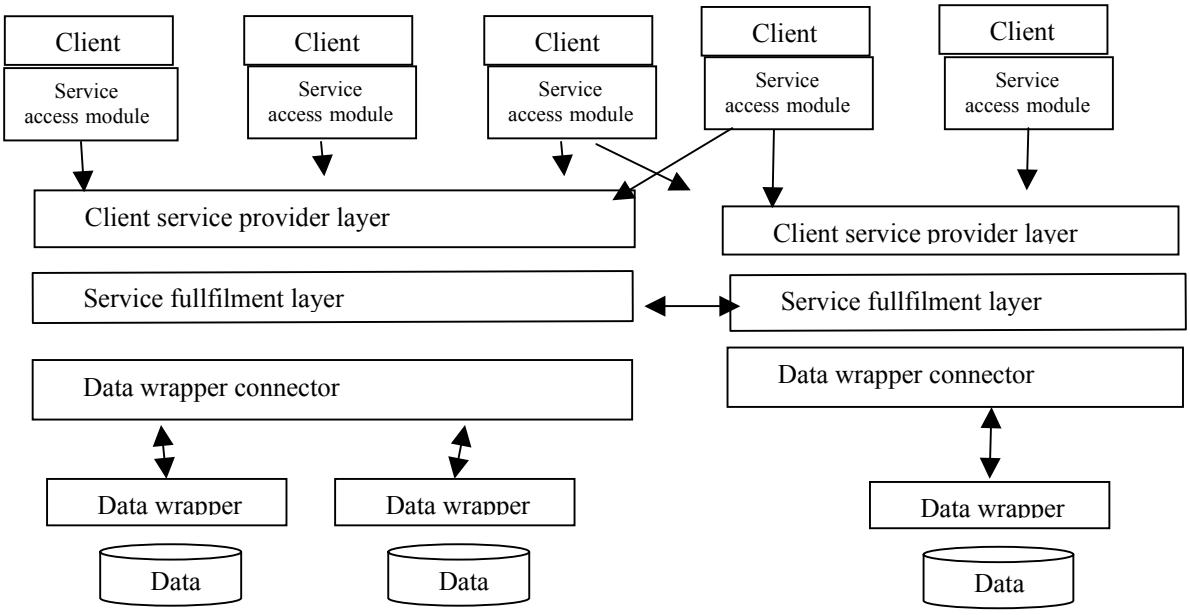
I think that Tim Berners-Lee is right in defining the semantic web as being the web weaved by machines, made understandable and meaningful for machines. But it is close to taking the lid off the Artificial Intelligence Pandora Box, and sends us plummeting into the abyss of defining how we ever would make humans understood by machines. *There is a theory, which states that if ever anyone discovers exactly what the universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another which states that this has already happened.* (Douglas Adams: **The Restaurant at the End of the Universe**)

The Semantic Intranet version 0.5 that I think can come to us much faster, is a sem.intranet visualized and navigated by people, thus saving us a lot of trouble in defining cross-ontologies, training robots, inferring from logic etc., while still gaining the benefit of working with semantic-enabled computer systems to the end-users.

Imagine working as a user on a semantically enabled system. Instead of being stuck with rigid inflexible data and fixed menu structures, you will have the opportunity to work with logical associations between elements. Or imagine to be a systems administrator configuring new service modules using components described with RDF and a standard for describing flow and logic, to tie components and services together. I call this vision "Associative Declarative Computing" – ADC.

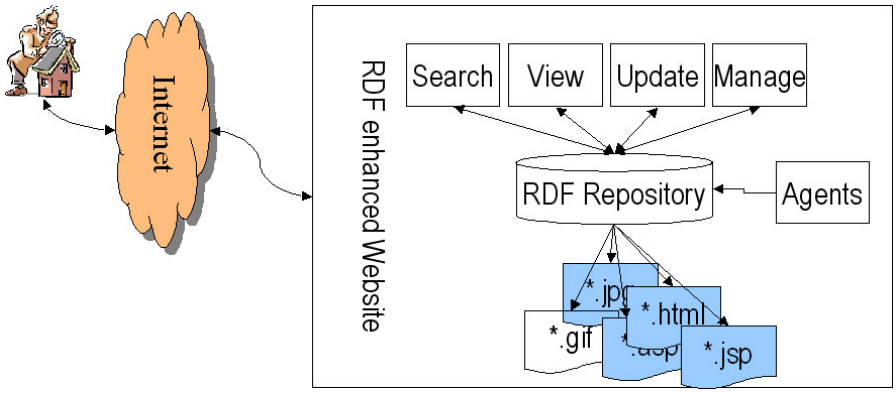
Some illustrated examples

Component model based architecture (or maybe even web-service architecture)



