SENG 42273
Semantic web and Ontological Engineering

Lecture 1 : Introduction

Course outline

Teaching and assessment

- Course credits- 3 credits
- No. of teaching hours- 45 hours (including tutorial sessions)
- Teaching & learning method- Lectures, supervised practical sessions and assignments
- Assessment- end of course examination, continuous assessments
- Reference- Introduction to the Semantic Web and Semantic Web Services by Liyang Yu, A Semantic Web Primer (third edition)

Course content

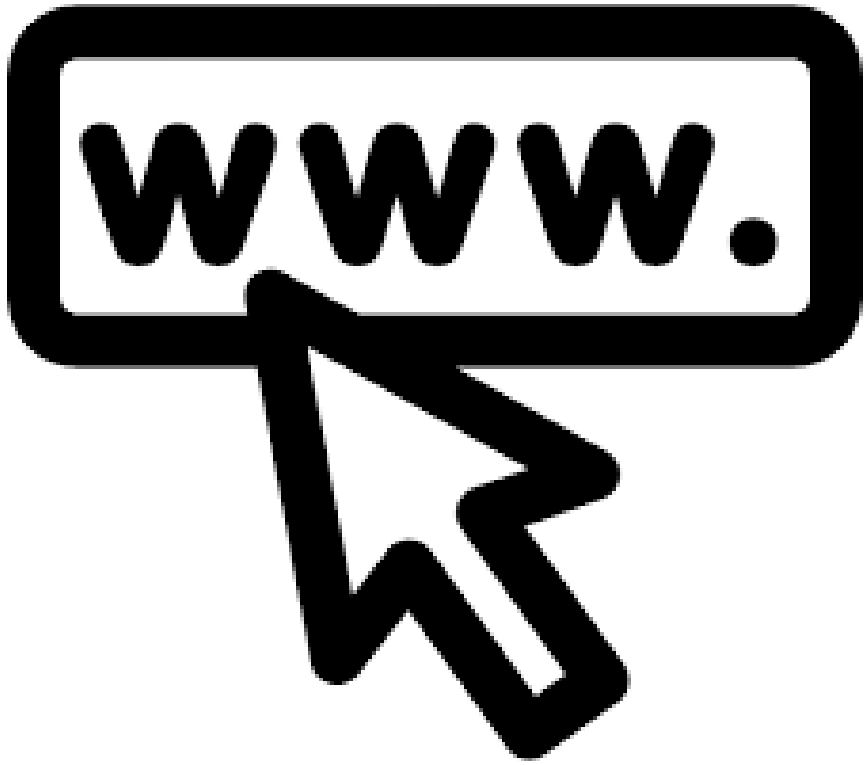
Part I: The World of the Semantic Web

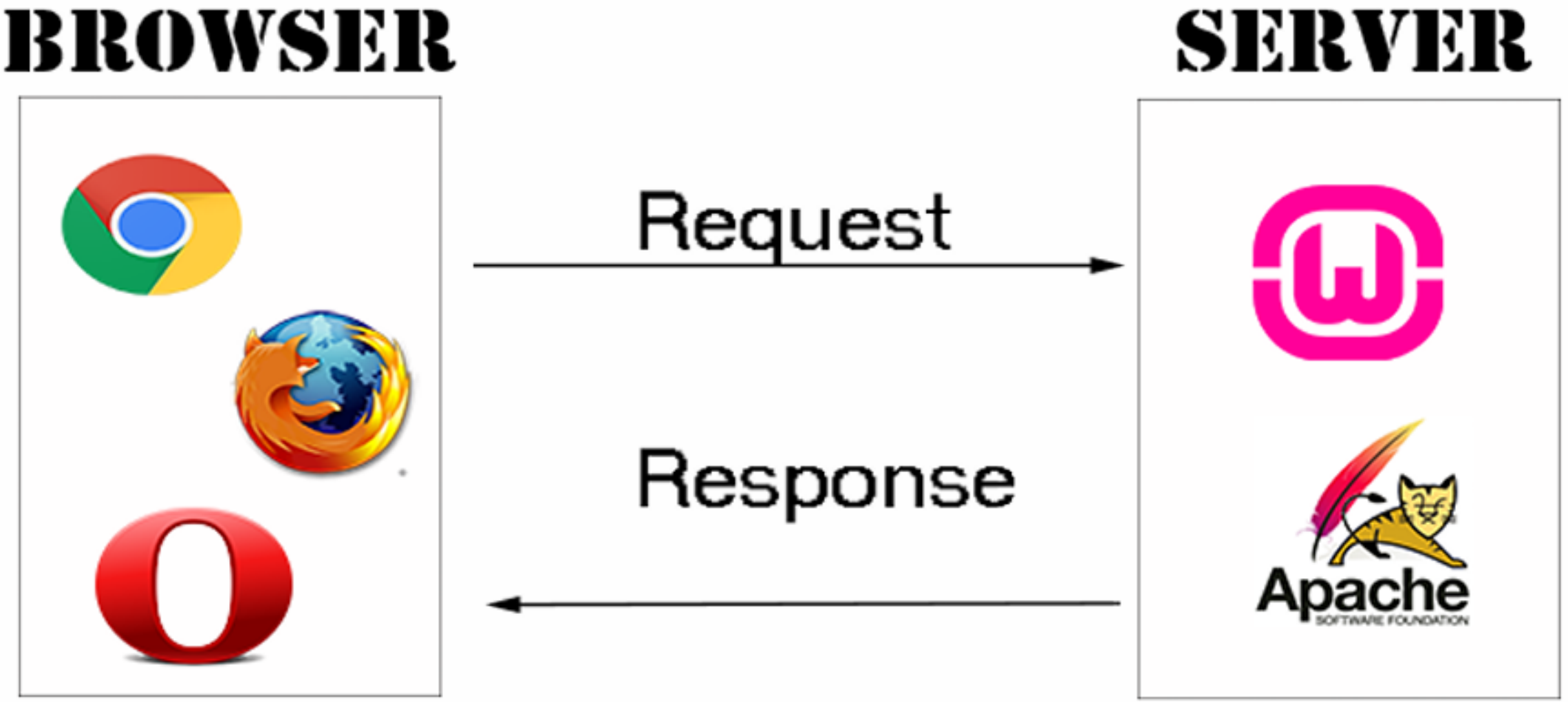
- From Traditional Web to Semantic Web
- Semantic web fundamentals

Part II: The Nuts and Bolts of Semantic Technology

- The Building Block of the Semantic Web: RDF
- RDFS, Taxonomy, and Ontology
- Web Ontology Language: OWL

Part III: The Semantic Web: Real-world Examples and Applications

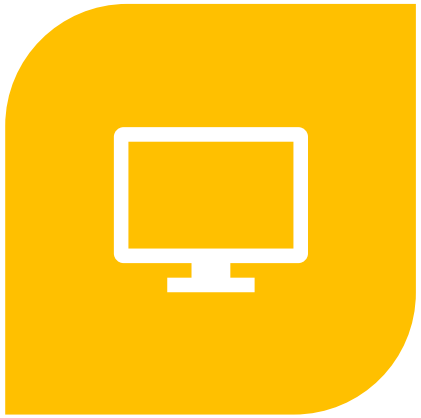

www.


BROWSER
SERVER
Request
Response
Apache
SOFTWARE FOUNDATION

How we use Internet?

SEARCH

INTEGRATION



WEB MINING

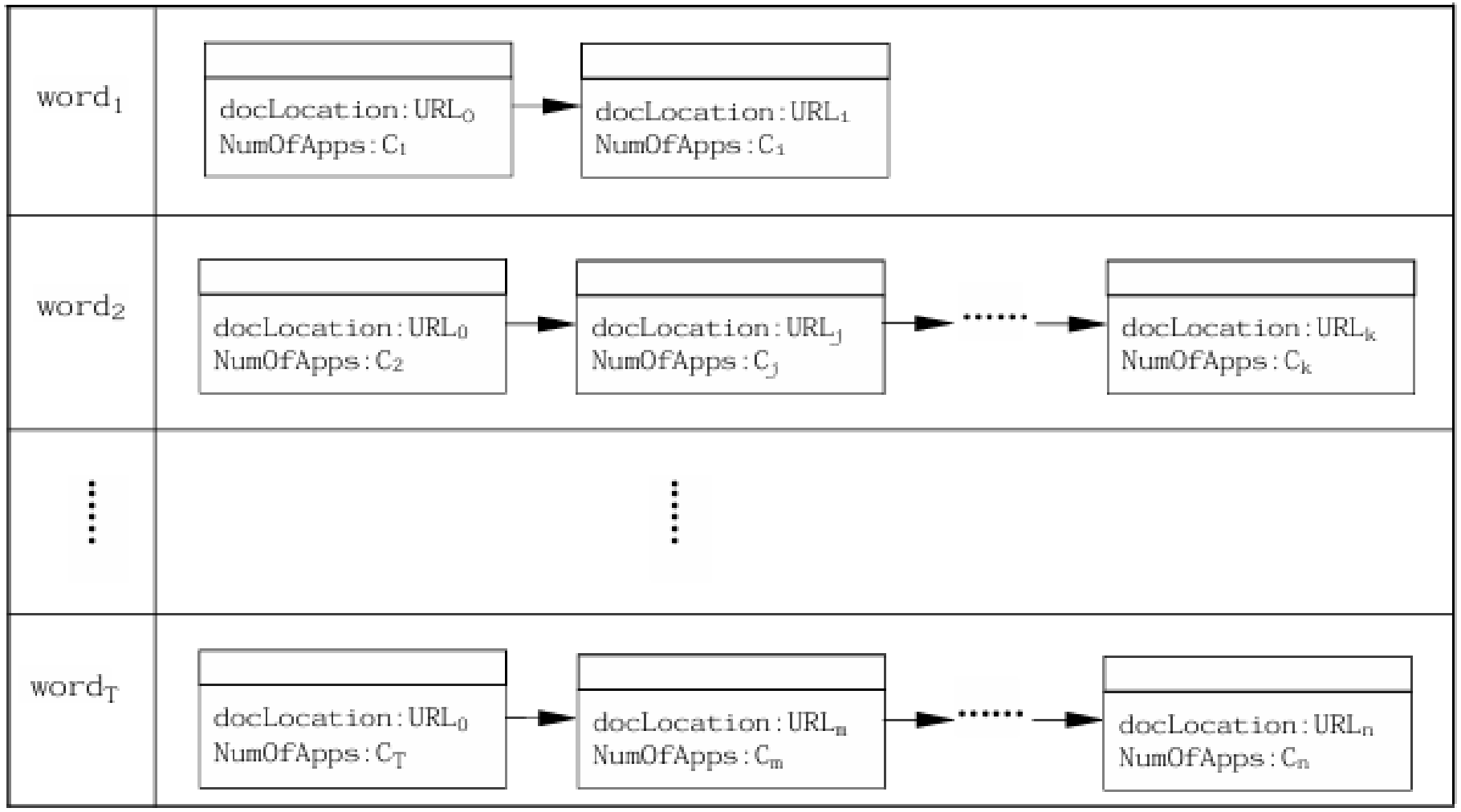
Search

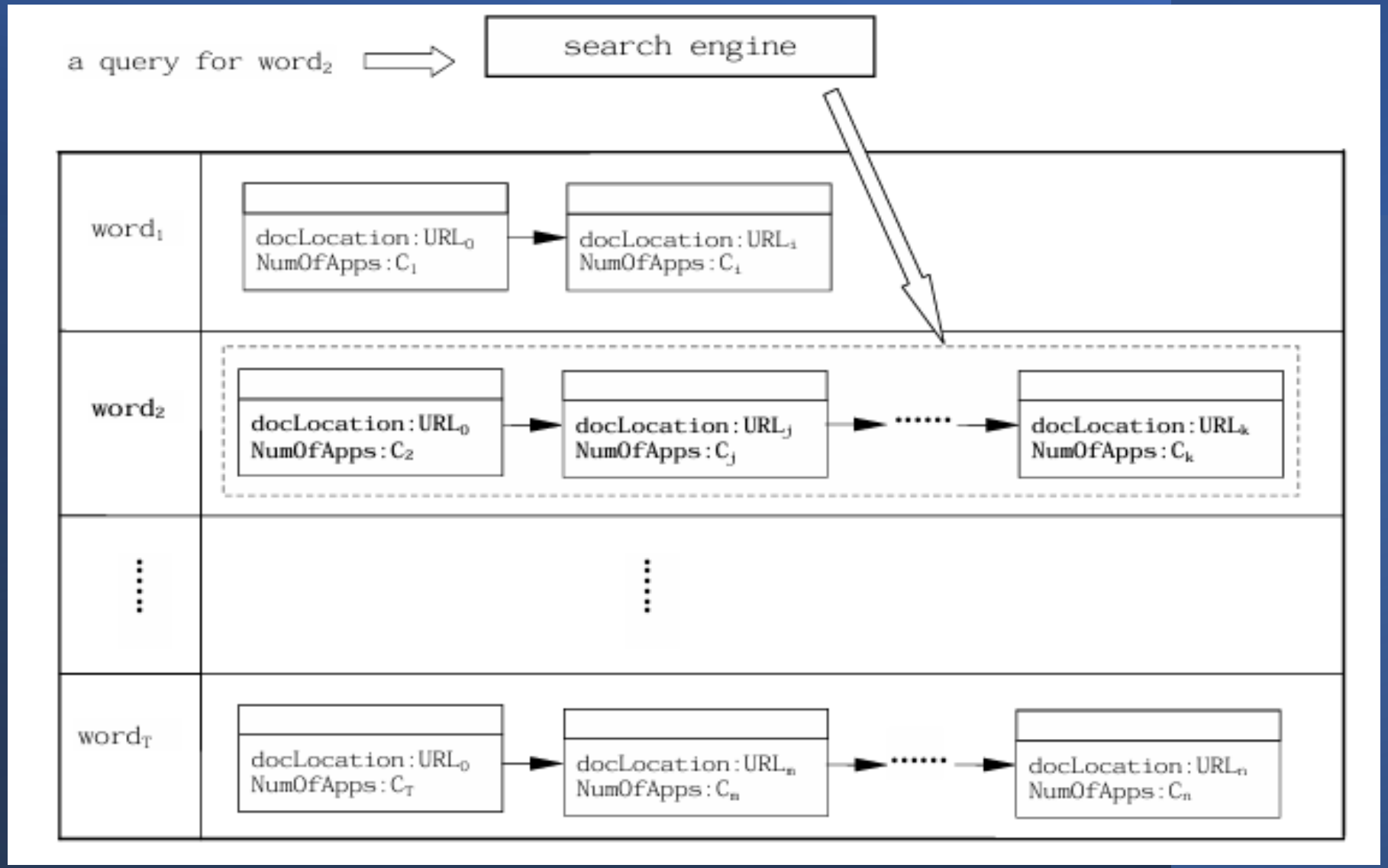
The goal is to locate and access information or resources on the Web.

Search engines implement their search based on which documents contain the given keyword.

Search Engine For The Traditional Web

1. Building the Index Table
   - Step 1: Build an index table for every single word on this page
   - Step 2: From the current page, find the first link (which is again a URL pointing to another page) and “crawl” to this link, meaning to download the page pointed to by this link.
   - Step 3: After downloading this page, start reading each word on this page, and add them all to the index table.
   - Step 4: Go to step 2, until no unvisited link exists

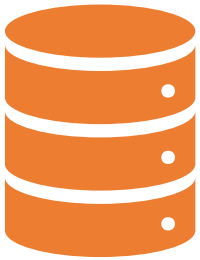

word1
docLocation:URL0
NumOfApps:C1
docLocation:URL1
NumOfApps:C1
word2
docLocation:URL0
NumOfApps:C2
docLocation:URLj
NumOfApps:Cj
docLocation:URLk
NumOfApps:Ck
wordT
docLocation:URL0
NumOfApps:CT
docLocation:URLm
NumOfApps:Cm
docLocation:URLn
NumOfApps:Cn


a query for word2
search engine
word1
docLocation:URL0
NumOfApps:C1
docLocation:URLi
NumOfApps:Ci
word2
docLocation:URL0
NumOfApps:C2
docLocation:URLj
NumOfApps:Cj
docLocation:URLk
NumOfApps:Ck
wordT
docLocation:URL0
NumOfApps:CT
docLocation:URLn
NumOfApps:Cn
docLocation:URLn
NumOfApps:Cn

Integration

- It means combining and aggregating resources on the Web so that they can be collectively useful.

Web Mining

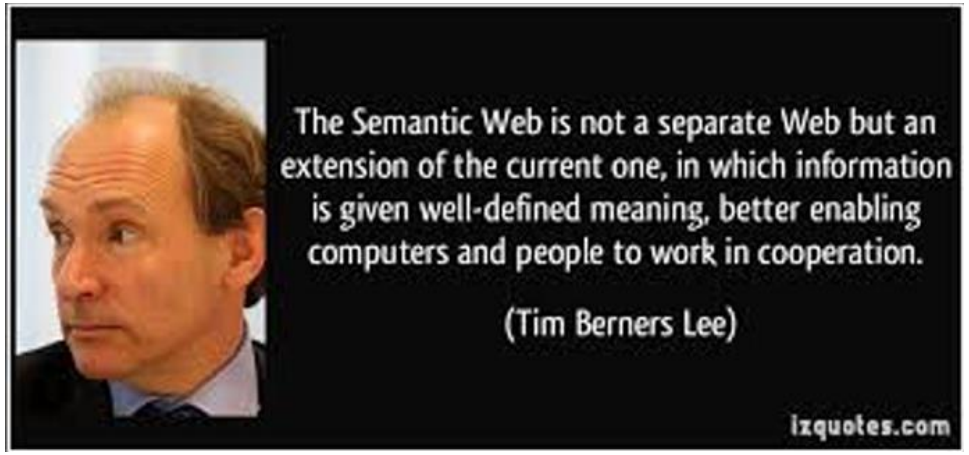


Data mining is the nontrivial extraction of useful information from large (and normally distributed) data sets or databases.

Web data mining refers to the activity of getting useful information from the Internet.

The Internet is constructed in such a way that its documents only contain enough information for the computers to present them, not to understand them.

Semantic Web



. . . the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration, and reuse of data across various applications.

— W3C Semantic Web Activity [12]

Metadata

- Data about data
- Data that describes information resources
- Structured data that machines can read and understand.
- To ensure metadata can be automatically processed by machines, some metadata standard is needed.
  - Dublin Core (DC) : It has 15 elements, which are called Dublin Core Metadata Element Set (DCMES)

Examples

Element Name	Element Description
Creator	This element represents the person or organization responsible for creating the content of the resource; e.g., authors in the case of written documents
Publisher	This element represents the entity responsible for making the resource available in its present form; it can be a publishing house, a university department, etc.
Contributor	This element represents the person or organization not specified in a creator element who has made significant intellectual contributions to the resource but whose contribution is secondary to any person or organization specified in a creator element; e.g., editor, transcriber, illustrator
Title	This element represents the name given to the resource, usually by the creator
Subject	This element represents the topic of the resource; normally, it will be expressed as keywords or phrases that describe the subject or content of the resource
Date	This element represents the date associated with the creation or availability of the resource
Identifier	This element is a string or number uniquely identifies the resource; examples include URLs, Purls, ISBN, or other formal names
Description	This element is a free text description of the content of the resource; it can be a flexible format, including abstracts or other content descriptions
Language	This element represents the language used by the document
Format	This element identifies the data format of the document; this information can be used to identify the software that might be needed to display or operate the resource; e.g., postscript, HTML, text, jpeg, XML

Example

```
<html>
<head>
<title>a joke written by liyang</title>
<meta name="DC.Title" content="a joke written by Liyang">
<meta name="DC.Creator" content="a joke">
<meta name="DC.Type" content="text">
<meta name="DC.Data" content="2004">
<meta name="DC.Format" content="text/html">
<meta name="DC.Identifier" content= http://www.codeproject.com/
script/profile/whos_who.asp?id=736920">
</head>
<body>
I decided to make my first son a medical doctor so that later on when
I am old and sick I can get medical care any time I need and for
free … in fact, better to make my second son a medical doctor, too,
so I can get a second opinion.
</body>
</html>
```

Summary

- The Semantic Web is an extension of the current Web; its main goal is to allow machine processing in a global scale.
- One way to accomplish this is to add metadata to the Web, as metadata is structured data, i.e., it is machine readable.
- DC schema seems simple, but it shows the key idea of adding metadata (meanings) to a given document.

Q & A