

# A history of markup languages

By Tibs / Tony Ibbs

This version for the Cambridge Write The Docs meetup, Feb 2018

Written using reStructuredText.

Converted to PDF slides using pandoc and beamer.

Source and extended notes at <https://github.com/tibs/markup-history>

# Timeline

- ▶ 1960s TYPSET and RUNOFF, GML
- ▶ 1970s roff, runoff, nroff/troff, T<sub>E</sub>X in SAIL
- ▶ 1980s Scribe, T<sub>E</sub>X in WEB/Pascal, L<sup>A</sup>T<sub>E</sub>X, SGML, TEI
- ▶ 1990s groff, HTML, setext, Docbook, WikiWikiWeb, StructuredText, XML
- ▶ 2000s reStructuredText, AsciiDoc, markdown

# The types of markup

Presentational or Semantic

...but also lightweight, and maybe programmable

# 1964: RUNOFF

.LINE LENGTH 60

.LEFT MARGIN 0

.PARAGRAPH 5

Call us on our toll free number

.CENTER

1-800-555-5555

and we will respond as soon as convenient.

## 1969: GML and 1986: SGML

**<td>** The Implication of SGML for the Preparation of Scientific Publications

**<au>** Joan M. Smith

**<ad>** **<al>** National Computing Centre, Oxford Road, Manchester M1 7ED

**<ab>** The &SGML (SGML) is a draft international standard for publishing.

...

**<h1>**Introduction

**<p>** The official title of SGML, currently, is ISO/DIS 8879,

**<ci>** Information Processing &end Text and Office Systems &end &SGML (SGML)

**</ci>**. **<ref>** ISO/DIS 8879 **<ci>** Information Processing &end Text and Office Systems &end &SGML (SGML). ISO, Geneva (1985). **</ref>**

...

**<p>**There are several points worthy of note here:

**<ul>**

**<li>** the normal publishing delay with ISO standards...

...

**</ul>**

# SGML DTD

*DTD for a list:*

```
<!--      ELEMENT MIN CONTENT      >
<!ELEMENT list      - - (item)+      >
<!ELEMENT item      0 0 (#PCDATA, (list)*) >
```

*and such a list:*

```
<list>
<item>First item</item>
<item>Second item</item>
<item>Last item</item>
</list>
```

1997: XML

*"XML is an application profile of SGML"*

## 1970s: roff, nroff, troff

.TH CORRUPT 1

.SH NAME

corrupt \- modify files by randomly changing bits

.SH SYNOPSIS

.B corrupt

[\fB\-n\fR \fIBITS\fR]

[\fB\-\-bits\fR \fIBITS\fR]

.IR file ...

.SH DESCRIPTION

.B corrupt

modifies files by toggling a randomly chosen bit.

.SH OPTIONS

.TP

.BR \-n " , " \-\-bits =\fIBITS\fR

Set the number of bits to modify. Default is one bit.



## 1990: groff

```
..INCLUDE mission-statement-strings.mom
.TITLE      "\*[Groff-Mission-Statement]
.SUBTITLE   "\*[2014]
.INCLUDE    mission-statement-style.mom
.PP
```

As the most widely deployed implementation of troff in use today, groff holds an important place in the Unix universe. Frequently and erroneously dismissed as a legacy program for formatting Unix manuals (manpages), groff is in fact a sophisticated system for producing high-quality typeset material, from business correspondence to complex, technical reports and plate-ready books. \\*[BU3]With an impressive record for backward compatibility, it continues to evolve and play a leading role in the development of free typesetting software.

1977/1978: T<sub>E</sub>X

`\name{Name Redacted}` wrote:

`\beginletter`

Thoughts on ``Why I like children's books':

`\beginlist`

`\item{\blob}` They aren't afraid to show a sense of wonder.

`\item{\blob}` They aren't `duty bound' to include love interest for the sake of it.

`\item{\blob}` They are rarely cynical, rarely bitter---but the best do not avoid tragedy and truth.

`\item{\blob}` They are willing to teach the simple lessons of being human---which adult books so often scorn, but which we all need to learn and relearn.

`\endlist`

1983: L<sup>A</sup>T<sub>E</sub>X

```
\begin{center}  
\rule{5in}{0.1mm}  
\end{center}
```

```
\section*{Captain Competent strikes again}
```

The superhero is a familiar concept in comics, science fiction and many other fields. However, I am more interested in what might be called 'the competent hero'. This is a subtler form of protagonist---a person who has attained {\em competence} in their daily life.

