



# Semantic web

Tapas Kumar Mishra

11CS60R32

# Agenda



## ■ Introduction

- What is semantic web
- Issues with traditional web search

## ■ The Technology Stack

- Architecture of semantic web

## ■ Meta Data

## ■ Main Tasks

- Knowledge Representation
  - XML
  - RDF
- Ontology
  - Taxonomy
  - Inference Rules

## ■ Conclusion

- And references



# Introduction

The **Semantic Web** is an extension of the current web in which information is given well-defined **meaning**, better enabling computers and people to **work in co-operation**.

[Tim Berners-Lee , 2001]

# Introduction contd...



- Traditional search

Displays the pages that contain the words without interpreting the meaning of those words.

birthplace sachin tendulkar

---

About 190,000 results (0.25 seconds)

---

[Sachin Tendulkar Profile](#)  
[www.yehhaicricket.com/india/sachinten/sachin.html](http://www.yehhaicricket.com/india/sachinten/sachin.html)  
Full Name: **Sachin** Ramesh **Tendulkar**. Date of Birth: April 24, 1973. **Place** Mumbai. Major Teams: India, Mumbai. Batting Style: Right -Hand Batsman ..

[Sachin Tendulkar - Wikipedia, the free encyclopedia](#)  
[en.wikipedia.org/wiki/Sachin\\_Tendulkar](http://en.wikipedia.org/wiki/Sachin_Tendulkar)  
On 24 May 1995, **Sachin Tendulkar** married Anjali, a paediatrician and daughter of a former Indian cricketer. In his **biography**, it is stated that "Bradman was most taken by Tendulkar's ..  
[List of international cricket centuries ... - Achievements of Sachin Tendulkar](#)

