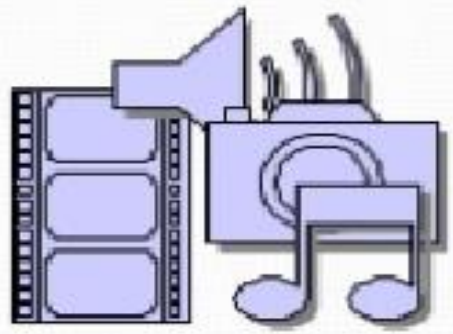
Data Mining vs. Web Mining

- **Traditional data mining**
  - data is structured and relational
  - well-defined tables, columns, rows, keys, and constraints.
- **Web data Mining**
  - Semi-structured and unstructured
  - readily available data
  - rich in features and patterns

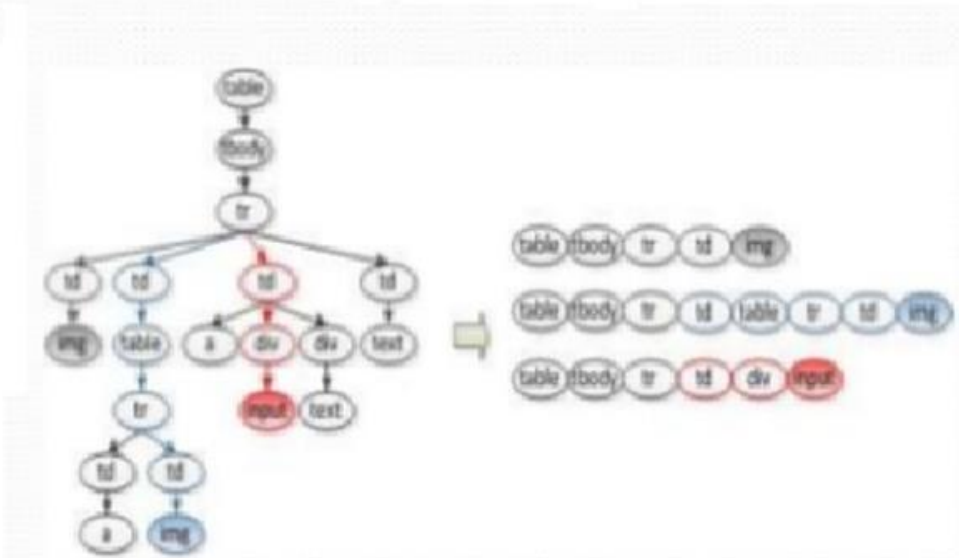


Classification of Web Mining Techniques

Web-Structure Mining

- Web Structure Mining is a tool used to identify the relationship between Web pages linked by information or direct link connection.

- Creating a model of web organization
- Classify web pages
- Create similarity measures between web pages
- Page Rank
- Hyperlink induced topic search(HITS)



Web Content Mining

- '***Process of information***' or resource discovery from content of millions of sources across the World Wide Web
  - E.g. Web data contents: text, Image, audio, video, metadata and hyperlinks
  - It is related to text mining

because much of

the web contents are texts.

Web Content Mining Techniques

Web-Usage Mining

- What is Usage Mining?

  Discovering user '*navigation patterns*' from web data.

  Prediction of user behavior while the user interacts with the web.

- Extends work of basic search engines.
- Performs mining on Web usage data or Web logs.
- Search Engines –
  - IR application.
  - Keyword based.
  - Similarity between query and document.
  - Crawlers.
  - Indexing.
  - Profiles.
  - Link analysis.

Web Crawler

- Program which browses WWW in a methodical, automated manner
- Copy in cache and do Indexing
- Starts from a seed url
- Searches and finds links, keywords
- Types of Crawler
  - Context focused
  - Focused
  - Incremental
  - Periodic