DRAFT 3, prepared by Jim Moore, 18 September 2004

ISO/IEC JTC 1/SC 22/WG 9 N 445
Explanatory Report re SC22 Resolution 04-15

At its 2004 plenary meeting, ISO/IEC JTC 1/SC 22 approved the following resolution:

Resolution 04-15: Coded Character Sets

JTC 1/SC 22 agrees that the proposed implementation of coded character set support described in document N 3758 agrees with the principles for coded character set support previously adopted by SC 22, notably resolution 02-24. JTC 1/SC 22 instructs WG 9 to consider referencing ISO/IEC TR 10176 Annex A in the revision of the Ada language standard.

For easy reference, Resolution 02-24 is quoted below:

Resolution 02-24: Recommendation on Coded Character Sets Support

JTC 1/SC 22 believes that programming languages should offer the appropriate support for ISO/IEC 10646, and the Unicode character set where appropriate.

Background

WG9 is currently preparing an amendment to the Ada Language Standard, ISO/IEC 8652:1995. The amendment will include changes in coded character support. Because the effect of changes in this area will pervade the entire standard and affect the treatment of other issues in the amendment, it was important to gain an early understanding of whether the proposed approach complies with the direction of SC22, notably Resolution 02-24. Therefore, WG9 prepared a document [22N3758] providing an overview of the proposed approach. The document was circulated to SC22 for a 60-day commenting period. (The comments received are summarized in 22N3807.) Furthermore, the WG9 document requested that the SC22 plenary meeting approve a resolution; Resolution 04-15 was the resolution actually approved by SC22 as a result of the request.

Purpose of this Explanatory Report

This explanatory report serves two purposes:
- In response to Resolution 04-15, this report explains that WG9 has considered referencing TR 10176 Annex A and has decided that it should not be referenced.
- In response to 22N3807, this report replies to the National Body comments offered on the approach outlined in 22N3758.

Response

First, the stated positions of the National Bodies are listed:
- Belgium replied to 22N3758, supporting the approach, supporting the resolution, and offering one comment. They did not attend the plenary, hence could not vote on the resolution.
- Canada replied to 22N3758 supporting the approach, and voted in favor of the resolution.

- Denmark replied to 22N3758 supporting the approach. They did not attend the plenary, hence could not vote on the resolution.
- France replied to 22N3758 supporting the approach. They did not attend the plenary, hence could not vote on the resolution.
- Germany replied to 22N3758 opposing the approach and offering comments. They did not attend the plenary, hence could not vote on the resolution.
- Japan voted in favor of the resolution.
- Korea voted in favor of the resolution.
- Netherlands was represented at the plenary by proxy, but did not vote on the resolution for lack of specific instructions.
- UK abstained on the resolution. (22N3807 shows them favoring the approach; but this may be an administrative error.)
- US voted in favor of the resolution.

In their response to 22N3758, Belgium offered a comment:

---

there is a small error approximately in the middle of document 22N3758: the first line in the syntax rule for "identifier_extend":

      identifier_extend ::= identifier_letter |

should be

      identifier_extend ::= identifier_start |

WG9 agrees with the comment and has made the correction.

In their response to 22N3758, Germany offered comments. The first comment is:

DIN strongly suggests that Ada follow the guidance given in Annex A "Recommended extended repertoire for user-defined identifiers" of TR 10176:2002 (4th edition) for the use of UCS characters in identifiers. This list of characters for identifiers is synchronized against ISO/IEC 10646-1:2000 and Unicode 3. Annex A will be kept in sync with future editions of ISO/IEC 10646 and the Unicode standard. Defining such a repertoire specifically for Ada is likely to invite numerous interoperability issues with other languages.

WG9 does not intend to diverge from TR 10176, except to the extent that TR 10176 itself diverges from the latest IS 10646 (which is the case at the moment). More precisely, WG9 intends to align strictly with IS 10646:2003. It must be noted that TR 10176 is a Technical Report while IS 10646 is a Standard. If one has to make a choice, one should conform with the Standard rather than with the Technical Report. And, it turns out that one MUST make a choice because there are important differences between the two:
- TR 10176 is still based on IS 10646:2000 while IS 10646:2003 has already been published for a year.
- There are considerable differences between the two editions of IS 10646, notably in supporting characters beyond the BMP (this might be significant for some languages, e.g. Korean).
- TR 10176 is a moving target. It is in its fourth edition already, and nevertheless needs additional revision to catch up with IS 10646. WG9 cannot afford to revise the Ada language and the vendors cannot afford to change the compilers each time TR 10176 changes. And WG9 cannot afford to delay the adoption of its amendment until TR 10176 has been revised; it would run out of interest, money, and the ISO time table before then.
- TR 10176 does not define case conversion tables, which are essential for a case-insensitive language like Ada. To get case conversion tables, WG9 would have to reference either IS 10646:2003 or Unicode, or would have to invent its own.

Germany provided an additional comment in their response to 22N3758:

> In its present form SC22 N3758 makes numerous references to character properties exclusively defined in the Unicode standard while maintaining that "This proposal is based on ISO/IEC 10646:2003. While this proposal contains references to Unicode, the amendment text will be carefully phrased to avoid such mentions".

The Ada language amendment needs character properties like categorization, as well as case conversion tables. These are mentioned in IS 10646:2003 as useful for implementations, with a reference to Unicode. Since the writers of IS 10646 thought that it was appropriate to say "see Unicode 4.0", WG9 does not know how it could follow a different lead. Our summary document IS N3758 was written in a manner to clarify the issues, hence it included mentions of Unicode. The actual text of the amendment will make specific references to the corresponding clauses of IS 10646:2003.

Finally, Germany provided a favorable comment:

> DIN lauds the intention to add new types to Interfaces.C that correspond to char16_t and char32_t as defined in upcoming ISO/IEC TR 19769.

WG9 appreciates the kind remarks.

Finally, WG9 responds to the final sentence of Resolution 04-15:

> JTC 1/SC 22 instructs WG 9 to consider referencing ISO/IEC TR 10176 Annex A in the revision of the Ada language standard.

WG9 has carefully considered referencing TR 10176 in its planned amendment. In WG9's response to the German NB comments (above), it is shown that TR 10176 is inadequate for the needs of a case-insensitive language, and conflicts with IS 10646. WG9 also notes the existence of Resolution 02-24 encouraging use of IS 10646. Finally, WG9 notes that the JTC1 Directives (5th Edition) discourage referencing documents other than international standards.

> N5.2 Preparation of the Referencing Explanatory Report
> It is the obligation of the JTC 1 Subcommittee which develops an IS containing normative references to specifications other than international standards, to provide for each proposed RS a related RER.

Referencing TR 10176 instead of IS 10646 would require justification in an explanatory report and review by JTC1.

Action

WG9 thanks SC22 and its national bodies for their consideration of this issue. Based on Resolution 04-15, WG9 will prepare its language amendment in accordance with the approach outlined in 22N3758. As instructed by the resolution, WG9 has considered a reference to TR 10176, but for reasons outlined above, has decided to reference IS 10646 instead.