LECTURE NOTES 5th September 2011

Gundlapalli Srinivas 09010166
B.Mani Shyam Patro 09010168

Fifth Normal Form

A relation schema R is said to be in **fifth normal form (5NF)** if for every JD $\bowtie \{R_1, \ldots, R_n\}$ that holds over R, one of the following statements is true:

- $R_i = R$ for some i, or
- The JD is implied by the set of those FDs over R in which the left side is a key for R.

The second condition deserves some explanation, since we have not presented inference rules for FDs and JDs taken together. Intuitively, we must be able to show that the decomposition of R into $\{R_1, \ldots, R_n\}$ is lossless-join whenever the **key dependencies** (FDs in which the left side is a key for R) hold. $\bowtie \{R_1, \ldots, R_n\}$ is a **trivial JD** if $R_i = R$ for some i; such a JD always holds.

The following result, also due to Date and Fagin, identifies conditions—again, detected using only FD information—under which we can safely ignore JD information.

> If a relation schema is in 3NF and each of its keys consists of a single attribute, it is also in 5NF.

The conditions identified in this result are sufficient for a relation to be in 5NF, but not necessary. The result can be very useful in practice because it allows us to conclude that a relation is in 5NF *without ever identifying the MVDs and JDs that may hold over the relation.*

Structured Data:

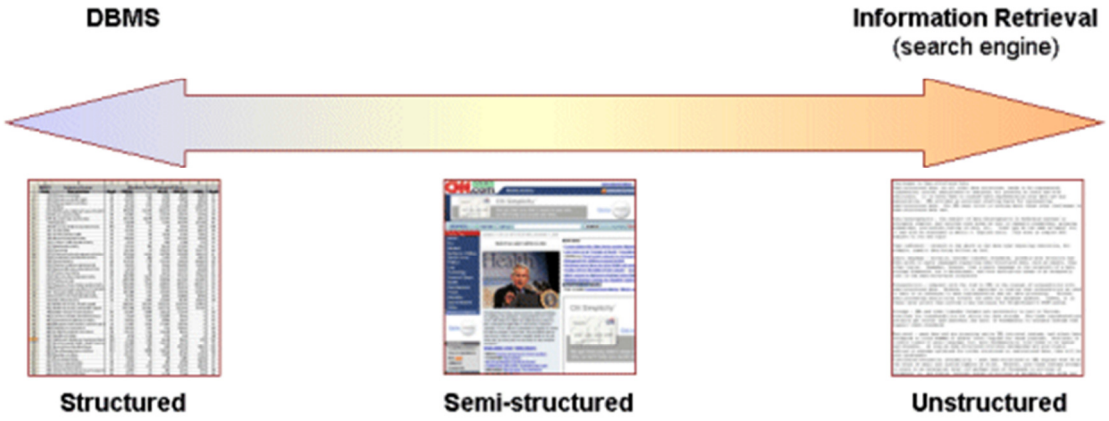
Data is organised into semantic chunks called entities. Entities in the same group have the same descriptions (or attributes), while descriptions for all entities in a group (or schema): a) have the same defined format; b) have a predefined length; c) are all present; and d) follow the same order. Structured data are what is normally associated with conventional databases such as relational transactional ones where information is organized into rows and columns within tables. You might look for all the rows containing a particular date or ZIP code or name – this is structured data, and it is organized and searchable by data type within the actual content. Spreadsheets are another example. Nearly all understood database management systems (DBMS) are designed for structural data.

Semi-structured Data:

Semi-structured data is a form of structured data that does not conform with the formal structure of tables and data models associated with relational databases but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Therefore, it is also known as schemaless or self-describing structure . Semi-structured data are organized in

semantic entities, similar entities are grouped together, entities in the same group may not have same attributes, the order of attributes is not necessarily important, not all attributes may be required, and the size or type of same attributes in a group may differ. To be organized and searched, semi-structured data should be provided electronically from database systems, file systems (*e.g.*, bibliographic data, Web data) or via data exchange formats (*e.g.*, EDI, scientific data, XML).

Semi-structured data is increasingly occurring since the advent of the Internet where full-text documents and databases are not the only forms of data any more and different applications need a medium for exchanging information.



Unstructured Data:

Unstructured Data refers to information that either does not have a pre-defined data model and/or does not fit well into relational tables. In this form, data can be of any type and do not necessarily follow any format or sequence, do not follow any rules, are not predictable, and can generally be described as 'free form'. This results in irregularities and ambiguities that make it difficult to understand using traditional computer programs as compared to data stored in fielded form in databases or annotated (semantically tagged) in documents.
The term is imprecise for several reasons:

- Structure, while not formally defined can still be implied.
- Data with some form of structure may still be characterized as unstructured if its structure is not helpful for the desired processing task, and
- Unstructured information might have some structure (semi-structured) or even be highly structured but in ways that are unanticipated or unannounced.

Examples of "unstructured data" may include books, journals, documents, metadata, health records, audio, video, files, and unstructured text such as the body of an e-mail message, Web page, or word-processor document.

Unstructured information accounts for more than 70%-80% of all data in organizations.

XML

Difference between XML and HTML?

XML was designed to transport and store data.

HTML was designed to display data.

XML is Not a Replacement for HTML

XML is a complement to HTML**.**

It is important to understand that XML is not a replacement for HTML. In most web applications, XML is used to transport data, while HTML is used to format and display the data.

XML is a software- and hardware-independent tool for carrying information.

XML Separates Data from HTML

If you need to display dynamic data in your HTML document, it will take a lot of work to edit the HTML each time the data changes.

With XML, data can be stored in separate XML files. This way you can concentrate on using HTML for layout and display, and be sure that changes in the underlying data will not require any changes to the HTML.

With a few lines of JavaScript code, you can read an external XML file and update the data content of your web page.

XML Simplifies Data Sharing

In the real world, computer systems and databases contain data in incompatible formats.

XML data is stored in plain text format. This provides a software- and hardware-independent way of storing data.

This makes it much easier to create data that can be shared by different applications.

XML Simplifies Data Transport

One of the most time-consuming challenges for developers is to exchange data between incompatible systems over the Internet.

Exchanging data as XML greatly reduces this complexity, since the data can be read by different incompatible applications.

XML Simplifies Platform Changes

Upgrading to new systems (hardware or software platforms), is always time consuming. Large amounts of data must be converted and incompatible data is often lost.

XML data is stored in text format. This makes it easier to expand or upgrade to new operating systems, new applications, or new browsers, without losing data.

XML Makes Your Data More Available

Different applications can access your data, not only in HTML pages, but also from XML data sources.

With XML, your data can be available to all kinds of "reading machines" (Handheld computers, voice machines, news feeds, etc), and make it more available for blind people, or people with other disabilities.

XML is Used to Create New Internet Languages

A lot of new Internet languages are created with XML.

Here are some examples:

- XHTML
- WSDL for describing available web services
- WAP and WML as mark-up languages for handheld devices
- RSS languages for news feeds
- RDF and OWL for describing resources and ontology
- SMIL for describing multimedia for the web