

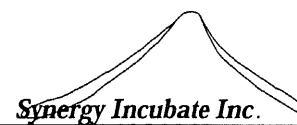


# SGML/XML

## *Re-Inventing the WEB*

8 Sept. 1998

**Masahiro Kikuta**  
*Synergy Incubate Inc.*

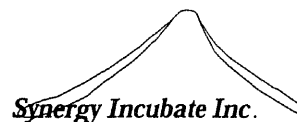


## ***What is happening on the NET?***

### **✓ The extraordinary growth of the World-Wide Web**

#### **<Key Technology>**

- HTTP: Hypertext Transfer Protocol
- URL: Uniform Resource Locator
- HTML: Hypertext Markup Language



## **HTTP**

### ✓ **HTTP. Hypertext Transfer Protocol**

- The client-server TCP/IP protocol used on the World-Wide Web for the exchange of HTML documents.

*Synergy Incubate Inc.*

## **URL**

### ✓ **URL:Uniform Resource Locator**

- A draft standard for specifying an object on the Internet, such as a file or a newsgroup. URLs are used extensively on the World-Wide Web. They are used in HTML documents to specify the target of a hyperlink.

*Synergy Incubate Inc.*

DR

# HTML

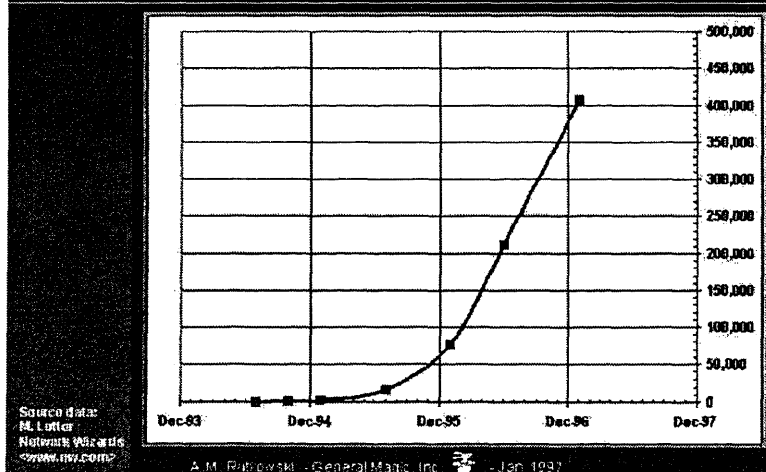
## ✓ HTML: Hypertext Markup Language

- A Hypertext document format used on the World-Wide Web. Built on top of SGML. "Tags" are embedded in the text. A tag consists of a "<", a "directive" (case insensitive), zero or more parameters and a ">". Matched pairs of directives, like "<TITLE>" and "</TITLE>" are used to delimit text which is to appear in a special place or style.

Synergy Incubate Inc.

