FRESNO STATE
Discovery. Diversity. Distinction.

# Web Programming (CSci 130)

Department of Computer Science
College of Science and Mathematics
California State University Fresno
H. Cecotti

# Learning outcomes

- **Web semantic**
  - What does it mean?
  - What is it for?
  - The relationships with other fields

# Rationale

- **Current stage**
  - ➢ Web programming == Web document
    - GUI / Web page ← → Data (from databases)
    - GUI / Query (client) → Class (JS) → jSON → Class (PHP) → MySQL (database)
    - Databases
  - ➢ **On the web**: request (e.g. search engine) → result
    - Not a result to the query **but** a list of documents related to the query
- **Current challenges**
  - Difficulties to find, present, access, or maintain available electronic information on the web
    - **Too much** information?
  - Need for a data representation to enable software products (agents) to provide intelligent access to **heterogeneous** and **distributed** information.

# Rationale

- **User query**
  - ➢ Simple answer(s)
    - ○ List of **documents** containing **data**
      1. Up to the user to find the most relevant documents
      2. Up to the user to extract the information
    - ○ **Output**: sorted list of documents
    - ○ Ideal output: 1 document that organizes all the extracted data
  - ➢ May involve several websites
    - ○ Example
      - Vacation = flight + hotel + car renting + activities …
    - ○ Change your address
      - → automatically updated everywhere where your address is mentioned

# Introduction

- "The Semantic Web is a major research initiative of the World Wide Web Consortium (W3C) to create a **metadata-rich** Web of resources that can describe themselves not only by how they should be displayed (**HTML**) or syntactically (**XML**), but also by the **meaning** of the metadata."
  - From W3C Semantic Web Activity Page


- "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 cooperation."
  - Tim Berners-Lee, James Hendler, Ora Lassila,
  - The Semantic Web, Scientific American, May 2001

# Ontology

- From the philosophy / metaphysics
  - The study of being or **existence**
- **Definition**
  - **A specification of a conceptualization**
  - A set of representational primitives with which to model a domain of knowledge or discourse
- Representation of information
  - Classes (sets)
  - Attributes (properties)
  - Relationships (or relations among class members).
  - Information
    - About the meaning of the elements
    - Constraints on their logically consistent application
- Database systems
  - level of abstraction of data models = hierarchical and relational models
    - **BUT** for modeling knowledge about individuals, their attributes, and their relationships to other individuals.

# Ontology

- Specification with languages that allow abstraction **away** from
  - Data structures and implementation strategies
- Languages of ontologies
  - Closer in expressive power to first-order logic
  - than languages used to model databases
    - → ontologies at the **semantic level**
    - → database schema are models of data at the **logical or physical level**
- Independence from lower level data models
  - used for integrating heterogeneous databases
  - → enabling **interoperability** among disparate systems
  - specifying interfaces to independent, knowledge-based services.
- Semantic Web standards
  - Ontologies → an **explicit** layer
    - Standard component of knowledge systems (KS)