# Introduction contd...

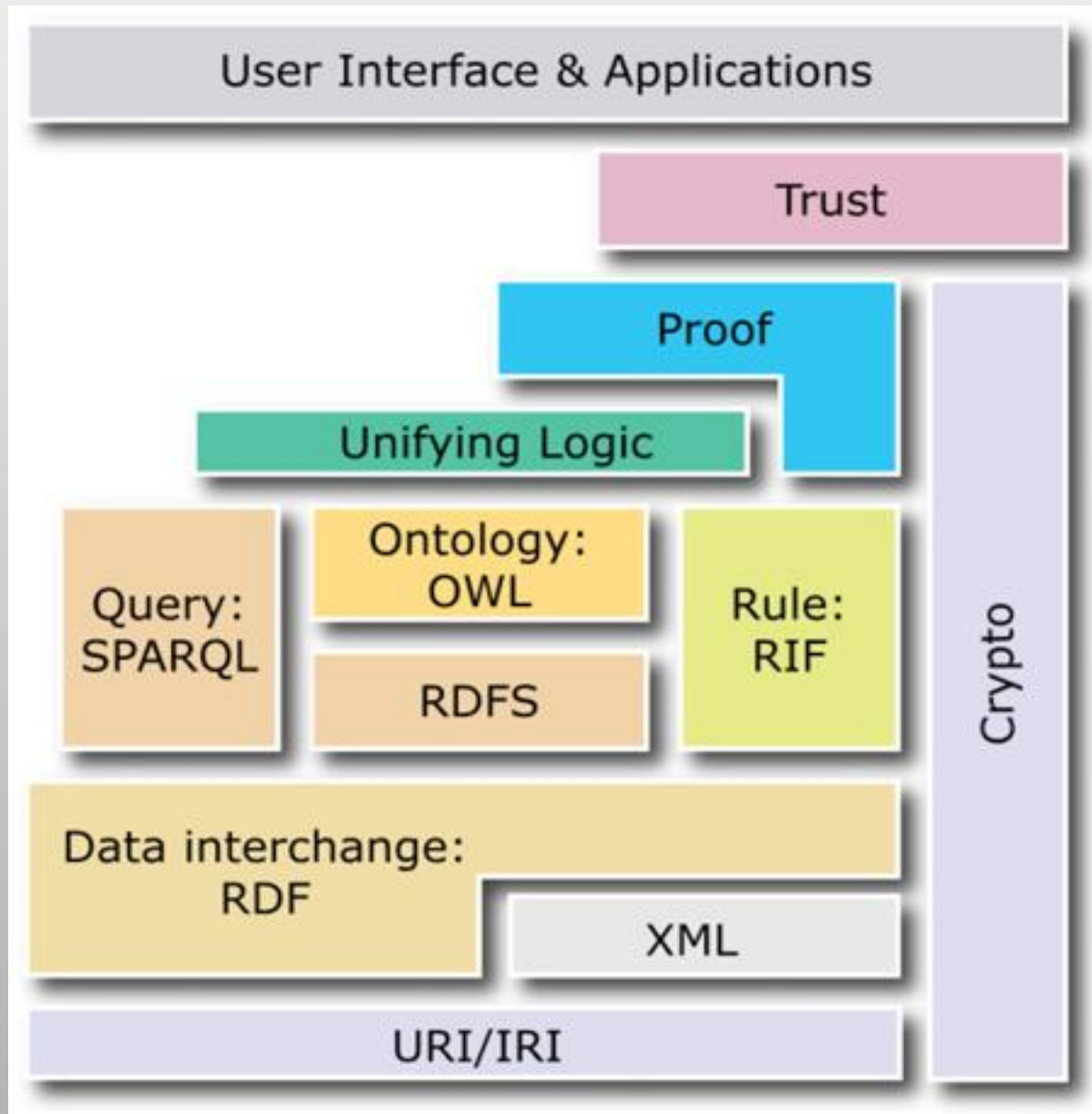


- Semantic Search

