FORMATTING SGML MANUSCRIPTS

- or -

MUCH ADO ABOUT NOTHING

UKTUG and BCS-EPSG meeting

(c) Copyright 1995

Jonathan Fine
203 Coldhams Lane
Cambridge
CB1 3HY

SPACES BETWEEN WORDS

The notes I handle no better than many pianists. But the pauses between the notes — ah, that is where the art resides.

Artur Schnabel (1885–1951)

Basic to quality are the spaces between words, the breaking of text into lines, and the breaking of lines into pages.

Another basic is the space separating vertically stacked elements.

Designers understand such things.

When the spacing between words or elements is wrong, there is no remedy to make the result good.

WHO OWNS WHAT?

The words belong to the author.

The spaces between the words belong to the formatter (which should be person and program working in harmony).

Each participant needs to respect the others.

Punctuation is a battlefield. It is neither word nor space, but shares qualities with both. Consider:

- Rules of style.
- Quote marks versus quote font.
- Depends on language.
- Interaction between space, punctuation and change of font.

MANUSCRIPT PROBLEMS

Authors are not yet always perfect. Just because it parse without error, that doesn't mean it is without error.

Should the messages

```
<mess>Hello world!</>
<mess> Hello world! </>
<mess> Hello  world ! </>
```

be formatted identically? And how should

```
one <bold> two </>
buckle <bold>my</> shoe
three<bold>four</>
close<bold> the </>door
```

be formatted? Is it the author, parser or formatter who fixes such problems?

WHAT IS TEX THE PROGRAM?

Between 1978 and 1982 the eminent Professor Donald Knuth of Stanford University wrote a very high quality typesetting program called TEX. Its source code is published as a book.

Low cost (or even free) versions are available for most machines, and they run identically. TEX is batch not WYSIWYG, and is programmable via macros.

Sometimes TEX can mean the entire system of fonts, macros and other software — and sometimes it means an input file syntax.

LaTEX is a popular TEX macro package with its own input file syntax.

WHAT IS SIMSIM?

SIMSIM is a TEX macro package which understands SGML. It is a platform upon which style files for formatting SGML manuscripts can be developed.

SIMSIM will run on PCs, Macintosh, Sun, UNIX, VMS and any other machine which supports TEX such as Acorn, Amiga, Alpha, and Atari. SIMSIM is truly portable software.

SIMSIM is the modern Arabic form of the Akkadian word for what we call sesame. Some words are ancient beyond our knowledge. They express our common human heritage.

SIMSIM has a magic power to remove obstacles and open doors.

THE FLAVOUR OF SIMSIM

The SGML declarations

```
<!ELEMENT par ANY>
<!ATTLIST par
        font (rm|bf|it) rm >
```

together with the code

```
def (par)  // links to <par>
{
  paragraph
  {
    // parameters go here
  }
    (par|font)  // attribute
}
def (par*rm)  // name token
// ... etc
```

tell SIMSIM what to do.

FIVE IMPORTANT QUESTIONS

When can I get SIMSIM?

I'm working on it! Hope for the first test release within months. This depends on clients' non-SGML requirements. Any takers?

Can SIMSIM do tables and math?

Yes. SIMSIM can be made to do anything TEX can do.

Does SIMSIM implement all of SGML?

No. For markup minimization, validation etc., use with a parser.

Is SIMSIM compatible with LaTEX?

Is LaTEX compatible with SGML?

What will it cost?

SIMSIM has been five years in the making