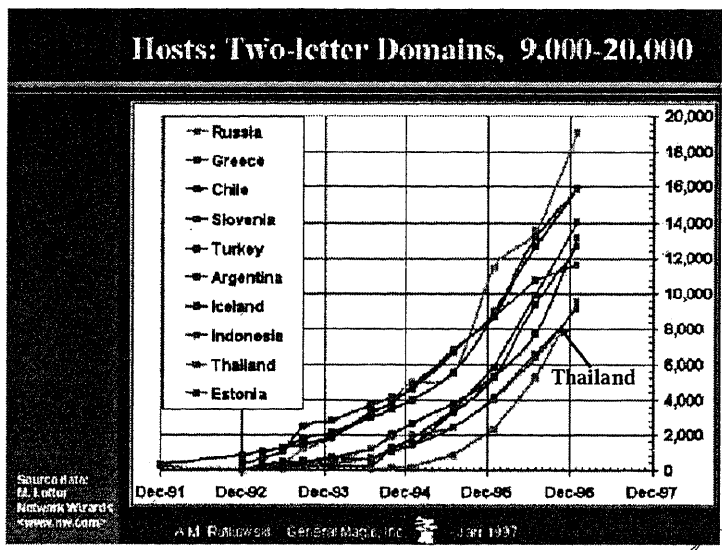
DR

## WWW-Prefixed Hosts



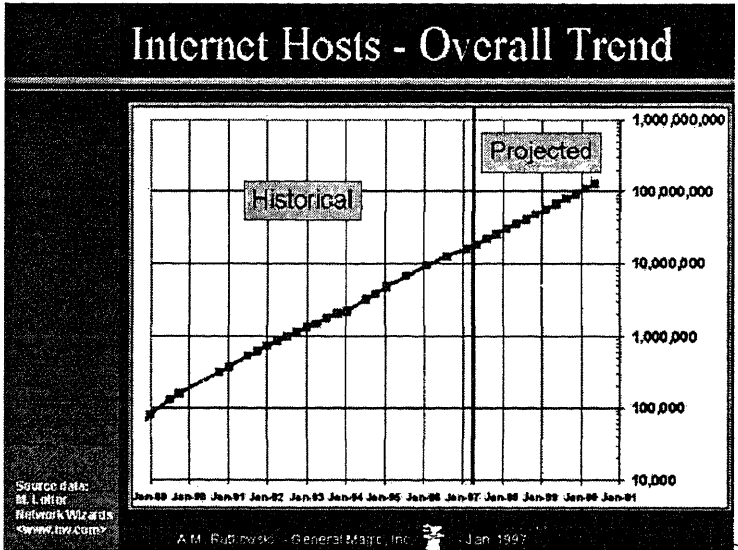
Synergy Incubate Inc.

DR



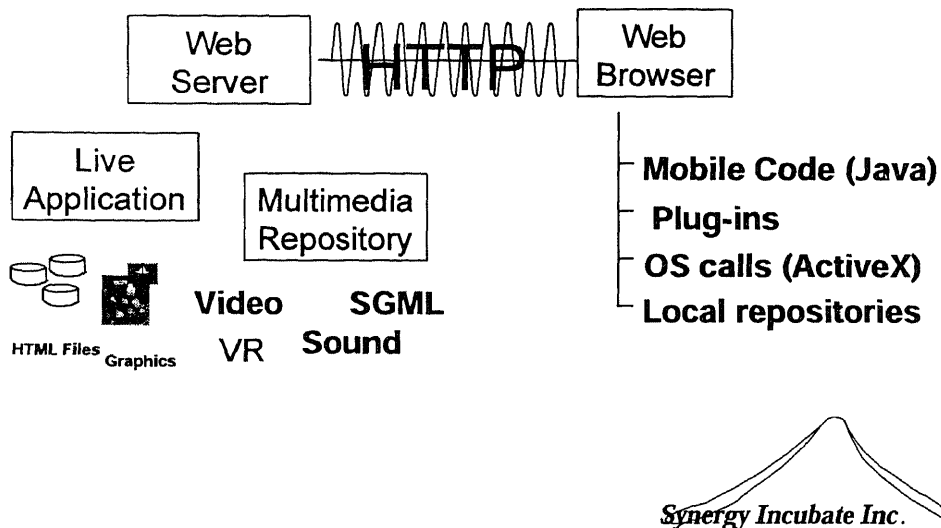
Synergy Incubate Inc.

DR



Synergy Incubate Inc.

## ***The World-Wide Web: Multidimensional Multimedia***



## ***What's wrong with HTML?***

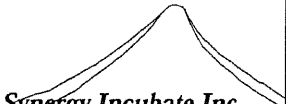
- ✓ **HTML was optimized for easy learning**
  - One tag set for all applications
  - Predefined semantics for each tag
  - Predefined data structures
  - No formal validation
- ✓ **HTML trades power for ease of use**

*Synergy Incubate Inc.*

..... 

## ***What's wrong with HTML?***

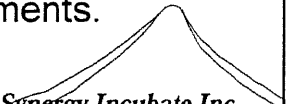
- ✓ **HTML is well suited to simple applications,**
- ✓ **But poorly suited to more demanding applications**
  - Large or complex collections of data
  - Data that must be used in many ways
  - Data with a long life cycle

  
*Synergy Incubate Inc.*

..... 

## ***SGML offers,***

- ✓ **Extensibility**
  - Authors can define new tag names and attribute names for documents by specifying their their syntax and semantics.
- ✓ **Structure**
  - Documents can be containers for other documents, with arbitrary nesting.
  - This allows complex documents to be constructed from simpler documents.