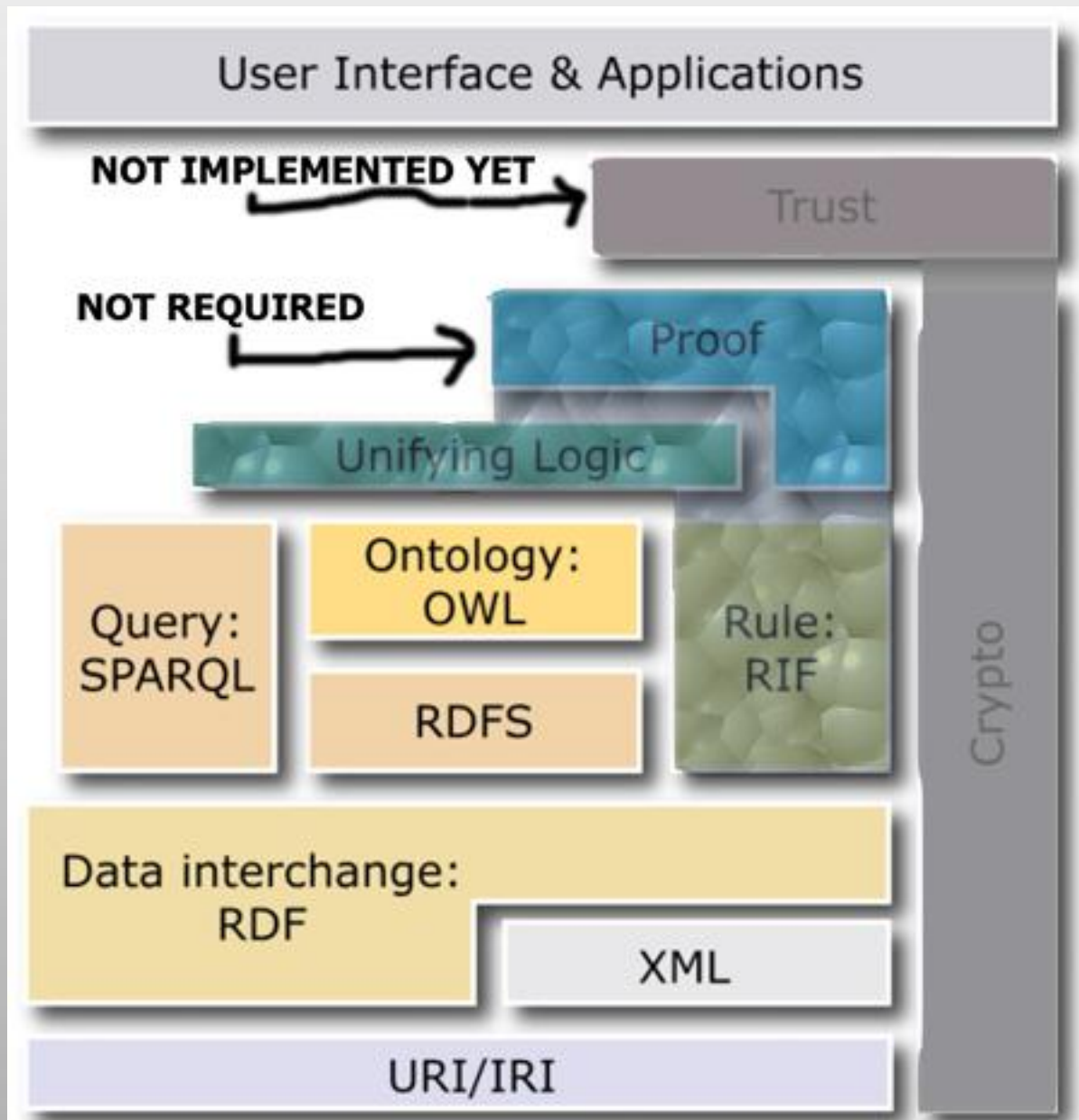
retrieves the meaning from the bag of words