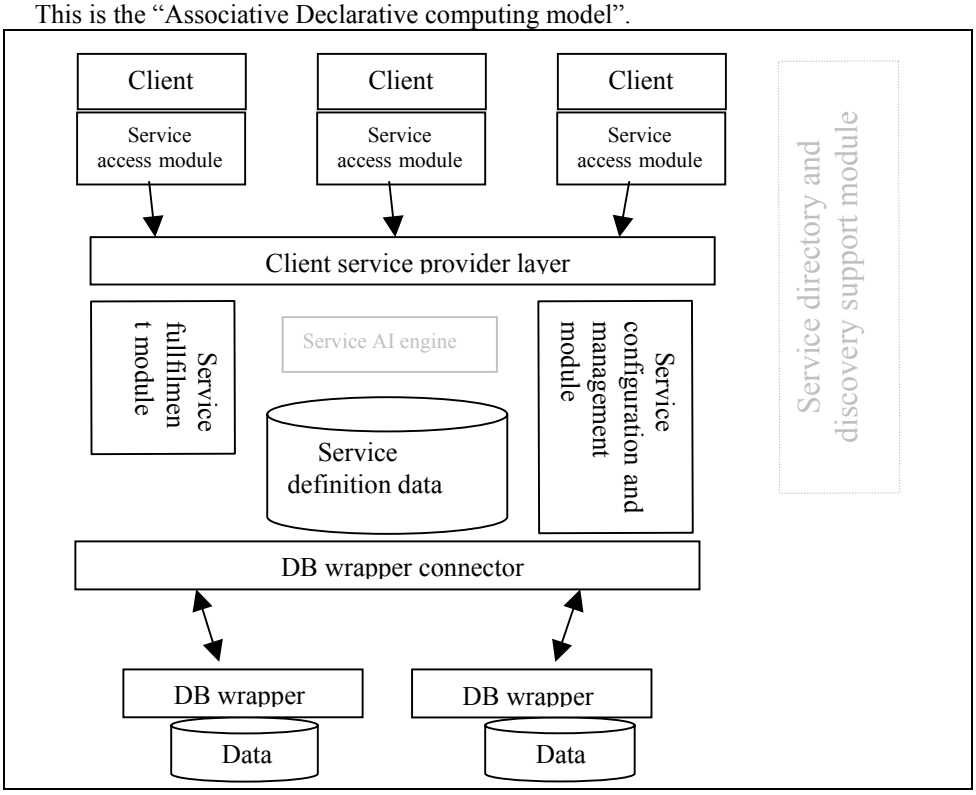
This is the current best-practice model, the n-tier model. Data is decoupled from fulfillment (business logic) layer, which again is decoupled from the client layer. Clients access the service fulfillment layer via a well defined API (format and protocol). Data wrappers are used for information hiding. In theory this models enables us to change service implementation as we like, since we do not compile the actual binary modules into the client software. But the use of components and even web-services doesn't solve all the issues – it can still be deemed a closely coupled system. And even worse – it is only at the infrastructure level that we are given "components" and "layers", very often I have seen logical monoliths (or stove-pipes) build on expensive n-tier infrastructures. When the connections are hard coded and determined at build-time instead of being inferred at runtime on the basis of logical rule processing, , and none or very few of the services and modules are used by more than 1 application, the model usually ends up with something that is quite inflexible to work with and generally a waste of time, money and other resources.

Last year at XML Finland 2001, I presented what I called a "Semantic web in a box" blueprint.



Where the sem.web in a box was aimed at building a RDF-enabled web-presence around the RDF-search concept, for users to access over the web, the new model aims at using sem.web to enhance as well web-services as the inner logic, and build solutions.

This is the "Associative Declarative computing model".



This model introduces a "declarative" service layer, where the inner workings of the service, as well as the access to backend/legacy systems are not hard coded, but defined with "Service Definition Data". The basic idea is to use W3C standards to define the available content classes and actions that can be applied to them. And leave the AI module out for a start.

It is the pragmatic approach, that will bring us closer to RDF tagging of multiple resources, as well as logic rule sets defined in machine usable form (such as for instance Ruleml), but with the use of a human administrator to make the intelligent matching process, and feed this back into to the model for use at runtime, so we have a model where we can reap the benefits of semantically described services.

