        A Summary of Unicode Consortium Procedures, Policies, Stability,
                            and Public Access

Status of this Memo

Copyright Notice

Abstract

   This memo describes various internal workings of the Unicode
   Consortium for the benefit of participants in the IETF. It is
   intended solely for informational purposes. Included are discussions
   of how the decision-making bodies of the Consortium work and their
   procedures, as well as information on public access to the character
   encoding & standardization processes.

1. Introduction

   This memo describes various internal workings of the Unicode
   Consortium for the benefit of participants in the IETF. It is
   intended solely for informational purposes. Included are discussions
   of how the decision-making bodies of the Consortium work and their
   procedures, as well as information on public access to the character
   encoding & standardization processes.

2. About The Unicode Consortium

   The Unicode Consortium is a corporation. Legally speaking, it is a
   "California Nonprofit Mutual Benefit Corporation", organized under
   section 501 C(6) of the Internal Revenue Service Code of the United
   States. As such, it is a "business league" not focussed on profiting
   by sales or production of goods and services, but neither is it
   formally a "charitable" organization. It is an alliance of member
   companies whose purpose is to "extend, maintain, and promote the
   Unicode Standard". To this end, the Consortium keeps a small office,
   a few editorial and technical staff, World Wide Web presence, and
   mail list presence.

The corporation is presided over by a Board of Directors who meet
annually. The Board is comprised of individuals who are elected
annually by the full members for three-year terms. The Board
appoints Officers of the corporation to run the daily operations.

Membership in the Consortium is open to "all corporations, other
business entities, governmental agencies, not-for-profit
organizations and academic institutions" who support the Consortium's
purpose. Formally, one class of voting membership is recognized, and
dues-paying members are typically for-profit corporations, research
and educational institutions, or national governments. Each such
full member sends representatives to meetings of the Unicode
Technical Committee (see below), as well as to a brief annual
Membership meeting.

3. The Unicode Technical Committee

The Unicode Technical Committee (UTC) is the technical decision
making body of the Consortium. The UTC inherited the work and prior
decisions of the Unicode Working Group (UWG) that was active prior to
formation of the Consortium in January 1991.

Formally, the UTC is a technical body instituted by resolution of the
board of directors. Each member appoints one principal and one or
two alternate representatives to the UTC. UTC representatives
frequently do, but need not, act as the ordinary member
representatives for the purposes of the annual meeting.

The UTC is presided over by a Chair and Vice-Chair, appointed by the
Board of Directors for an unspecified term of service.

The UTC meets 4 to 5 times a year to discuss proposals, additions,
and various other technical topics. Each meeting lasts 3 to 4 full
days. Meetings are held in locations decided upon by the membership,
frequently in the San Francisco Bay Area. There is no fee for
participation in UTC meetings. Agendas for meetings are not
generally posted to any public forum, but meeting dates, locations,
and logistics are posted well in advance on the "Unicode Calendar of
Events" web page.

At the discretion of the UTC chair, meetings are open to
participation of member and liaison organizations, and to observation
by others. The minutes of meetings are also posted publicly on the
"UTC Minutes" page of the Unicode Web site.

All UTC meetings are held jointly with the INCITS Technical Committee
L2, the body responsible for Character Code standards in the United
States. They constitute "ad hoc" meetings of the L2 body and are

usually followed by a full meeting of the L2 committee. Further
information on L2 is available on the official INCITS web page.

4. Unicode Technical Committee Procedures

The formal procedures of the UTC are publicly available in a document
entitled "UTC Procedures", available from the Consortium, and on the
Unicode web site.

Despite the invocation of Robert's Rules of Order, UTC meetings are
conducted with relative informality in view of the highly technical
nature of most discussions. Meetings focus on items from a technical
agenda organized and published by the UTC Chair prior to the meeting.
Technical items are usually proposals in one of the following
categories:

   1. Addition of new characters (whole scripts, additions to
      existing scripts, or other characters)

   2. Preparation and Editing of Technical Reports and Standards

   3. Changes in the semantics of specific characters

   4. Extensions to the encoding architecture and forms of use

Note: There may also be changes to the architecture, character
properties, or semantics. Such changes are rare, and are always
constrained by the "Unicode Stability Policies" posted on the Unicode
web site. Significant changes are undertaken in consultation with
liaison organizations, such as W3C and IETF, which have standards
that may be affected by such changes. See sections 5 and 6 below.