A screenshot of the WolframAlpha search interface. At the top, the WolframAlpha logo is displayed in red and orange, with the text 'computational... knowledge engine' to its right. Below the logo is a search input field with the text 'birthplace sachin tendulkar'. To the right of the input field is a link labeled 'Examples'. Below the input field is a section titled 'Input interpretation:' which shows two boxes: 'Sachin Tendulkar' and 'place of birth'. Below this is a section titled 'Result:' which displays the text 'Mumbai, Maharashtra, India'.

# Semantic Web Stack



# Real world implementation



# Metadata



- The first form of semantic data on the Web was metadata :  
information about information
- Basically include:
  1. Means of creation of the data
  2. Purpose of the data
  3. Time and date of creation
  4. Creator or author of data
  5. Placement on a computer network where the data was created
  6. Standards used



# Metadata Contd..



Example :

- a meta element specifies name and associated content attributes describing aspects of the HTML page.

```
<meta name="keywords" content="wikipedia,encyclopedia">
```

- default charset for plain text is simply set with meta:

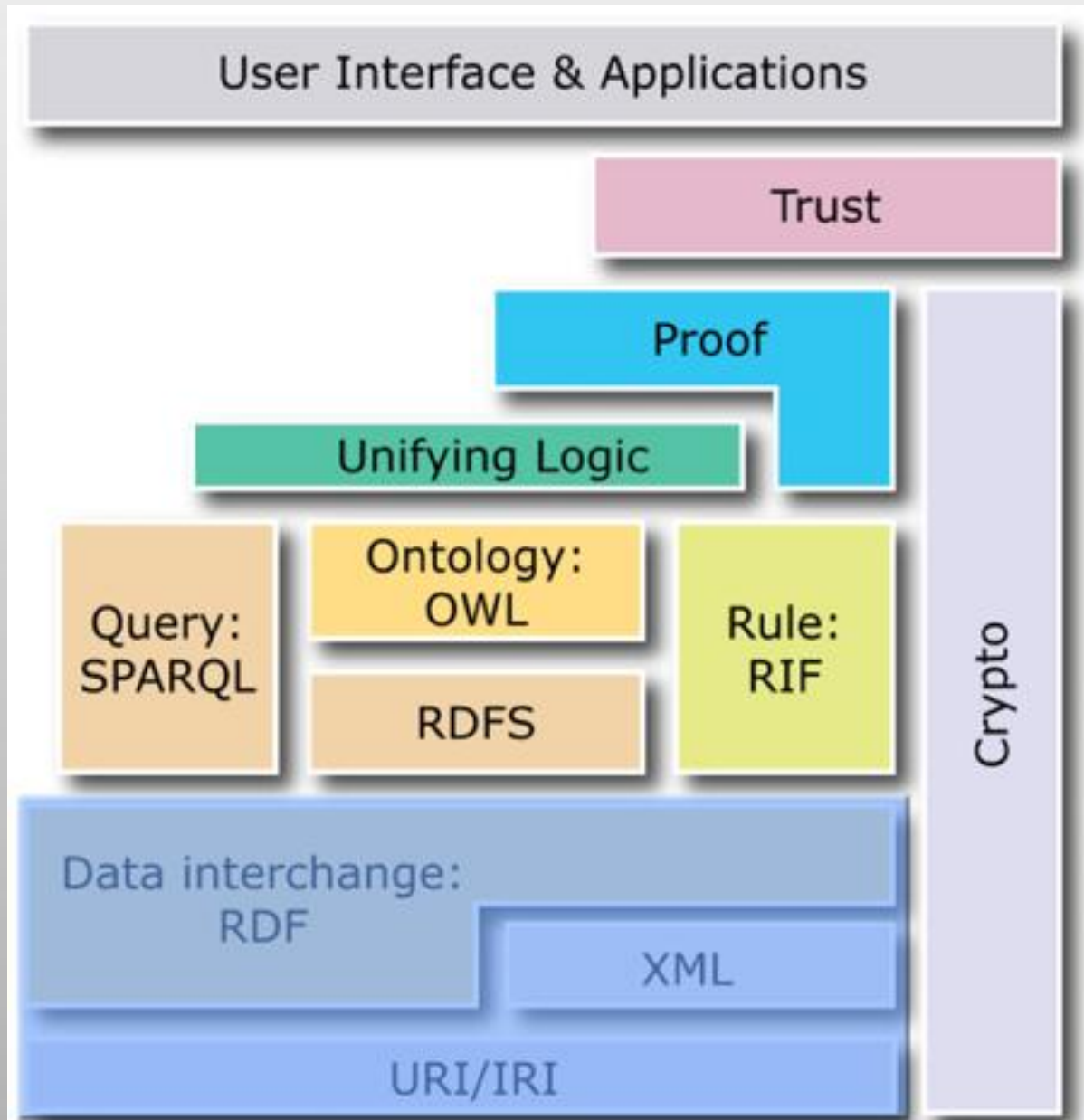
```
<meta http-equiv="Content-Type" content="text/html charset=UTF-8" >
```

# Semantic Web main tasks



- Knowledge Representation:
  - Metadata annotation
  - Description of resources using standard languages
- Search:
  - Retrieve relevant information according to user's query / interest / intention
  - Use metadata (and possibly content) in a "smart" way (i.e. "reasoning" about the meaning of annotations)