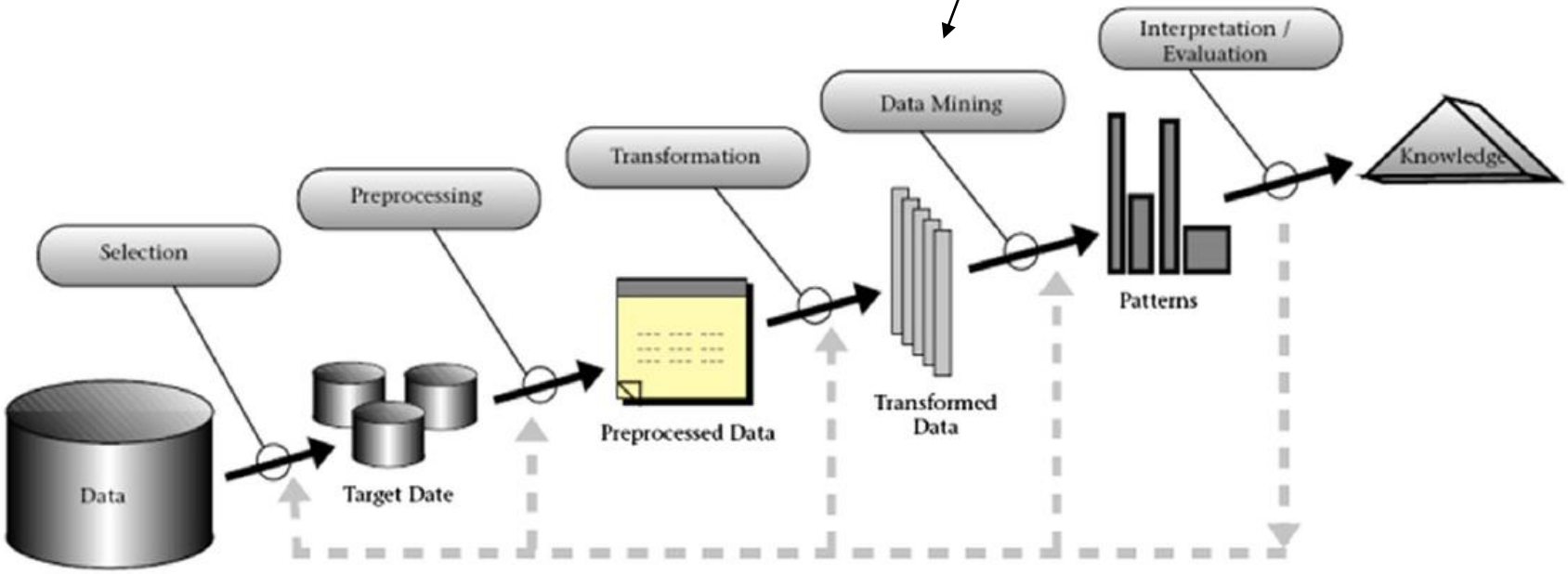
# Ontology

- Domain ontologies
  - ➢ Express background knowledge about the application domain
    - the domain of the data at hand on which KDD and data mining are performed.
- Ontologies for data mining process
  - ➢ Define knowledge about the data mining process
    - Steps and algorithms + their possible parameters
- Metadata ontologies
  - ➢ Describe **meta knowledge** about the data at hand
    - Provenance information
      - Example: the processes used to construct certain datasets
- Connections with Symbolic Artificial Intelligence
  - ➢ Manipulation of symbols
  - ➢ Manipulation of abstract compositional representations whose elements stand for objects and relations

# Data mining in the web

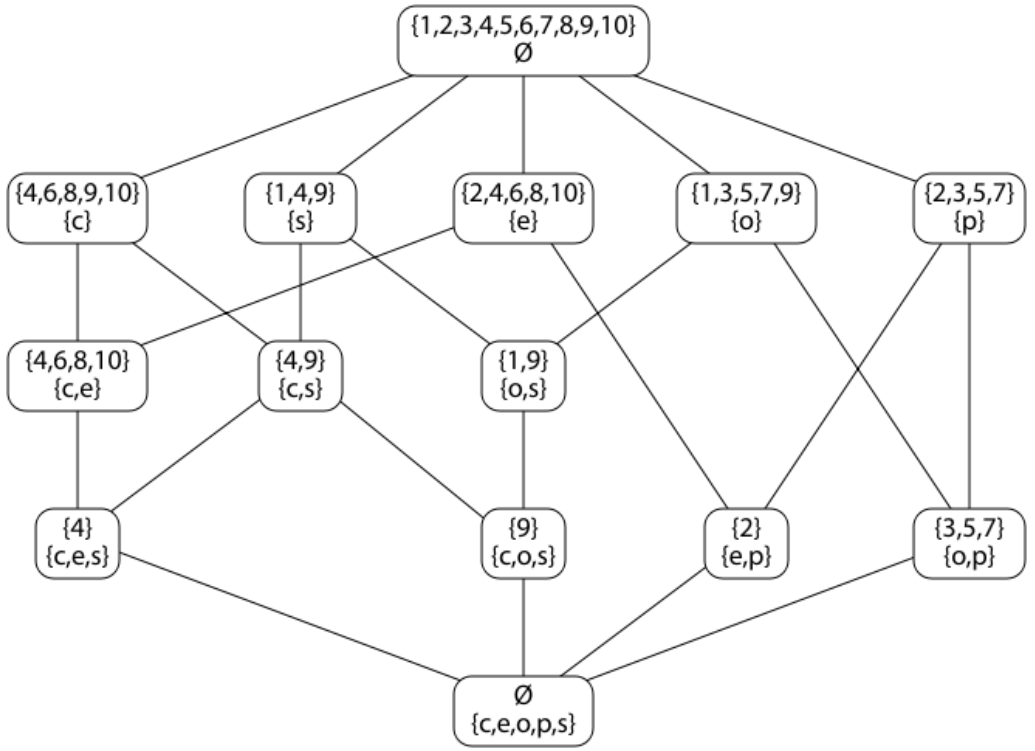
- Data knowledge discovery
  - Request → Data
  - Data → Knowledge → User