Typical outputs of the UTC are:

   1. The Unicode Standard, major and minor versions (including the
      Unicode Character Database)

   2. Unicode Technical Reports

   3. Stand-alone Unicode Technical Standards

   4. Formal resolutions

   5. Liaison statements and instructions to the Unicode liaisons to
      other organizations.

For each technical item on the meeting agenda, the general process is
as follows:

   1. Introduction by the topic sponsor

   2. Proposals and discussion

   3. Consensus statements or formal motions

   4. Assignment of formal actions to implement decisions

5. Unicode Technical Committee Motions

   Technical topics of any complexity never proceed from initial
   proposal to final ratification or adoption into the standard in the
   course of one UTC meeting. The UTC members and presiding officers
   are aware that technical changes to the standard have broad
   consequences to other standards, implementers, and end-users of the
   standard. Input from other organizations and experts is often vital
   to the understanding of various proposals and for successful adoption
   into the standard.

   Technical topics are decided in UTC through the use of formal
   motions, either taken in meetings, or by means of thirty-day letter
   ballots. Formal UTC motions are of two types:

   1. Simple motions

   2. Precedents

   Simple motions may pass with a simple majority constituting more than
   50 percent of the qualified voting members; or by a special majority
   constituting two-thirds or more of the qualified voting members.

   Precedents are defined, according to the UTC Procedures as either

      (A) an existing Unicode Policy, or

      (B) an explicit precedent.

   Precedents must be passed or overturned by a special majority.

   Examples of implicit precedents include:

   1. Publication of a character in the standard

   2. Published normative character properties

      3. Algorithms required for formal conformance

   An Explicit Precedent is a policy, procedure, encoding, algorithm, or
   other item that is established by a separate motion saying (in
   effect) that a particular prior motion establishes a precedent.

   A proposal may be passed either by a formal motion and vote, or by
   consensus. If there is broad agreement as to the proposal, and no
   member wishes to force a vote, then the proposal passes by consensus
   and is recorded as such in the minutes.

6. Unicode Consortium Policies

   Because the Unicode Standard is continually evolving in an attempt to
   reach the ideal of encoding "all the world's scripts", new characters
   will constantly be added. In this sense, the standard is unstable:
   in the standard's useful lifetime, there may never be a final point
   at which no more characters are added. Realizing this, the
   Consortium has adopted certain policies to promote and maintain
   stability of the characters that are already encoded, as well as
   laying out a Roadmap to future encodings.

   The overall policies of the Consortium with regard to encoding
   stability, as well as other issues such as privacy, are published on
   a "Unicode Consortium Policies" web page. Deliberations and encoding
   proposals in the UTC are bound by these policies.

   The general effect of the stability policies may be stated in this
   way: once a character is encoded, it will not be moved or removed and
   its name will not be changed. Any of those actions has the potential
   for causing obsolescence of data, and they are not permitted. The
   canonical combining class and decompositions of characters will not
   be changed in any way that affects normalization. In this sense,
   normalization, such as that used for International Domain Naming and
   "early normalization" for use on the World Wide Web, is fixed and
   stable for every character at the time that character is encoded.
   (Any changes that are undertaken because of outright errors in
   properties or decompositions are dealt with by means of an adjunct
   data file so that normalization stability can still be maintained by
   those who need it.)

   Once published, each version of the Unicode Standard is absolutely
   stable and will never be changed retroactively. Implementations or
   specifications that refer to a specific version of the Unicode
   Standard can rely upon this stability. If future versions of such
   implementations or specifications upgrade to a future version of the
   Unicode Standard, then some changes may be necessary.

Property values of characters, such as directionality for the Unicode
Bidi algorithm, may be changed between versions of the standard in
some circumstances. As less-well documented characters and scripts
are encoded, the exact character properties and behavior may not be
well known at the time the characters are first encoded. As more
experience is gathered in implementing the newly encoded characters,
adjustments in the properties may become necessary. This re-working
is kept to a minimum. New and old versions of the relevant property
tables are made available on the Consortium's web site.