It will give us the platform for loosely coupled services and components, without the initial burden of choosing a rule markup standard.

Since systems are not compiled, but defined, there will be no more syntax errors and hidden bugs in the code. There will only be the logical behavior that is defined in the rule markup as source of potential wrong results. And then these will not be blamed on the software, but on the defined logical rules.

Additional benefits of this model will be that disciplines such as inventory management, deployment management and asset management will be easier to automate, since the RDF description are ready and available for reuse.

Next Step – 4GL sem.web

I believe that we stand with all the pieces of the puzzle in our hands, and just need the catalyst that will start the reaction, and create new tools. In the words of William Gibson: The future is here. It's just not widely distributed yet.

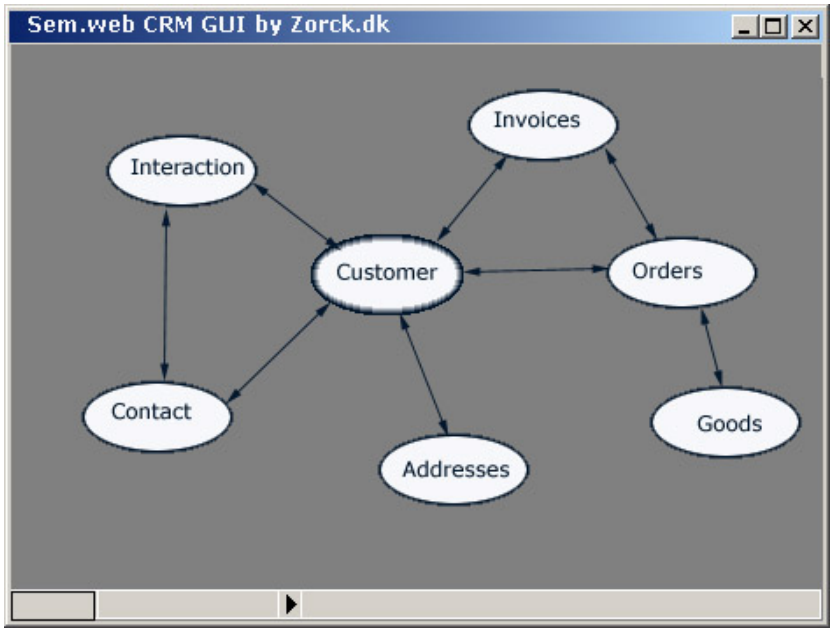
When the Semantic web really takes off, one of the areas that I look forward to see development in, is a new generation of 4GL tools, specifically designed to build associative declarative systems.

For decades now all software has evolved around the basic elements of data and commands. First it was pure command line systems, then came menu-based software. And we are still in menu based software, although it has moved from pure text-based menus (think 3270, Unix, DOS etc) to graphical menus, such as Mac and Windows. It is still the menu pattern; pick some action element and execute it.

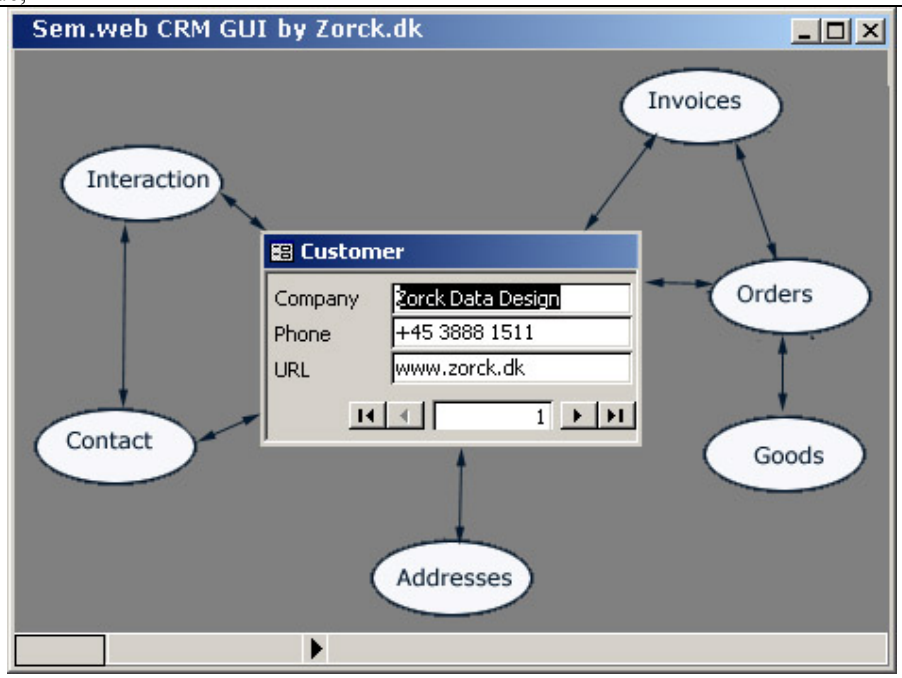
So users are still burdened with the task of learning the semantics of a lot of different arcane menus and commands, as well as remembering which program to start to do what: Start the invoicing module if you need to do invoicing, start the calendar if you need to book a meeting, start the budgeting module if you need to do budgeting.

I envision a 4GL tool build on top of the Semantic Web, where the information found in RDF and ontologies will be used as the foundation of application models. This will facilitate GUI's where user will navigate between associated elements, such as customers, contacts, meetings, orders etc. as exemplified in the diagram below, and act on these. Data elements and their associations become the major GUI paradigm, not the menus and commands.

The information model will become the virtual architecture of all systems.



Sample "sem.web" GUI. Resources and their relationship are shown in the "navigation GUI". The user can double-click on any element (as well as connections) to toggle from model to list/detail view, or entirely switch over to "listview/detailview" mode,



If you come from the RDBMS/SQL world, you might begin to think "Hey, this is basically an ER-model based GUI or even EER", and you've got a point. The difference I see is that the ER-model is very close to the RDBMS, whereas the sem.web model is a purely artificial model of relationships, where some might even be indirectly determined by ontologies behind the actual RDF documents.

Concluding remarks

I believe that the semantic web project is here to stay. And that it will succeed on a global scale in a timeframe of 10-15 years.

I also believe that we need to find some intermediate pragmatic steps, which will utilize the basic semantic web standards, and generate business value in "normal" computing fields, such as software development and Enterprise IT systems.

In this paper I have outlined 2 such areas, where a pragmatic approach to RDF , Semantic web and eventually logic markup standards can add huge benefits without too much work, and thus help paving the way in Enterprise computing.

One area is "component based computing build on sem.web enabled middleware" and the other area is "sem.web GUI for user applications".

As long as we stay inside the Intranet boundaries, we can leave out digital signatures and trust – some of the difficult parts.

And as long as we use human administrators to configure the "lines between the dots", instead of using inference engines to do the logical reasoning, we can get acceptable logic and proof, without entering the AI fields.

Additional work needs to be done in the area of defining standards for mapping Enterprise IT ressources to the sem.web model, as well in refining the vision of using these models to architecture complete Enterprise IT systems.

We need to follow the development in web-services and web-service orchestration, as well as development in knowledge represenatation standards, and map these areas together to form the foundation of Enterprise IT systems.

Link section:

http://www.daml.org/2001/06/swday-ontologies/Ontologies-talk-060401_files/frame.htm - Ontoliges: What they are and why you should care; A nice primer on ontologies

http://conferences.oreillynet.com/etcon2002/ - O'Reilly Emerging Technology Conference

http://iswc2002.semanticweb.org/ - First International Semantic Web Conference

http://www.zorck.dk/tsw - The Semantic Web at Zork.dk

http://www.gca.org/papers/xmleurope2000/papers/s11-01.html - The TAO of Topic Maps, Steve Peppers great introduction to ISO Topic Maps

http://technetcast.ddj.com/tnc_catalog.html?item_id=1085 - Tim Berners Les web cast available from Dr. Dobb

http://www.brunel.ac.uk/~emstjja2/ - Great mathematical theories for state estimation, Constraint logic programming and network estimation. One of the sharpest minds I've met.

http://logic.stanford.edu/selt/selt.html- Stanford Encyclopedia of Logic Technology

http://www.daml.org/ - The DARPA Agent Markup Language home

http://www.dfki.uni-kl.de/ruleml/- RuleML home

http://www.cordis.lu/ist/ka3/iaf/index.htm - Information Society Technologies Programme (EU program)

http://www-2.cs.cmu.edu/~softagents/daml_Mmaker/daml-s_matchmaker.htm - DAML-S matchmaker homepage

http://www.w3.org/2001/sw/ - Semantic web home at W3C

http://www.ontoweb.org/ - Ontoweb, a European Union founded propject about Ontology-based information exchange for knowledge management and electronic commerce

Flipside:

http://www.users.zetnet.co.uk/s.billen/skynet.htm - Background info on Cyberdyne Systems Skynet (for a short timeout-session, I do not think that it ever will happen – HAL will stop them)