Csci 198 project ? ☺



# Example

- "Concept" lattice
  - Objects
    - integers (1 to 10)
  - Attributes
    - composite (c)
    - square (s)
    - even (e)
    - odd (o)
    - prime (p).



# Current view of data

- JSON/XML
  - Nodes of information with a **parent** node
  - Root at the top with **no** parent
- **Data graph**
  - No root, no hierarchy

# Resource Description Framework (RDF)

- **RDF** → part of the W3C specifications
  - Written in XML
- General method for conceptual description of information
  - Implemented in **web resources**
  - Used in knowledge management applications
- A framework for describing and interchanging metadata
  - Data describing the web resources
  - Gives machine understandable semantics for metadata →
    - better precision in resource discovery than full text search
    - assisting applications as schemas evolve
    - interoperability of metadata

# RDF

- Paradigm shift
  - Hierarchical/Relational databases → **Graph** databases
    - Directed, labeled graph
    - The edges represent the named link between two resources
- Expressions (triple)
  - **Subject**
    - Resource (e.g. earth)
      - An URI
  - **Predicate**
    - Aspect of the resource (e.g. is shaped like)
    - Relationship between Subject and Object
  - **Object**
    - (e.g. a sphere)
- OOP
  - earth = object , shape = property, value = sphere

# RDF

- eRDF
  - Embedded RDF (in HTML)
    - The subject of the triple is the current HTML page
    - The object of the triple is the current HTML page
    - The subject is a unique identifier on the current page
    - The object is a unique identifier on the current page
- RDFa markup
  - https://www.w3.org/TR/html-rdfa/#extensions-to-the-html5-syntax
- Turtle
  - A textual representations of an RDF graph
    - https://www.w3.org/TR/turtle/

# RDF

- Examples
  - For the description of
    - Properties of shopping items
    - Time schedules for different events
    - Content and rating for web images
    - Content of search engine
- Syntax and examples
  - RDF_examples.xml

# Conclusion

- To always think about the semantic/meaning of each element
  - Definition of atomic elements representing “data”
    - That can be reused/linked with other elements
- Web semantic
  - A Pandora box?
    - RDF, Turtle, Sparkle…
    - Linked with AI
      - Symbolic - knowledge extraction (Graph, Lattices,…) Case Base Reasoning
      - Machine learning/Pattern recognition for extraction of information from images, data
        - Classification , Clustering …
- **Data mining**
  - Using Semantic Web based approaches, Semantic Web Technologies,
    - + Linked Open Data to support the process of knowledge discovery
  - Using data mining techniques to mine the Semantic Web (Semantic Web Mining)
  - Using machine learning techniques to create and improve Semantic Web data.

# Conclusion

- AI & Machine learning for
  - Adding knowledge to the web content
  - Labelling images → automatic creation of alt=…
  - Connecting elements
    - Text ← → Multimedia files

- Addition of Machine Learning to Web Programing

# Further reading

- Berners-Lee, T., Hendler, J. and Lassila, O.  The Semantic Web, Scientific American, May 2001.
  - http://www.w3.org/2001/sw/
- Gruber, T. R., A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199-220, 1993.
- Gruber, T. R., Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Int. Journal Human-Computer Studies, 43(5-6):907-928, 1995.

- Proceedings of the International Semantic Web Conference

- + Links on Canvas

# Lab of week 16

- Support for the Project
  - ➢Lab session
    - Goal(s) + Target(s) + Problem(s)
      - Web page organization
      - Database queries
      - POST/GET
  - ➢**You must know at what stage you are**
  - ➢What should not happen: