

1 From: timbl@nxoc01.cern.ch
 2 Subject: Comments on P2.2 Functional Spec v2
 3 To: rare-wg3@SURFNET.NL
 4 Date: Mon, 22 Apr 91 09:57:36 GMT+0100

5
 6 Here are comments on version 2 of the P2.2 functional specification
 7 (D10), for what they are worth.

8
 9 The comments marked * are looking ahead to when one might want to use
 10 the data as part of a hypertext system, and I would encourage their
 11 inclusion out of foresight.

12
 13 GENERAL:

14
 15
 16 The philosophy that the P2.2 server should be "central" is
 17 suspicious. No facility should regard itself as central, or it will
 18 have big problems fitting in with other facilities which do the same.
 19 That which is "central" is very subjective.

20
 21 * There should be provision for a non-interactive fast enquiry. This
 22 probably just means turning off any paging of the output, for
 23 example, and formatting the result with (SGML preferably) tags for
 24 parsing.

25
 26 NIGGLEY:

27
 28 2.2.9 "...able to answer 80% of questions correctly about a word
 29 processing package with which they are familiar ...".
 30 This doesn't mean anything - what questions? Can the user make up
 31 his/her own questions? There are questions and questions. If there is
 32 no way to be precise, don't pretend. I would just remove the
 33 sentence.

34
 35 3.2.1.20 "User manuals"

36
 37 For CONCISE itself, do you mean, or for anything?

38
 39 * 3.2 para 1: "...includes pointers..."

40
 41 If those pointers are to remote electronic data, would it not be wise
 42 to put them either in the database, or in a form in the text in which
 43 they can be extracted, using a consistent representation?

44
 45 3.2 para 4: "cross links will be added to the structure".

46
 47 This is good (quite hypertexty). Is there an acyclic constraint on
 48 the web? I hope not. Can one distinguish "cross-links" from "links"?

49
 50 3.2.4: redundancy limited to 30%.

51
 52 Why this limit? The quality of the information includes its
 53 completeness. My copy of the Geneva phone book is 99.99% redundant,
 54 but I would hate unused entries to be deleted! Explain to me why the
 55 limit is needed.

56
 57 * 3.2.7.9 "the expiry date of the information"

58
 59 Another useful concept is the "shelf life" of any copy one might
 60 take. That is, how often I ought to go back and see whether the
 61 information has changed. For an RFC it's years, for a phone book it's
 62 24hrs.

63
 64 * 3.2.7.12 "The type of data".

65
 66 I notice that different format files of the same document (xxx.txt,
 67 xxx.ps) must be treated as different nodes. I suppose some convention
 68 for making links to the source, and then having all derived files
 69 below that, for example, will be needed. I point out that having a
 70 set of (data type, filename) for different representations of the
 71

72 same thing is an alternative.
 73

74
 75 3.2.7.2 "a name"

76 One can distinguish between an identifier, a menu item, and a title.

77
 78
 79 An identifier (like a filename) would form part of a path to the
 80 document. If you quote the name of a document, then you give its
 81 path. Do not underestimate the amount of access which will be made by
 82 people quoting a full path name which they have on the back of an
 83 envelope. Unless you rule out this method and only allow access by
 84 quoting the unique numeric ID, then it is wise for the identifier to
 85 be short and have no spaces (like a nice filename).

86
 87 On a menu, by contrast, a little phrase is better. Ideally it makes
 88 sense IN THE CONTEXT OF THE MENU. The menu item can be long and
 89 contain spaces.

90
 91 It is also useful to have a title. This is a human-readable string
 92 which describes the document IN A VERY WIDE CONTEXT. You would
 93 use this as a running head if you were printing the document. You
 94 would put it on the title bar of a window if you were displaying the
 95 document.

96
 97 For example, the minutes of our last meeting could have

98
 99 unique id 1734572
 100 identifier min910415
 101 RARE/WG3/menu item Minutes of the meeting of April 15th
 102 Title Minutes of the 15-Apr-91 meeting of RARE WG3
 103

104 The menu item can default to the title, and the title can default to
 105 the identifier, and that can even default to the unique id :- (but
 106 its not so user friendly.

107
 108 (You have probably discussed this at length already!) Which do you
 109 mean here?
 110

111
 112 3.3.10

113 I agree with the many who have said that the times quoted are too
 114 long. In my experience, most people are content with 3 sec mean, 10
 115 sec max retrieval time (and that includes any transmission delays).
 116 They are not happy with longer. I would guess one should aim for a
 117 search to be mean 0.3, max 3s without transmission. An indexed search
 118 ought to be able to achieve this. Without any mention of the size of
 119 the database, of course, none of the figures mean a lot. An
 120 irritatingly slow system won't do anything for the reputation of
 121 anybody, or the networks. (Yes, I know the larger figures are
 122 contractual requirements. Perhaps the document should also contain
 123 some more realistic figures so that one doesn't forget to aim higher)

124
 125 3.4.4 Interactive access

126 I assume that full-screen mode will not be implemented, no cursor
 127 control facilities on a terminal will be used. Is this correct
 128

129 4.3, last para: for "parts" read "parents".

130
 131 4.3.15: Rights to make links

132
 133 As I understand it, a link is monodirectional, from one node to
 134 another. It does not appear in or affect the destination node. In
 135 this case I propose that anyone who has the rights to alter a node
 136 should have the right to make a link FROM it. (By analogy, anyone who
 137 writes a paper is allowed to make a reference to anyone else's
 138 published paper without asking their permission.)
 139

140 4.4.3.2 "Not all IP functions will be available through RFC987
 141 gateways"

142 Excuse my ignorance, but why not? Apart from multiple body parts

143 returned, will there be a serious lack of functionality?
144 _____ 5
145 .2.1 Statistics
146 It is also very useful to have a list of failed keyword searches.
147 People are often loth to send a complaint message to the help desk,
148 and may just leave concluding that the information is not there.
149 Given a list of failed searches, one can improve the keywording of
150 nodes, and keep track of the things which users would like but
151 haven't got.
152
153 5.5.5 "sufficient spare disk storage space to store all the data"
154 Now is the time for someone to make an estimate. If 300MB should
155 cover it, say so. It affects one's perception of the sort of data to
156 be stored.
157
158 5.7.1 "trialed"? "tried". In a trial, one tries something.
159 _____
160 7.2.1 "The programs will be in ANSI C"
161 For easy portability, a K&R or Berkley C option is useful.
162 This tends to means mainly putting in alternative function headings,
163 #ifdef __STDC__ . As we are obviously (5.6paral etc) talkig about a
164 Sun, but I have just ported wwv to a Sun which did not have an ANSI
165 C compiler. To port ANSI-C tends to imply a port of Gnu gcc...
166
167 7.3.1 "The time required to port...of the operating system..."
168 By which you mean sun-os, Unix V, or unix? Say.
169 I feel it is a pity not to specify portability to BSD when the
170 product is largely aimed at the academic community.
171
172 7.3.4 See comment on 2.2.9.
173 _____
174 7.3.4 Reference to level-7 contract. The others seem to have been
175 removed.
176
177 8 "D37"
178 A handy list of those numbered deliverables referred to would be nice
179 in the references.
180
181
182
183
184
185
186
187