Normative and some informative data about characters is kept in the
Unicode Character Database (UCD). The structure of many of these
property values will not be changed. Instead, when new properties
are defined, the Consortium adds new files for these properties, so
as not to affect the stability of existing implementations that use
the values and properties defined in the existing formats and files.
The latest version of the UCD is available on the Consortium web site
via the "Unicode Data" heading.

Note on data redistribution: Unlike the situation with IETF
documents, some parts of the Unicode Character Database may have
restrictions on their verbatim redistribution with source-code
products. Users should read the notices in files they intend to use
in such products. The information contained in the UCD may be freely
used to create derivative works (such as programs, compressed data
files, subroutines, data structures, etc.) that may be redistributed
freely, but some files may not be redistributable verbatim. Such
restrictions on Unicode data files are never meant to prohibit or
control the use of the data in products, but only to help ensure that
users retrieve the latest official releases of data files when using
the data in products.

7. UTC and ISO (WG2)

The character repertoire, names, and general architecture of the
Unicode Standard are identical to the parallel international standard
ISO/IEC 10646. ISO/IEC 10646 only contains a small fraction of the
semantics, properties and implementation guidelines supplied by the
Unicode Standard and associated technical standards and reports.
Implementations conformant to Unicode are conformant to ISO/IEC
10646.

ISO/IEC 10646 is maintained by the committee ISO/IEC JTC1/SC2/WG2.
The WG2 committee is composed of national body representatives to
ISO. Details on the ISO organization may be found on the official
web site of the International Organization for Standardization (ISO).

Details and history of the relationship between ISO/IEC JTC1/SC2/WG2
and Unicode, Inc. may be found in Appendix C of The Unicode Standard.
(A PDF rendition of the most recent printed edition of the Unicode
Standard can be found on the Unicode web site.)

WG2 shares with UTC the policies regarding stability: WG2 neither
removes characters nor changes their names once published. Changes
in both standards are closely tracked by the respective committees,
and a very close working relationship is fostered to maintain
synchronization between the standards.

The Unicode Collation Algorithm (UCA) is one of a small set of other
independent standards defined and maintained by UTC. It is not,
properly speaking, part of the Unicode Standard itself, but is
separately defined in Unicode Technical Standard #10 (UTS #10).
There is no conformance relationship between the two standards,
except that conformance to a specific base version of the Unicode
Standard (e.g., 4.0) is specified in a particular version of a UTS.
The collation algorithm specified in UTS #10 is conformant to ISO/IEC
14651, maintained by ISO/IEC JTC1/SC2, and the two organizations
maintain a close relationship. Beyond what is specified in ISO/IEC
14651, the UCA contains additional constraints on collation,
specifies additional options, and provides many more implementation
guidelines.

8. Process of Technical Changes to the Unicode Standard

Changes to The Unicode Standard are of two types: architectural
changes, and character additions.

Most architectural changes do not affect ISO/IEC 10646, for example,
the addition of various character properties to Unicode. Those
architectural changes that do affect both standards, such as
additional UTF formats or allocation of planes, are very carefully
coordinated by the committees. As always, on the UTC side,
architectural changes that establish precedents are carefully
monitored and the above-described rules and procedures are followed.

Additional characters for inclusion in the The Unicode Standard must
be approved both by the UTC and by WG2. Proposals for additional
characters enter the standards process in one of several ways:
through...

    1. a national body member of WG2

    2. a member company or associate of UTC

    3. directly from an individual "expert" contributor

The two committees have jointly produced a "Proposal Summary Form" that is required to accompany all additional character proposals. This form may be found online at the WG2 web site, and on the Unicode web site along with information about "Submitting New Characters or Scripts". Instructions for submitting proposals to UTC may likewise be found online.

Often, submission of proposals to both committees (UTC and WG2) is simultaneous. Members of UTC also frequently forward to WG2 proposals that have been initially reviewed by UTC.