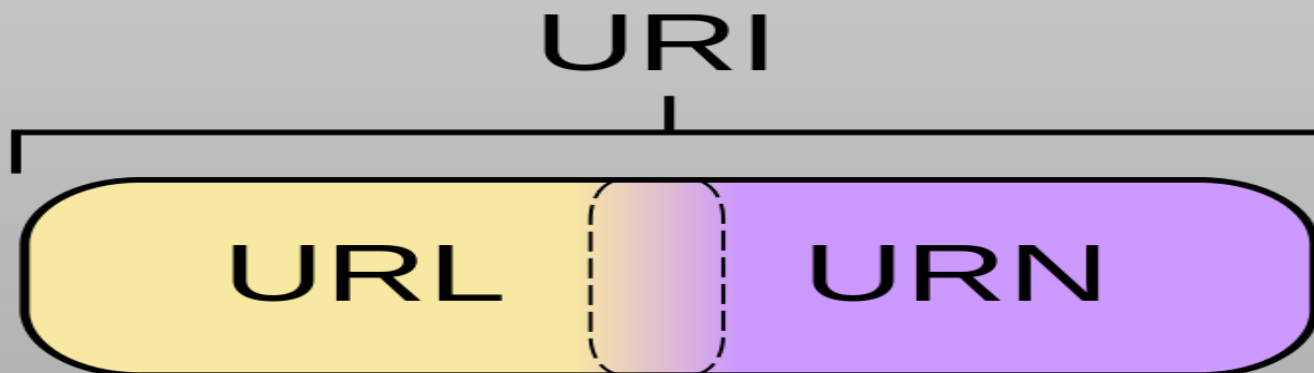
# Knowledge Representation



# URI



- string of characters used to identify a name or a resource on the Internet
- Categorized into 2 types
  - URL → Uniform Resource Locator
  - URN → Uniform Resource Name
- URN defines an item's identity, while the URL provides a method for finding it
- Example:
  - URN ----> ISBN of books,  
ISBN 0486275574 cites, unambiguously, a specific edition of Shakespeare's play Romeo and Juliet.



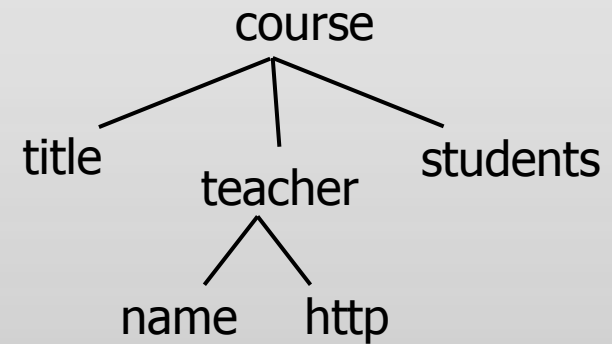
# XML



- User definable and domain specific markup

```
<course date="...">
  <title>...</title>
  <teacher>...
    <name>...</name>
    <http>...</http>
  </teacher>
  <students>...</students>
</course>
```

=



- XML provides an elemental syntax for content structure within documents
- But associates no semantics with the meaning of the content contained within.

# RDF(Resource Description Framework)



- corner stone of the Semantic Web technology stack
