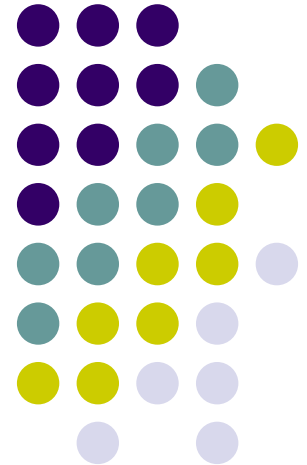


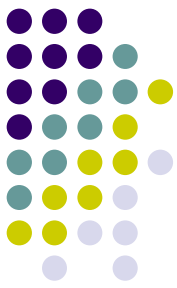
# Semantic Web Mining:

## Using Association Rules for Learning an Ontology

---

Presented By : Amgad Madkour





# Agenda

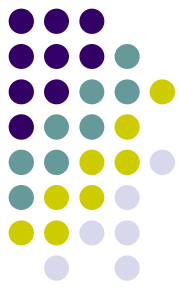
- Semantic Web Mining aim
- Web Mining overview
- Semantic Web overview
- Ontology Building
- Learning an Ontology using WM



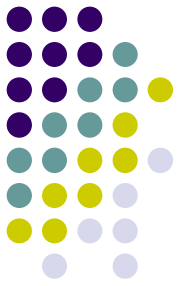
# What is the aim ?

- To **improve**, on one hand, the **results** of Web Mining by exploiting the new semantic structures in the Web
- To make use of Web Mining, on the other hand, for building up the Semantic Web.

# How can both of them help each other ?



- A backbone of semantic web are ontologies which at present are often **hand-crafted**.
- The challenge is to learn ontologies and/or instance of their concepts , in a **semi automatic way** by using web mining techniques such as **association rules** .
- Conversely , background knowledge – in the form of ontologies , or in other forms – can be used to improve the process and results of web mining.

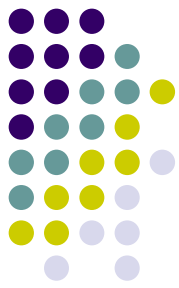


# Web Mining



# What is Web Mining ?

- Web mining is the application of data mining techniques to the content, structure, and usage of Web resources. This can help to discover global as well as local structure (“models” or “patterns”) within and between Web pages
- Web mining can profit from given structure on data (as in database tables), but it can also be applied to semi-structured or unstructured data like free-form text



# Web Mining Types : Content

- Web Content Mining is a form of text mining
- The primary Web resource that is being mined is an individual page.
- Web content mining can take advantage of the semi-structured nature of Web page text.
- Web content mining can be used to detect co-occurrences of terms in texts

# Web Mining Types : Structure



- Usually operates on the hyperlink structure of Web pages.
- The primary Web resource that is being mined is a set of pages, ranging from a single Web site to the Web as a whole.
- Exploits the additional information that is (often implicitly) contained in the structure of *hypertext*.
- Identification of the relative relevance of different pages that appear equally pertinent when analyzed with respect to their content in isolation.

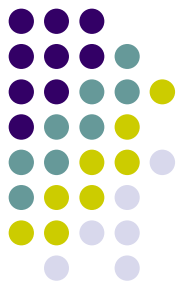




# Web Mining Types : Usage

- The primary Web resource that is being mined is a record of the requests made by visitors to a Web site, most often collected in a Web server log.
- The actual behavior of those who use these resources may reveal additional structure than the one in structure mining.
- It is also used for static site improvement by identifying navigational pattern of the user inside a Website
- It is useful to combine Web usage mining with content and structure analysis in order to “make sense” of observed frequent paths and the pages on these paths

# Extracting Semantics from The Web



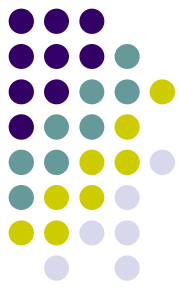
- The effort behind the Semantic Web is to add semantic annotation to Web documents in order to access knowledge instead of unstructured material, allowing knowledge to be managed in an automatic way.
- Web Mining can help to learn definitions of structures for knowledge organization (e.g., ontologies) and to provide the population of such knowledge structures.
- All approaches discussed are semi-automatic. They assist the knowledge engineer in extracting the semantics, but cannot completely replace him/her

# Exploiting Semantics for Web Mining : Web Content Mining



- We preprocess the input data (e. g. text) and apply ontology-based heuristics for feature selection and feature aggregation.
- Based on these representations, we compute multiple clustering results using k-Means.
- The results can be characterized and explained by the corresponding selection of concepts in the ontology.