  
*Synergy Incubate Inc.*

.....  
JK

## ***SGML offers,***

### **✓ Validation**

- If desired, any SGML document can reference a description of its grammar so applications can validate that the document conforms to its specified structure.

IEEE Internet Computing  
By Rohit Khare/ Adam Rifkin

  
*Synergy Incubate Inc.*

.....  
JK

## ***SGML can be***

- ✓ Safest of all formats for high-value data (ISO standard 8879)
- ✓ Best available format for search/retrieval.
- ✓ Easily converted to HTML.

**But.....**

  
*Synergy Incubate Inc.*



## ***What' wrong with SGML***

- ✓ Designed before the PC revolution.
- ✓ Large, cumbersome specification (500pages)
- ✓ Difficult for people to read and understand
- ✓ Difficult for computers to process and manipulate
- ✓ High entry cost

*Synergy Incubate Inc.*



## ***Motivation to XML***

- ✓ SGML has been here since 1986, and
- ✓ Many of us think it's a good idea, but
- ✓ It hasn't caught on that fast, and
- ✓ The Web is here, and
- ✓ We need it even more than ever!

*Synergy Incubate Inc.*





## ***What is XML?***

- ✓ **Next step in Web evolution**
- ✓ **Goes beyond the limitations of HTML**
- ✓ **Will create new Web applications**
  - database exchange
  - Distribution of processing to clients
  - Client-side manipulation of views into the data
  - Customization of information by intelligent agents
  - Management of document collections

*Synergy Incubate Inc.*



## ***Why XML?***

- ✓ **For a new generation of Web applications:**
  - ***Extensibility:*** Users can define new tags as needed
  - ***Structure:*** Hierarchical data can be modeled to any level of complexity
  - ***Validation:*** Data can be checked for structural correctness
  - ***Media independence:*** The same content can be published in multiple media

*Synergy Incubate Inc.*



## ***XML Design Goals***

- 1.XML shall be straightforwardly usable over the Internet**
- 2.XML shall support a wide variety of applications**
- 3.XML shall be compatible with SGML**
  - Existing SGML tools read & write XML
  - XML documents are SGML documents
  - The same parse can be generated
  - Similar expressive power

*Synergy Incubate Inc.*



## ***XML Design Goals***

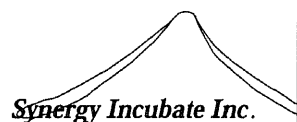
- 4.It shall be easy to write programs which process XML documents.**
- 5.The number of optional features in XML is to be kept to the absolute minimum, ideally zero**
- 6.XML documents should be human-legible and reasonably clear**

*Synergy Incubate Inc.*



## ***XML Design Goals***

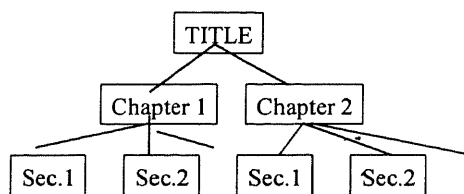
- 7. The XML design should be prepared quickly**
- 8. The design of XML shall be formal and concise**
- 9. XML documents shall be easy to create**
- 10. Terseness is of minimal importance**



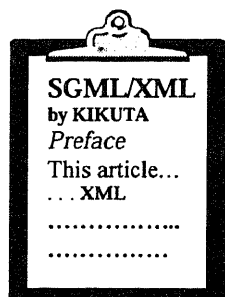
## ***Document is composed of***

These slides are designed for XML tutorial seminar in IJWDL '98,.....

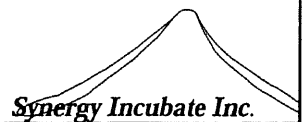
**CONTENT**



**STRUCTURE**

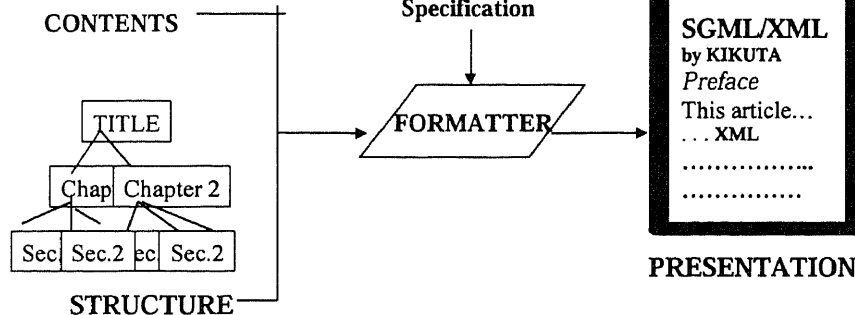


**PRESENTATION**



## XML allow us:

These slides are designed for XML tutorial seminar in IJWDL'98,.....