In general, a proposal that is submitted to UTC before being submitted to WG2 passes through several stages:

   1. Initial presentation to UTC

   2. Review and re-drafting

   3. Forwarding to WG2 for consideration

   4. Re-drafting for technical changes

   5. Balloting for approval in UTC

   6. Re-forwarding and recommendation to WG2

   7. At least two rounds of international balloting in ISO

About two years are required to complete this process. Initial proposals most often do not include sufficient information or justification to be approved. These are returned to the submitters with comments on how the proposal needs to be amended or extended. Repertoire addition proposals that are submitted to WG2 before being submitted to UTC are generally forwarded immediately to UTC through committee liaisons. The crucial parts of the process (steps 5 through 7 above) are never short-circuited. A two-thirds majority in UTC is required for approval at step 5.

Proposals for additional scripts are required to be coordinated with relevant user communities. Often there are ad-hoc subcommittees of UTC or expert mail list participants who are responsible for actually drafting proposals, garnering community support, or representing user communities.

The rounds of international balloting in step 7 have participation both by UTC and WG2, though UTC does not directly vote in the ISO process.

Occasionally a proposal approved by one body is considered too
immature for approval by the other body, and may be blocked de-facto
by either of the two. Only after both bodies have approved the
additional characters do they proceed to the rounds of international
balloting. (The first round is a draft international standard during
which some changes may occur, the second round is final approval
during which only editorial changes are made.)

This process assures that proposals for additional characters are
mature and stable by the time they appear in a final international
ballot.

9. Public Access to the Character Encoding Process

While Unicode, Inc. is a membership organization, and the final say
in technical matters rests with UTC, the process is quite open to
public input and scrutiny of processes and proposals. There are many
influential individual experts and industry groups who are not
formally members, but whose input to the process is taken seriously
by UTC.

Internally, UTC maintains a mail list called the "Unicore" list,
which carries traffic related to meetings, technical content of the
standard, and so forth. Members of the list are UTC representatives;
employees and staff of member organizations (such as the Research
Libraries Group); individual liaisons to and from other standards
bodies (such as WG2 and IETF); and invited experts from institutions
such as the Library of Congress and some universities. Subscription
to the list for external individuals is subject to "sponsorship" by
the corporate officers.

Unicode, Inc. also maintains a public discussion list called the
"Unicode" list. Subscription is open to anyone, and proceedings of
the "Unicode" mail list are publicly archived. Details are on the
Consortium web site under the "Mail Lists" heading.

Technical proposals for changes to the standard are posted to both of
these mail lists on a regular basis. Discussion on the public list
may result in a written proposal being generated for a later UTC
meeting. Technical issues and other standardization "events" of any
significance, such as beta releases and availability of draft
documents, are announced and then discussed in this public forum,
well before standardization is finalized. From time to time, the UTC
also publishes on the Consortium web site "Public Review Issues" to
gather feedback and generate discussion of specific proposals whose
impact may be unclear, or for which sufficiently broad review may not
yet have been brought to the UTC deliberations.

Anyone may make a character encoding or architectural proposal to
UTC. Membership in the organization is not required to submit a
proposal. To be taken seriously, the proposal must be framed in a
substantial way, and be accompanied by sufficient documentation to
warrant discussion. Examples of proposals are easily available by
following links from the "Proposed Characters" and "Roadmaps"
headings on the Unicode web site. Guidelines for proposals are also
available under the heading "Submitting Proposals".

In general, proposals are publicly aired on the "Unicode" mail list,
sometimes for a long period, prior to formal submission. Generally
this is of benefit to the proposer as it tends to reduce the number
of times the proposal is sent back for clarification or with requests
for additional information. Once a proposal reaches the stage of
being ready for discussion by UTC, the proposer will have received
contact through the public mail list with one or more UTC members
willing to explain or defend it in a UTC meeting.

10. Acknowledgements

Thanks to Mark Davis, Simon Josefsson, and Ken Whistler for their
extensive review and feedback on previous versions of this document.

11. Security Considerations

This memo describes the operational procedures of an organization;
the procedures themselves have no consequences for Internet Security.

12. Author's Address

Rick McGowan
c/o The Unicode Consortium
P.O. Box 391476
Mountain View, CA 94039-1476
U.S.A.

Phone: +1-650-693-3921
Web: http://www.unicode.org/

13. Full Copyright Statement

Intellectual Property

Acknowledgement