# Mining



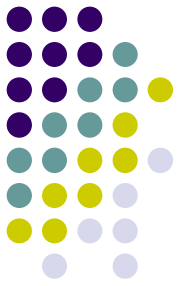
content mining



structure mining



usage mining

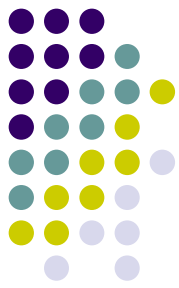


# Semantic Web



# What is Semantic Web ?

- Huge amount of data is interpretable by humans only; machine support is limited.
- Berners-Lee suggested to enrich the Web by machine-processable information which supports the user in his tasks
- To reach this goal the Semantic Web will be built up in different levels, the one we care about is ontologies



# What is an Ontology ?

- An **ontology** is a formal **explicit description of concepts** in a domain of discourse , **properties** of each concept describing various features and **instances** of the concept
- An ontology together with a set of individual **instances** of classes constitutes a **knowledge base**.

# Example :

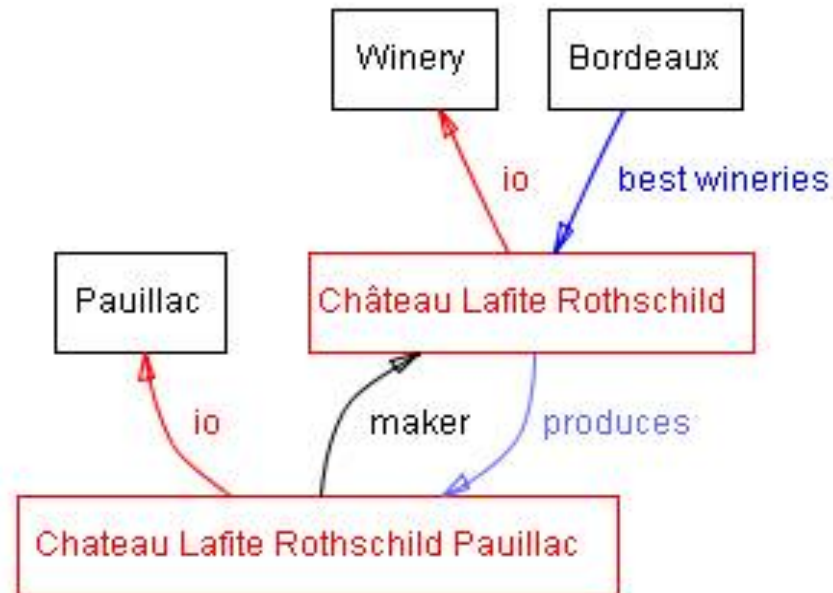


Figure 1. Some classes, instances, and relations among them in the wine domain. We used **black for classes** and **red for instances**. **Direct links** represent **properties** and internal links such as instance-of and subclass-of.

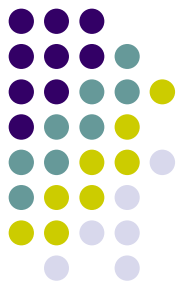


# Ontology vs. XML

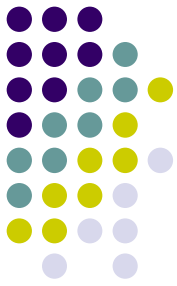


- An ontology differs from an XML schema in that it is a knowledge representation, not a message format.
- The XML specification is not designed to support reasoning outside the transaction context where ontology does.

# A Simple Knowledge Engineering Methodology

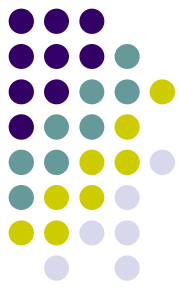


- There are some fundamental rules in ontology design:
  - There is **no one correct way to model a domain**
  - Ontology development is necessarily **an iterative process**.
  - Concepts in the **ontology should be close to objects**.



# Ontology Building

# Step 1: Determine the domain and scope of the ontology



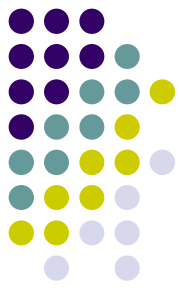
- We suggest starting the development of an ontology by defining its **domain** and **scope**. That is, answer several basic questions:
  - What is the **domain** that the ontology will cover?
  - **For what** we are going to use the ontology?
  - For what types of questions the information in the ontology should provide **answers**?
  - **Who will use** and maintain the ontology?

## Step 2. Consider reusing existing ontologies



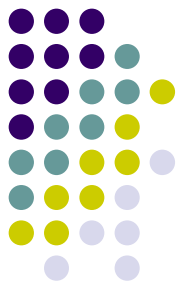
- Check if we can **refine** and **extend** existing sources for our particular domain and task
- Reusing existing ontologies **may be a requirement** if our system needs to interact with other applications that have already committed to particular ontologies or controlled vocabularies

# Step 3. Enumerate important terms in the ontology



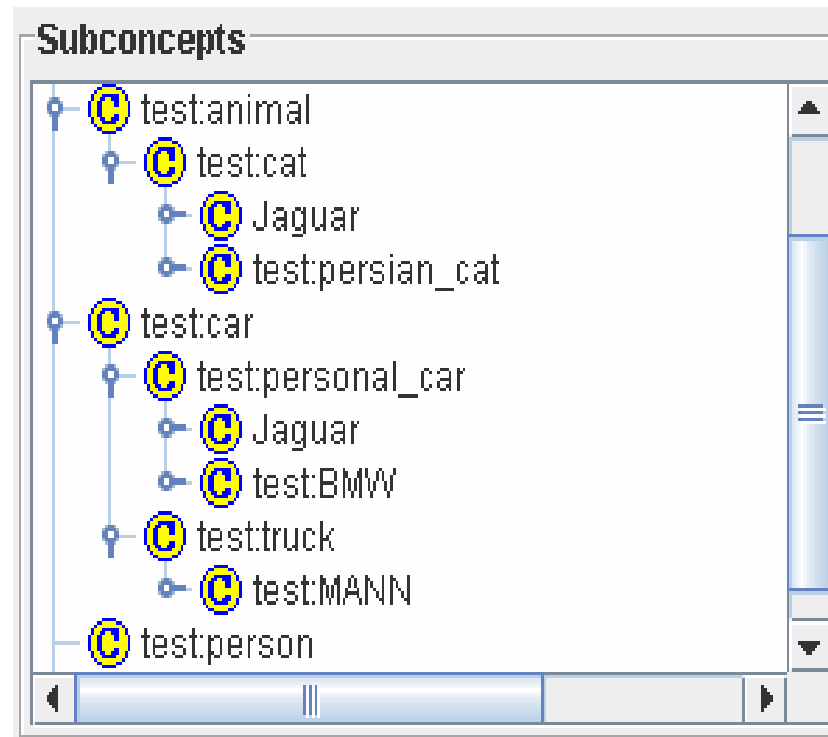
- It is useful to write down a list of all terms we would like either to make statements about or to explain to a user
- It is important to **get a comprehensive list** of terms **without worrying** about overlap between concepts they represent, relations among the terms, or any properties that the concepts may have, or whether the concepts are classes or slots..

# Step 4. Define the classes and the class hierarchy



- A **top-down** development process starts with the definition of the most general concepts in the domain and subsequent specialization of the concepts.
- A **bottom-up** development process starts with the definition of the most specific classes with subsequent grouping of these classes into more general concepts
- A **combination** development process: We define the more salient concepts first and then generalize and specialize them appropriately.

# Example : Concept Hierarchy

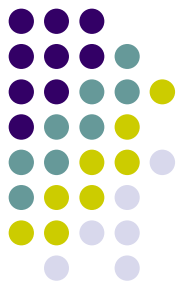






# **Learning an Ontology Using Web Mining**

# Step 1 : Extract the Terms from the Corpus ( Term Extraction )



- We obtain the **corpus** of the domain that we want to create the ontology for
- We start with extracting terms using term extraction techniques
- We specify the **linguistic filter** using Part Of Speech tags
- The **output** will be a list of concepts that the knowledge engineer can include inside his/her model in order to create the ontology whether in a automatic or semi automatic way.

## Step 2: Apply Association Rules to the extracted terms

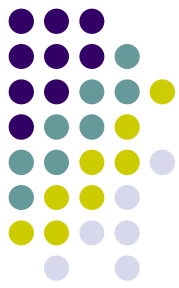


- We specify the minimum support level and minimum confidence level to the terms that have already been extracted as concepts
- The system rechecks the corpus using Part Of Speech again but now matching the extracted concepts with the corpus to determine the association rules between them
- The result then shows the support and confidence between the resulting concepts



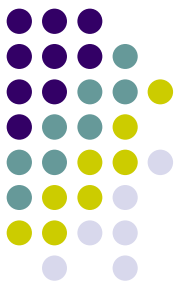
## Step 3 : Extract Instances

- We can help the knowledge engineer in automatically or semi-automatically extracting instances of the concepts that we proposed so far
- The system will not accurately be able to determine all instances from the corpus there for the aid of the knowledge engineer is required in order to obtain a better description of the ontology



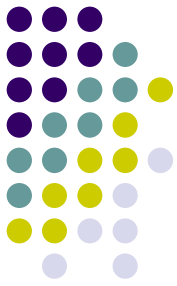
# Extra Refining Steps

- We can **apply pruning** to the concepts to remove any unrelated concepts that could populate our ontology
- As we mentioned the process that we present is **semi automatic and needs the knowledge engineering** to refine the resulting set of concepts that the system proposed



# Conclusion

- A mutual relationship between semantic web and web mining.
- The task of learning ontologies can be simplified for the knowledge engineer using web mining techniques such as association rules.
- We showed a complete run-through on how to create ontology using the web mining approach in an automatic/semi-automatic manner.



**Thank You !**