as needed now or in the future. Another distinct advantage is the ability to have one system supports multiple specific schemas in parallel.

2.4 The Contextual Information Component

The Contextual Information Component ("Contextual Component "or "Entity") is another unique aspect of PrXML. Almost every document is a collection of different linked ideas. Contextual Components are logical units (or atoms) of information within a given document that define these ideas. The rules that expose the Contextual Components in a given documents are defined by the information owner or by the specific content application. For example, a component can be an article in a journal or newspaper, a picture in a magazine or even a section or a paragraph in a report.

Contextual Component extraction and tagging enables management and direct access to any piece of information within a document, independent of other components instead downloading large files. While at the same time, information is always maintained in the context of its original document.

Information Components improve the search engine indexing process, thus increasing the ~~relativity~~ rankings of search results. For example; If a user searches all articles that are related to the key words "food" and "oil" and "olive". Regularly he will get all of the PDF documents that include such key words even if they are not related one to each other, (for example a page that consists of one article about "Olive trees and gardening" and anther article that talks about "Oil fields in Iraq".) With Information Component based searches, the user will get search results only on articles or advertisements (components) that include the specific key words.

2.5 The PrXML Structure

Contextual Components (Entities) also contain their own internal structure and are defined as "Elements". For example; in the case of a newspaper, if an entity is defined as an article then its elements might be a title, headline and byline etc. In the case of a book, an entity might be a chapter so its elements would be a chapter title, body and graphs. An element is still a unit of information but it takes a collection of elements together to make a complete Entity.

A further breakdown of the structure includes the element's "nuclear elements"—called "Primitives". Primitives are the fundamental building blocks of the PrXML schema. A Primitive is a defined rectangular region on a page that contains text or graphic information such as, a word, map or paragraph. Articles, chapters and other elements are made up of a collection of rectangular Primitives.

The metadata tags in each Entity's XML, collects knowledge about the Entity's environment and its relationships to other components; information like the section the Entity belongs to, the page it is on, its size, the articles it is related to, etc. Each word XML tag contains attributes such as page coordinates and font style information (supplied as *Font Style Gallery* references).