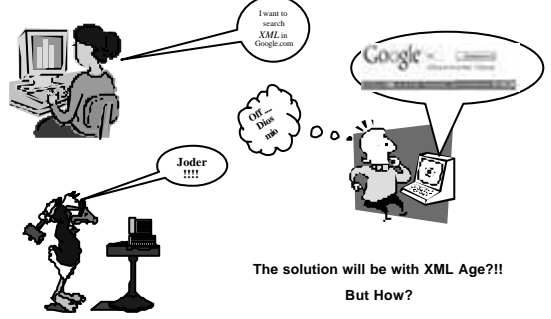


# How to Structure and Access XML Documents With Ontologies

This Presentation prepared  
By  
Jamil Itmazi, Granada Univ.  
Supervised by  
Dr. José Samos Jiménez  
26/5/2003

Jamil Itmazi / Granada Univ.

## The Problem of searching Internet/ The web



Jamil Itmazi / Granada Univ.

### • Introduction to XML

• What is XML? a markup language much like HTML AND stands for EXtensible Markup Language, XML uses a Document Type Definition (DTD) or an XML Schema to describe the data, XML with a DTD or XML Schema is designed to be self-descriptive.

XML was designed to describe data and focus on what data is.  
XML was designed to store, carry, and exchange data.  
XML was not designed to display data.  
HTML was designed to display data and focus on how data looks.  
XML can Separate Data from HTML  
With XML, your data is stored outside your HTML.

XML DTD : The purpose of a DTD is to define the legal building blocks of an XML document. It defines the document structure with a list of legal elements.  
XML Schema : is an XML based alternative to DTD, describes the structure of an XML document, supported from W3C as an alternative to DTD called XML Schema.

XML is going to be everywhere  
all future applications will exchange their data in XML.

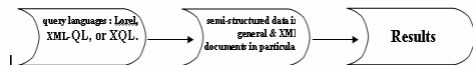
Jamil Itmazi / Granada Univ.

But

where XML sources are presented without an explicit agreement on the semantics of certain tags & document structures, the task of correctly interpreting the data is more difficult.

TO survey of XML many lists vocabularies standards, showing that these standards have some amount of overlap.  
the biggest advantage of XML (its extensibility) is at the sometime its biggest handicap. XML is so flexible that XML documents cannot be automatically provided with an underlying semantics.

- Then, how access to XML documents is currently realized
- by XML query languages.
- Access to semi-structured data is currently realized by a number of query languages, e.g. Lorel, Lorel for XML, XML-QL, or XQL.



Jamil Itmazi / Granada Univ.

### An example in XML-QL

We are looking for people with knowledge of the SGML language:

- WHERE
- <people>
- <persons>
- <name>SP</name>
- <know how>SGML</know-how>
- </person>
- <people IN 'some URL'
- CONSTRUCT SP

The query seek the content of the name sub element of person, iff the person has another sub element know-how containing "SGML", we (only) get Peter and Hans as answers.

- <skills>
- </people>
- </person>
- <name>Peter</name>
- <know-how>SGML</know-how>
- </person>
- </person>
- <name>Hans</name>
- <know-how>SGML</know-how>
- <know-how>Perl</know-how>
- </person>
- </people>
- </seminars>
- <seminar>
- <topic>SGML</topic>
- <participant>
- <name>Dieter</name>
- <name>Gisela</name>
- </participant>
- </seminar>
- </seminars>
- </skills>

Jamil Itmazi / Granada Univ.

But !!!, The given XML document also contains facts about seminars and if we need more search about another facts, we need more query.  
all possible encodings of the relevant knowledge has to be guessed by the information seeker to formulate such a query.

Neither XML & RDF define a standard vocabulary for describing semantics of documents  
Neither XML nor RDF define a standard structure for describing semantics of documents  
Define standard structure or mappings between different structures

When adding semantics to XML, we can solve some of the problems arising here.  
To see the power of ontologies, a relationship between the ontology and the structure of XML documents has to be established.

The ontology has 2 purposes in supporting queries :

- *it defines* a common vocabulary to enable structure independent queries.
- *it provides* additional background knowledge to improve the quality of answers.

In our example, facts like seminars, is part of the ontology's background knowledge.  
Another benefit comes from the modelled concept taxonomy.

Assume Tom is a programmer :

```
<programmer>
  <name>Tom</name>
  <know-how>SGML</know-how>
  <know-how>Java</know-how>
  <know-how>C++</know-how>
</programmer>
```

in languages for querying XML, Tom would not be retrieved by any of the given queries (but !!! everybody knows that programmers are persons).

Jamil Itmazi / Granada Univ.

**Okay ! current query languages are strong when the location of wanted data is known in advance or at least if the information seeker is aware of the document structure. But the additional knowledge cannot directly be made available to seekers.**

- Our approach enable integrated access to different XML document types via the use of ontologies.
- An ontology provides a formal, shared specification of concepts, their relationships, and other realities of some domain. This knowledge is independent of any representation.
- will providing information integration capabilities to XML processing by defining a semantic underpinning for XML documents based on ontologies .
- Thus, the content of these documents can be uniformly accessed and integrated, and Our approach enables true semantic queries to the contents of collections of XML documents and relieves the information seeker from the burden of knowing the structure of documents The ontology provides a conceptual view to XML documents.

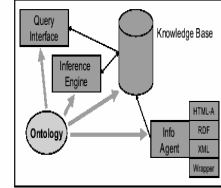
Defining the Context : The framework for doing this is set up by ONTOBROKER (built in 1997), an information integration and retrieval system in which we will use ontologies to access information contained in XML documents.

The ONTOBROKER uses techniques such as ontologies to provide access to different & distributed semi-structured documents. Underlying all parts of ONTOBROKER is an ontology that defines the vocabulary for annotating documents,

The principle architecture of ONTOBROKER consists of 4 independent parts :

Fig. 1. A Sketch of ONTOBROKER's architecture.

- 1-The query interface allows information seekers to pose queries to the knowledge.
- 2-The knowledge base contains all basic facts in Frame Logic.
- 3-An inference engine provides services that derive additional knowledge through inference rules on demand or in advance.
- 4- The info agent can access : HTML-A documents (a extension of HTML), XML documents .etc.



## Discussion

- Ontologies provide a compact, formal & conceptually way of describing the semantics of XML documents.
- By deriving DTDs from an ontology the document structure is grounded on a true semantic basis and thus, XML documents become adequate input for semantics-based processing.
- By providing a conceptual foundation for XML we achieve at the same time away to access sets of differently structured XML documents rather independently of their actual linear representation.
- The ontology provides a shared vocabulary that integrates the different XML sources, that makes the information uniformly accessible, and thus mediates between the conceptual terms used by an information seeker and the actual mark-up used in XML documents.

**Is the Problem of searching Solve, I hope, But for sure every technology come with it's problems!!**

