

SGML Entity Management

Manoj Srivastava <srivasta@debian.org>

17 February 2002

Abstract

This document provides a guidelines for implementation dependent Entity management of SGML entities for Debian systems. This defines the mapping of *External Public Identifiers* to *System Identifiers*. (In other words, this document covers the use of `/usr/lib/sgml`, entity naming, and catalog files.

Copyright Notice

Copyright ©1998 Manoj Srivastava

This manual is free software; you may redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2, or (at your option) any later version.

Contents

1	Introduction and Scope	1
2	Proposed mapping of public identifiers to system identifiers	3
3	Location of miscellaneous files	5
4	Examples	7

Chapter 1

Introduction and Scope

This document was written by Manoj Srivastava <srivasta@debian.org> with contributions from Mark W. Eichin <eichin@kitten.gen.ma.us> and Adam P. Harris <aph@debian.org>. This document is part of the `sgml-base` package.

This guideline is intended to be interpreted as SGML sub-policy (not official policy). While this document does not carry the weight of official policy, it is sufficient basis for the submission of bugs against a package. This may change at a latter date.

Entity Management is generally left up to the implementation, and hence there are no extant Standards that cover this. However, lack of an established convention would prevent different segments of the SGML subsystem from co-operating with each other, hence it is important that a policy be established so that SGML package maintainers may depend on other parts of the system behaving consistently.

Chapter 2

Proposed mapping of public identifiers to system identifiers

SGML can refer to an external file (really an entity) with an *external identifier*: this is a *public identifier* or a *system identifier*, or both.

A typical public identifier looks like

```
PUBLIC "ISO 8879-1986//ENTITIES Added Latin 1//EN"
```

where ISO 8879-1986 is the owner, ENTITIES is the text class and Added Latin 1 is the text description, and EN is language.

A system identifier looks like

```
SYSTEM "htmlplus.dtd"
```

where `htmlplus.dtd` is a system-specific identifier.

To map external identifiers to file names, one should first try the system identifier, as a file name, and then search entity catalog files and then search the list of file names derived from the public identifier. The catalog format is according to SGML/Opens resolution on entity management. The catalog consists of a series of entries and comments. A comment is delimited by `--` like in a markup declaration.

The fallback derivation of the file name is modelled after the `sgmls` environment variable `SGML_PATH`, and Emacs `psgml` mode's `sgml-public-map` variable. There does not seem to be any official standards (this is left to the implementation), so this standard is simply an abstraction of real-world practice of SGML tools, and shall now be the standard for Debian systems, since this is the convention currently followed by all applications currently in Debian.

Contiguous white space is compacted to a single space and replaced with an underscore (_); the characters / to % are also replaced with _. The text class is down-cased. The language specifier (i.e., //EN) and anything following it should be removed.

Chapter 3

Location of miscellaneous files

There are a number of other files, though not entities referenced by Document instances, are still required by the SGML subsystem to parse or validate the document. These files are also covered by this document.

1. Declaration Files: Any declaration file should be put in `/usr/lib/sgml/declaration`
2. Notations: These files go in `/usr/lib/sgml/notation`.

Chapter 4

Examples

A few public and system identifiers pairings are shown below.

```
PUBLIC "ISO 8879-1986//ENTITIES Added Latin 1//EN"  
/usr/lib/sgml/ISO_8879-1986/entities/Added_Latin_1
```

```
"ISO 8879-1986//ENTITIES Added Math Symbols: Arrow Relations//EN"  
/usr/lib/sgml/ISO_8879-1986/entities/Added_Math_Symbols:_Arrow_Relations
```

```
"-//IETF//DTD HTML Level 3//EN//3.0"  
/usr/lib/sgml/IETF/dtd/HTML_Level_3.0
```

```
"-//IETF//DTD HTML Strict Level 3//EN"  
/usr/lib/sgml/IETF/dtd/HTML_Strict_Level_3
```

```
"-//USA-DOD//DTD Table Model 951010//EN"  
/usr/lib/sgml/USA-DOD/dtd/Table_Model_951010
```

The first four actually exist in Debian.