- 1999, first publication
- directed and labeled graphs as data model
- everything is univocally identifiable with a Uniform Resource Identifier(URI)
  - a web page, a person, a book, an intangible thing



# RDF CONTD...

A statement is a triple

- ❑ Subject –predicate –object
- ❑ Subject: a resource
- ❑ Predicate: a verb / property / relationship
- ❑ Object: a resource, or a literal string

Relationships between things could be expressed with a directed, labeled graph where

- nodes could be resources or XMLSchema-typed values and
- relationships are identified also by URIs

# RDF CONTD...



Author = D.West

Diagram:



Simple RDF assertion triple :

**triple (hasAuthor, URI, D.West)**

RDF in XML syntax:

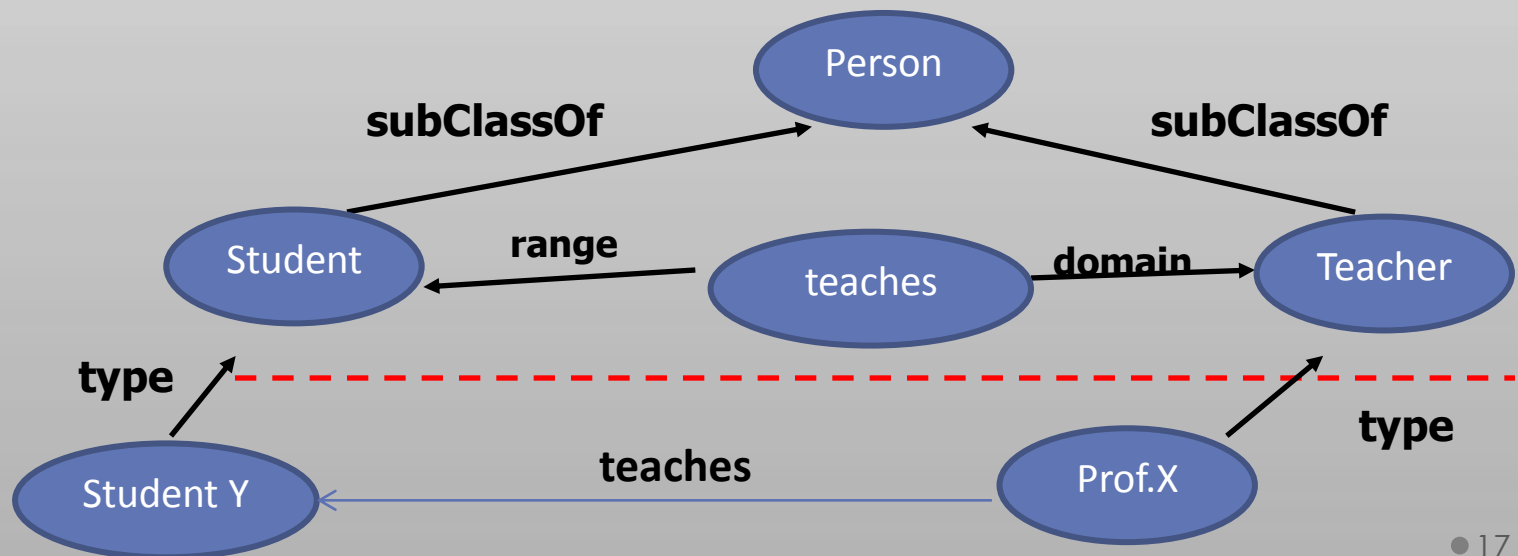
```
<RDF xmlns="http://www.w3.org/TR/ ... " >  
<Description about="http://www.w3.org/semweb/intro">  
<Author>D.West</Author>  
</Description>  
</RDF>
```



