A new way to think about text

- structure not formatting
- portable representation
- enhanced content

What it is

- Markup in plain ASCII
- ISO 8879:1986
- 332 definitions; 70 basic
- history: GML
- DTD, Entity, Element

History

- Specific Coding (format-17)
- late 1960's: Generic Coding (header)
- 1969: GML (IBM, Goldfarb, Mosher, Lorie) for law offices
- Invented 1974 by Charles F. Goldfarb
- 1978: ANSI committee
- 1983: GCA standard
- 1985: final ISO draft
- 1987: DoD CALS

Acronyms

- SGML - Standard Generalized Markup Language
- DSSSL - Document Style Semantics Specification Lang.
- DTD - Document Type Description
- CALS - Computer-aided Acquisition and Logistic Support
- ISO - is *not* an acronym

What you can do with it

- smart editing
- automatic generation
- databases
- typed searching
- output processing

What it looks like

<tag> </tag>
<tag attr=value>
< & >

sgml-lecture.sgml

```
<!doctype slideshow SYSTEM "slides.dtd">
<slideshow>
  <title>Introduction to SGML</title>
  <author>Mark Eichin, SIPB</author>
  <date>IAP 1998</date>
  <slide>
  <header>A new way to think about text</header>
  <bullet>structure not content</bullet>
  </slide>
```

slides.dtd

```
<!ENTITY lt     SDATA "<">
<!ENTITY gt     SDATA ">">
<!ENTITY amp    SDATA "&">
<!element slideshow o o (title?, author?, date?, slide*)>
<!element (header|title|author|date) - - (#PCDATA)>
<!element slide - - (header, (bullet|subbullet|url|raw|example)+)>
<!element bullet - - (#PCDATA)>
<!element subbullet - - (#PCDATA)>
<!element url - - (#PCDATA)>
<!element raw - - (#PCDATA)>
<!element example - - (#PCDATA)>
```

sgmlspm

- perl interface to add code to tags

```
sgml('<SLIDE>', sub {  &inittag;
  output "%%Page: $pageno\n";
  output "($savetitle) showtitle\n";
  output "($pageno) showpageno\n"; $pageno++; });
sgml('</SLIDE>', sub {
  output "finishpage\n";
});
```

HTML

- Simple DTD - 70 elements
- historically, no invalid pages
- comment syntax

<-- comment -- -- another one -->

Docbook

`http://www.oreilly.com/davenport/`

- Software Documentation
- Very rich (400+ tags)
- Industry Consortium

Other DTDs

- QWERTZ
- sgml-tools/linuxdoc
- TEI (Text Encoding Initiative)

DSSSL

- R4RS Scheme
- Document Style Semantics Specification Language
- ISO/IEC 10179:1996

XML (Extensible Markup Language)

- W3C effort
- Stripped down SGML
- XML parser in java
- Web Spiders/Robots
- Microsoft CDF (Channel Definition Format)
- XSL (Extensible Stylesheet Language)

Athena tools

`add sgml`

- jade, nsgmls, stylesheets

Commercial Tools

- Frame 5.5+SGML (?)
- WordPerfect 8
- OmniMark (LE)

Other Stuff

- HyTime: ISO 10744
  Hypermedia/Time-based Structuring Language
- mathematics
- MIME-SGML
- OCLC, LOC
- Chemistry, Physics, Astronomy

Pointers

`http://www.mit.edu/iap/iap-sgml.html`

`http://www.sil.org/sgml/sgml.html`

`ftp://ftp.lysator.liu.se/pub/sgml/psgml-1.0`

`http://www.jclark.com/jade/`

- comp.text.sgml