XML specifies the content and structure of document in a way that allows particular presentation.

*Synergy Incubate Inc.*

## Relationship to SGML

- ✓ XML is a simplified subset of SGML:
  - Powerful
  - No limits on namespace of structural depth
  - Easy to implement
  - Small enough for Web browsers
- ✓ The translation from SGML to XML is straightforward

*Synergy Incubate Inc.*



## ***What is the role of XML?***

### **✓ SGML**

- For large publication systems
- High-value information

### **✓ XML**

- For structured delivery over the web
- In many cases will be adequate for authoring

### **✓ HTML**

- For delivery over the web
- Low-value information

*Synergy Incubate Inc.*



## ***Advantages of XML***

- ✓ **Conformant with existing international standard (SGML)**
- ✓ **Complete extensibility -- no tag limitations**
- ✓ **Full internationalization**
- ✓ **Validation and editorial control**
- ✓ **Ability to model any kind of hierarchical data**

*Synergy Incubate Inc.*

## ***Advantages of XML***

- ✓ **Automatic generation of links and navigational aids**
- ✓ **Increased speed of access to essential data**
- ✓ **Print and online versions from same source**
- ✓ **Dynamically user-configurable views**
- ✓ **Next-generation hypertext capabilities**

*Synergy Incubate Inc.*

## ***The XML family***

- ✓ **XML (eXtensible Markup Language)**
  - A subset of SGML(ISO 8879) designed for easy implementation
- ✓ **XLL(eXtensible Linking Language):**
  - A set of standard hypertext mechanisms based on HyTime(ISO/IEC 10744)and the Text Encoding Initiative(TEI)

*Synergy Incubate Inc.*

## ***The XML family***

### **✓ XSL(eXtensible Stylesheet Language)**

- A standard stylesheet language for structured information formed by subsetting DSSSL (ISO/IEC 10179), designing an alternative syntax, and incorporating key CSS concepts

*Synergy Incubate Inc.*

## ***Industry specific tag language***

- ✓ CML:Chemical Markup Language
- ✓ CDF:Channel Definition Format
- ✓ OFX:Open Financial Exchange
- ✓ HDML:Handheld Device Markup Language
- ✓ MathML:Mathematics Markup Language
- ✓ PGML:Precision Graphics Markup Language

*Synergy Incubate Inc.*



## ***The XML applications can***

- ✓ use Web clients to mediate between multiple heterogeneous databases
- ✓ distribute the load from Web servers to their clients
- ✓ use Web clients to present different views of the same data
- ✓ employ agents to tailor information discovery and filtering to the customized needs

*Synergy Incubate Inc.*



## ***User view of the data***

- ✓ **Different views of whole documents**
  - Novice view vs. expert view
  - Outline vs. content
  - Generated Tables of Contents
- ✓ **Different views of document components**
  - Bar graph vs. pie chart
  - Totals by region vs. totals by business units

*Synergy Incubate Inc.*




.....



## ***XML require additional mechanisms***

- ✓ **Programs, applets, or scripts designed for a specific tag set**
- ✓ **Industry agreements on the processing of specific tags**
- ✓ **Tag-sensitive components.**
- ✓ **Stylesheets**



*Synergy Incubate Inc.*

.....



## ***The future of the Web***

- ✓ **View Selection: Letting The User Decide**
- ✓ **Web Agents: Data That knows About Me**
- ✓ **Users no longer tied to a proprietary format**
- ✓ **An end to domination of the market by a few big companies**
- ✓ **An end to domination of the market by a few big countries**

**Jon Bosak** Chairman of XML WG, W3C



*Synergy Incubate Inc.*