## 1980: Scribe

```
@Heading(The Beginning)
```

```
@Begin(Quotation)
```

```
    Let's start at the very beginning, a @i(very good  
    place) to start
```

```
@End(Quotation)
```

*which can also be written:*

```
@Heading(The Beginning)
```

```
@(Quotation
```

```
    Let's start at the very beginning, a @i(very good  
    place) to start
```

```
)
```

## 1987: TEI

```
<lg type="sestina">
<lg type="sestet" rhyme="ababab">
<l>I saw my soul at rest upon a
      <rhyme label="a" xml:id="A">day</rhyme></l>
<l>As a bird sleeping in the nest of
      <rhyme label="b" xml:id="B">night</rhyme>,</l>
<l>Among soft leaves that give the starlight
      <rhyme label="a" xml:id="C">way</rhyme></l>
<l>To touch its wings but not its eyes with
      <rhyme label="b" xml:id="D">light</rhyme>;</l>
<l>So that it knew as one in visions
      <rhyme label="a" xml:id="E">may</rhyme>,</l>
<l>And knew not as men waking, of
      <rhyme label="b" xml:id="F">delight</rhyme>.</l>
</lg>
```

## 1991: HTML

```
<!DOCTYPE html>
<html>
  <head>
    <title>This is a title</title>
  </head>
  <body>
    <p>Hello world!</p>
  </body>
</html>
```

## 1991: Docbook

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE article PUBLIC "-//OASIS//DTD Simplified DocBook XML V1.0//EN"
"http://www.oasis-open.org/docbook/xml/simple/1.0/sdocbook.dtd">
<article>
  <title>DocBook Tutorial</title>
  <articleinfo>
    <author>
      <firstname>Adrian</firstname> <surname>Giurca</surname>
    </author>
    <date>April 5, 2005</date>
  </articleinfo>
  <section>
    <title>What is DocBook ?</title>
    <para>DocBook is an SGML dialect developed by O'Reilly and HaL
      Computer Systems in 1991.</para>
  </section>
</article>
```

## 1991: setext

This is the title. There can be only one.

=====

Body text must be indented by two spaces.

A subheading

-----

**Bold words** and *italic* are supported. Underlined words are also supported. `Backquoted words` are not touched.

> This text will be represented using a monospaced font.

\* This text will have a bullet mark before it.

*.. Two dots introduce text that can be ignored.*

*.. Two dots alone mean the logical end of text.*

..



## 1994/1995: wikiwikiweb

Paragraphs are not indented.

- \* This is a list item

- \*\* This is a sub-list item

Indented text is monospaced.

We have `'emphasis'`, `'''bold'''`, `''''bold italic''''`, and a `LinkToAnotherPage`.

But we can `A''''''voidMakingAWikiLink`.

No HTML, tables, headers, maths, scripts.

No links within a page.

## 1996: StructuredText

This is a heading

This is a paragraph. Body text is indented.

- This is a list item. Words can be *\*emphasized\**, *\_underlined\_*, ***\*\*strong\*\**** or 'inline' - yes, that's using single quotes [1].

o This is a list item as well.

This is a sub-heading

Sub-section body text is indented even further. This indented body text makes the sub-heading a heading.

.. [1] Or we could use ``backquotes``.

## 2001/2002: reStructuredText

This is a heading

=====

This is a paragraph. Body text is not indented.

- This is a list item. Words can be *\*emphasized\**, **\*\*strong\*\*** or ``teletype`` - yes, that's paired backquotes [1]\_.
- This is a list item as well.

This is more of the second list item. It is indented appropriately.

This is a sub-heading

-----

Sub-section body text is not indented either.

.. [1] Note the indentation inside the list item.

## 2002: AsciiDoc

= This is a title heading

This is a paragraph. Body text is not indented.

- This is a list item. Words can be *\_italic\_*, **\*bold\*** or +mono+ - yes, that's paired plus-signs.

- This is a list item as well.

+

This is more of the second list item. It is "`joined on`" by the  
`+`.footnote:[Note the quotation marks around *\_joined on\_*.]

== This is a sub-heading

Sub-section body text is not indented either.

## 2004: markdown

# This is a heading

This is a paragraph. Body text is not indented.

- This is a list item. Words can be *\*emphasized\**, **\*\*strong\*\*** or ``inline`` - that's single backquotes.
- This is a list item as well.

    This is more of the second list item. Its first line must be indented by 4 spaces or a tab.

## This is a sub-heading

Sub-section body text is not indented either.

(No footnotes. But `<tt>HTML</tt>` is allowed.)

# Fin

- ▶ 1960s TYPSET and RUNOFF, GML
- ▶ 1970s roff, runoff, nroff/troff, T<sub>E</sub>X in SAIL
- ▶ 1980s Scribe, T<sub>E</sub>X in WEB/Pascal, L<sup>A</sup>T<sub>E</sub>X, SGML, TEI
- ▶ 1990s groff, HTML, setext, Docbook, WikiWikiWeb, StructuredText, XML
- ▶ 2000s reStructuredText, AsciiDoc, markdown

Written using reStructuredText.

Converted to PDF slides using pandoc and beamer.

Source and extended notes at <https://github.com/tibs/markup-history>