Geneva Document Specification

Interstellar Ventures

Saturday, 19 September 2015

Table of Contents

This is a draft standard, this notice will disappear once the specification is final.

A *Geneva document* is an ordered collection of *elements*. Geneva defines the following *element types*:

- Pargraph

- Listing

- Table

- Plaintext

- Media

- Section

1   Rich Text

A central component of all element types is *rich text*. Rich text is defined as a sequence of *text tokens*, each made up of a variable number of character strings and an attribute to signify its appearance. There are five different types of text tokens:

Token	Description
plain *s*	Render *s* in regular font.
bold *s*	Recommends to render *s* in bold font.
italic *s*	Recommends to render *s* in italic font.
fixed-width *s*	Recommends to render *s* in fixed-width font.
url *s*	Interpret *s* as a *Uniform Resource Locator*.
url *s*, *u*	Interpret *u* as a *Uniform Resource Locator* and *s* as its label.

Table 1. Text token types.

The occurrence of *whitespace characters* in text token strings is restricted by the following rules:

- All whitespace character sequences are to be reduced to a single *space character* (ASCII `0x20` or equivalent).

- For all token types except the plain type, discard prefixes and suffixes of whitespace character sequences.

- For the first and last text tokens in a rich text sequence, discard prefixes and suffixes of whitespace character sequences respectively.

At least the following conceptual characters have to be recognized as whitespace:

- *Space*

- *Tab*

- *Newline* (including *Carriage Return*)

- *Vertical Tab*

- *Page break*

2   Element Types

A **paragraph** consists of exactly one rich text sequence. It signifies a self-contained piece of text.

A **listing** consists of a finite sequence of rich text sequences. It signifies an ordered group of self-contained text pieces.

A **table** consists of a two-dimensional matrix of rich text sequences and a single rich text sequence being its description. It signifies a tabular relation of the matrix of rich text pieces.

A **plaintext** element consists of a verbatim character string and a single rich text sequence being its description. It signifies a sequence of characters which has to be preserved as is except for whitespace prefixes and suffixes (including newlines).

A **media** element consists of an *Unique Resource Locator* string and a single rich text sequence being its description. It signifies the embedment of an external resource.

A description as mentioned above, is a piece of text elaborating the contents of a given element.

A **section** consists of a Geneva document and a single rich text sequence being its heading. It signifies a continuous subsequence of the document, introduced by a headline (the heading).

3   Formal Definition

The table below defines a Geneva document formally using the modifed BNF syntax described in ANSI Common Lisp's *Notational Conventions*.[1]

Symbol	Expression
document	document-element*
document-element	pargraph | listing | table | plaintext | media | section
paragraph	text-token+
listing	rich-text+
table	rich-text table-row+
table-row	rich-text+
plaintext	rich-text string
media	rich-text string
section	rich-text document-element*
rich-text	text-token*
text-token	*A text token, see "Rich Text"*
string	*A character string*

Table 2. Formal definition of a Geneva document.

- 1. *ANSI Common Lisp: Notational Conventions* (`http://users-phys.au.dk/harder/Notational-Conventions.html`)