# RDF(S): RDF Schema



- Defines **vocabulary** for RDF
- Organizes this basic vocabulary terms and the relations between those terms
  - Class, subClassOf, type
  - Property, subPropertyOf
  - domain, range



# RDF CONTD...



- triples of RDF form webs of information about related things.
- the URIs ensure that concepts are not just words in a document but are tied to a unique definition that everyone can find on the Web.

Example:

a database of info of people , including their addresses.

RDF can specify that "(field 5 in database A)  
(is a field of type)(zip code),"

Query: Find people living in city with zipcode x

# RDF CONTD...



- Problem with RDF:
  - Synonym problem:
    - two databases may use different identifiers for what is in fact the same concept, such as zip code. A program that wants to compare or combine information across the two databases has to know that these two terms are being used to mean the same thing.

# Ontology



- Problems with RDF
  - two databases may use different identifiers for what is in fact the same concept, such as *zip code*
- Solution is Ontology....

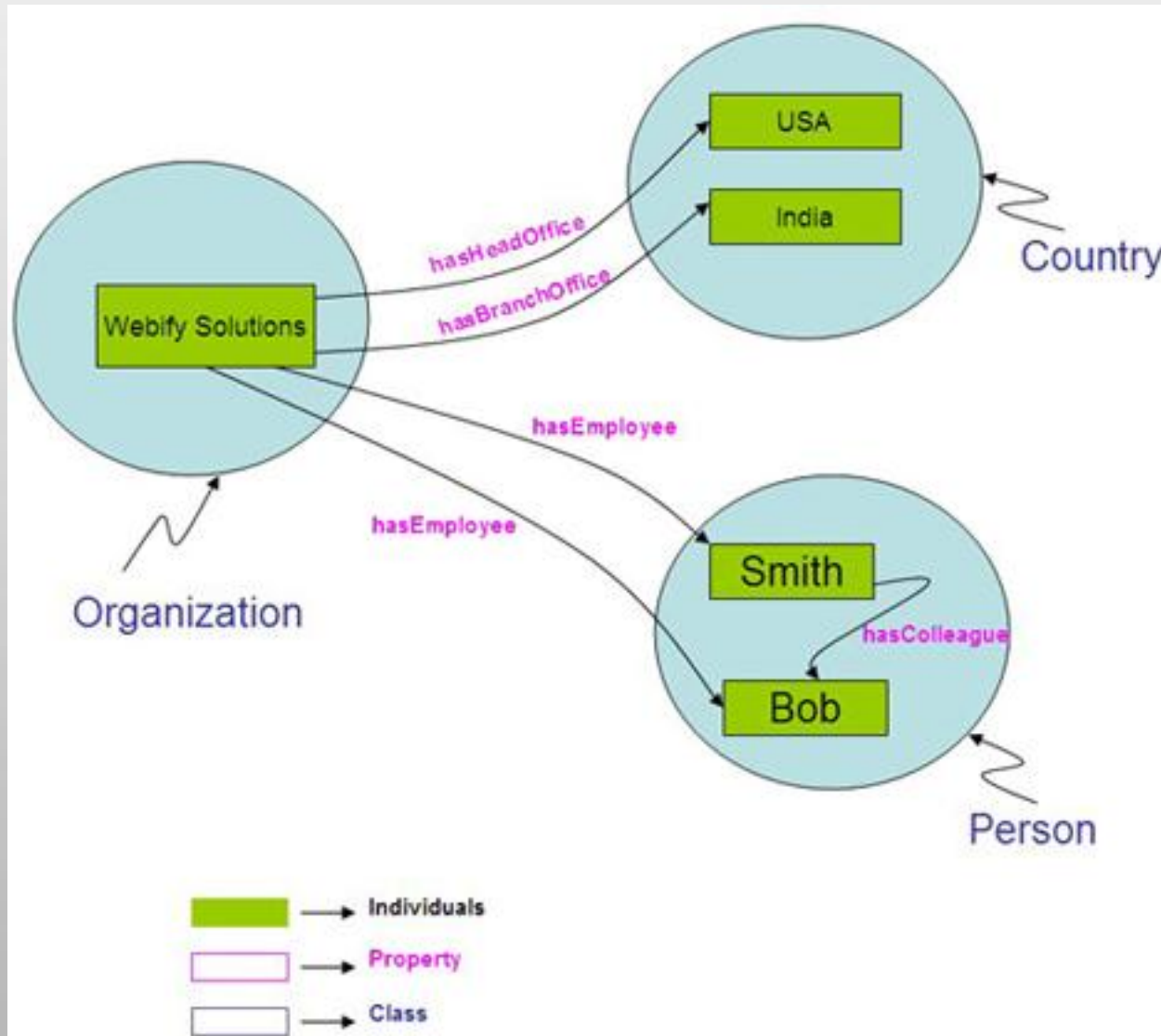
An ontology is an explicit description of a domain

- concepts
- properties and attributes of concepts
- constraints on properties and attributes
- individuals (often, but not always)

Web ontology consists of

- Taxonomy
- Inference Rules

# Ontology Contd...

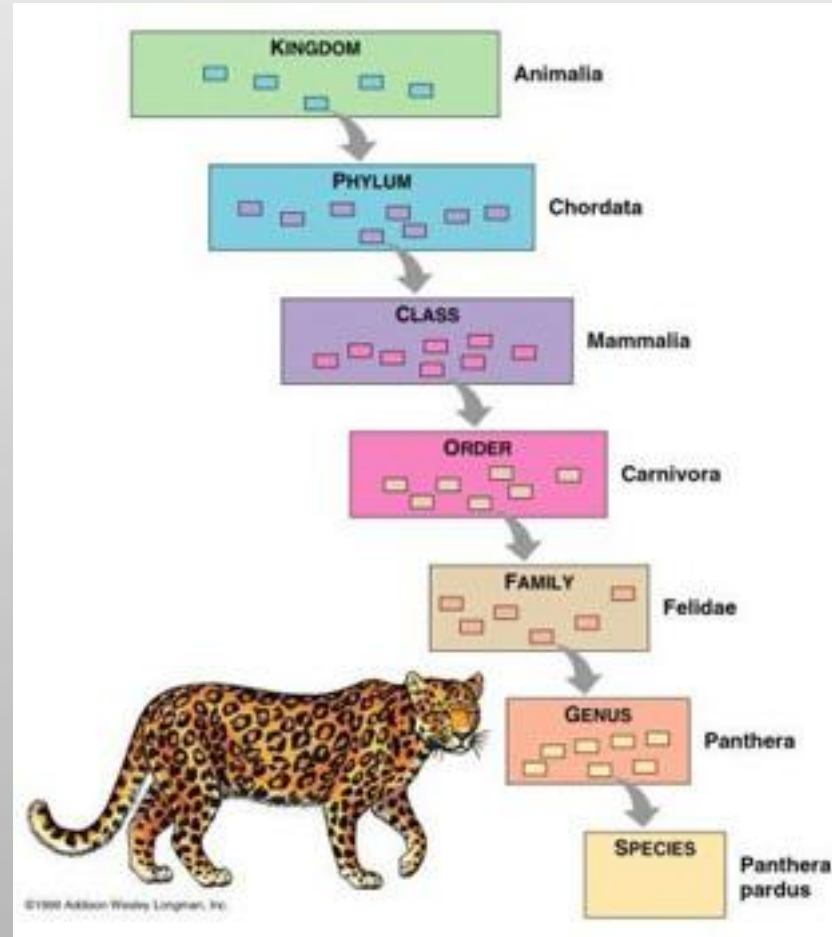


ontology that describes the Webify Solutions organization

# Taxonomy



- Defines classes of objects and relations among them

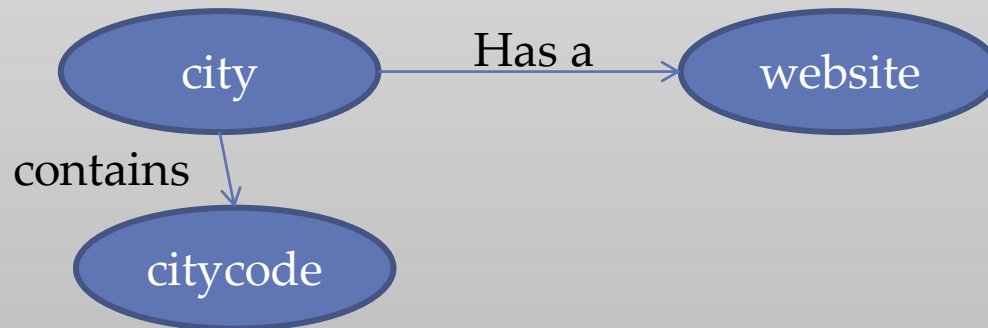


# Taxonomy Contd...



HOW Taxonomy helps:

- classes, subclasses and relations among entities are a very powerful tool for Web use
- city codes must be of type city and cities generally have Web sites, we can discuss the Web site associated with a city code even if no database links a city code directly to a Web site.



# Inference Rules



- Allows us to infer conclusions based on rules and facts available in the knowledge base
- Example:
  - An ontology may express the rule "If a city code is associated with a state code, and an address uses that city code, then that address has the associated state code."
  - A program could then readily deduce, for instance, that a IIT KGP address, being in Kharagpur, must be in West Bengal, which is in the India, and therefore should be formatted to indian standards.



# Inference Rules contd..



- this solve RDF's synonym problem.
- Example:
  - an ontology that defines addresses as containing a zip code and another ontology uses postal code.
  - The program could then use a service that takes
    - a list of postal addresses (defined in the first ontology) and
    - converts it into a list of physical addresses (the second ontology) by recognizing and removing post office boxes and other unsuitable addresses.



# PRO's of Semantic web

- Accuracy of web search
- Tackle complicated questions
- Inpage answer to query

# CONCLUSIONS AND FUTUREWORK



- Implementation of encryption layer
- Standard for retrieval
- Standard for metadata

# References



1. Berners-Lee, Tim; James Hendler and Ora Lassila (May 17, 2001). "The Semantic Web". Scientific American Magazine
2. [www.en.wikipedia.org/wiki/Semantic\\_Web](http://www.en.wikipedia.org/wiki/Semantic_Web)
3. Tim Berners-Lee, with Mark Fischetti. Harper San Francisco, 1999. "Weaving the Web"
4. <http://www.w3.org/2001/sw/>
5. [www.SemanticWeb.org/](http://www.SemanticWeb.org/)
6. James Farrugia, University of Maine, Orono, ME. " Model-theoretic semantics for the web". ACM New York, NY, USA ©2003



Thank you