

Approved Minutes - UTC #73 & L2 #170 joint meeting
Palo Alto, CA - August 4-5, 1997

1. Administrative Issues

1A. UTC Membership Roll Call

Quorum: 10

Present: Apple Computer, Inc.; Digital Equipment Corporation; Hewlett-Packard Company; IBM Corporation; Justsystem Corporation; Microsoft Corporation; NCR Corporation; Novell, Inc.; The Research Libraries Group, Inc.; Silicon Graphics, Inc.; Spyglass, Inc.; Sun Microsystems, Inc.; Sybase, Inc.; Unisys Corporation; Xerox Corporation.

Not Present: Booz, Allen & Hamilton, Inc.; Gamma Productions, Inc.; Oracle Corporation; Reuters, Ltd.

List of attendees: See appendix.

1B. Declaration of Joint Meeting

1C. Approval of Joint meeting Agenda

#73-M1 Motion: To approve agenda as amended.

Moved by Jenkins, seconded by Winkler

Unanimous.

1D. Approval of Minutes and Review of Action Items

Motion [#73-M2]: To approve the draft minutes of the joint meeting of the UTC [meeting #72] and NCITS Subgroup L2 [meeting #169].

Moved by Freytag, seconded by Jenkins.

Motion approved: 13 for, 0 against, 1 abstention (NCR)

Action 73-1 for Aliprand: Check with companies that have not participated in meetings to find out why.

Action 73-2 for McGowan: Ask Bruce Paterson to review Western Music Symbols proposal

Action 73-3 for Winkler: Copy action items for the editorial committee onto a separate list.

Action 73-4 for Aliprand/Winkler: Include numbered action items in the meeting minutes.

Action 73-5 for Aliprand: Include "Property Related Errata" as standing item on agenda.

Action 73-6 for Freytag: Put sample implementation of compression algorithm on Unicode Web site.

Action 73-7 for Jenkins: Tell Peter Edberg that revised transcoding proposal is wanted for December meeting.

Action 73-8 for Aliprand: Put transcoding proposal on agenda for December meeting.

1E. Registration of new documents

1F. Meeting Calendar

Freytag stated that WG2 is headed towards self-imposed deadline of considering proposals, so agenda is likely to be full, thus a third meeting is needed.

Action 73-9 for Winkler: Winkler to investigate alternative dates for December meeting.

Action 73-10 for Moore: Determine availability of rooms for other dates in December.

Action 73-11 for McGowan/Jenkins: Reserve room for UTC/L2 on February 25-27.

Action 73-12 for Winkler: Reserve room for UTC/L2 meeting on April 20-22.

Action 73-13 for Aliprand: Send out revised meeting dates.

1G. Appointment of Unicode Consortium representative for IRG meeting 10

Adams: When there is more than one delegate, one should be appointed Head of Delegation. L2 also sends a representative.

Motion [#73-M3]: That the UTC endorse Glenn Adams as HOD for the Unicode Consortium. Moved by Freytag, seconded by Jenkins
Unanimous.

Motion [#73-M4]: That the UTC endorse other UTC representatives attending IRG meeting 10 as alternates for the Unicode Consortium, with the HOD to appoint a substitute from the Unicode alternates if he is unable to attend.

Moved by Freytag, seconded by Jenkins.

Unanimous.

Hideki Hiura (Sun) and Michael Kung (Oracle) plan to attend the IRG meeting.

Recommendation to L2: That L2 appoint Glenn Adams to HOD for the US delegation, with John Jenkins and Michael Kung as alternates.

2. Interaction with SC2, part 1

2A. Report from Unicode liaison and L2 IR about the WG2 meeting in Crete

Freytag reported that the Crete meeting was very successful as far as Unicode [and L2] are concerned. Representatives showed flexibility; successful ad hoc Meeting on Sinhalese. DAMS 5-7 approved. ISO 10646 and Unicode Version 2.0 are now in synch (almost a year after the publication of Version 2.0). Risky, but successful. Strategic thing: WG2 agreed not to reconfirm ISO/IEC10646-1, but to revise it. (This recommendation was approved by SC2.)

2B. Version 3.0 and its relationship to Revision of ISO/IEC 10646

[L2/97-150] RESOLUTION M33.4 (5 year review of 10646)

Ksar: Note that each resolution has target date for publication. This will help synchronization between Unicode and ISO 10646. Ethiopic, Canadian syllabics, and Cherokee are essentially frozen (due to new rules); also the Character-Glyph model. Going to DAM ballot; DTR ballot for C-G model. Two rules on coding among WG2 resolutions, M33.11 helps with compression algorithm.

Freytag: Regarding WG2 procedures: pDAM has two rounds of technical votes. Objection to particular symbols needs to feed into L2 and Unicode liaison position statements.

Ksar: Split into 2 pieces, as previous decision could not be withdrawn.

Action 73-14 for All: Send comments on keyboard symbols to Freytag before Crete meeting

Adams: Reminder to use most recent version of proposal summary form -- N1502.

It is difficult to undo WG2 decisions. Opposition is not enough.

Ksar: Whenever we make a proposal – need to do homework.

“Accept into Unicode Standard” needs to be rephrased for WG2. Problem of having to back out of decisions.

Ksar: WG2 resolved not to reconfirm 10646 as is, but to republish including all of the approved amendments (AMs) up to now; plus, other up to time that Vertical Extension codes tables will be published (March 1999).

Version 3.0 – we want to be in synch with the next publication of 10646. Excellent goal.

Action 73-20 for Ksar: Send dates for revision of ISO 10646 publication to “unicore.”

Freytag: Defining steps for publication of Version 3.0. Content not yet defined. Future action: To report to UTC on content and coordination of publication schedules. Default goal is to be in synch with next edition of 10646.

Ksar: Met with ISO Secretariat in Geneva. Left copy of Unicode 2.0 Action Item for them: Could same tool be used? What can be used, what cannot? Editorial issues, not technical. Initial impression that layout could be modified. Biggest issue (up to WG2 to decide) whether to include 5 Han columns. Another option: One volume or more on CD. May combine pDAMs in future for ballot. Tool that AFII used is not readily available, they may not have fonts.

Whistler: Given past experience, will take approximately a year to complete a publication cycle. Scope: to include all the various pDAMs and beyond, with cut off point up to and including vertical extension. Content of book will not include Plane 1 – will only cover BMP. Keep content in synch with 10646.

Suignard: Issue of 10646.

Whistler: Relate to RFC for ACAP needs.

Freytag: Role of technical reports? If publishing of material off BMP is problematic, could issue as UTRs.

Uma: Will be pressure from WG2 side to publish firm Plane 1 additions; e.g., Byzantine musical symbols.

Freytag: Problem of overhang of properties, and description of script for Plane 1 proposals. Unicode anticipated approval of amendments by about a year. About a year from now, we could publish Version 3.0.

Uma: Exposure is any proposals not on plate today.

Whistler: Anything at FPDAM stage now, closing in summer, with aim for fall publication.

2C. ISO/IEC 10646, Part 2 [L2/97-180]

Uma: Part 2 has been initiated. Will contain both Planes 1 and 2. Ksar is acting editor. Part 1 revision and Part 2 publication are two separate efforts. Part 1 tentative cutoff date is March for anything sent to pDAM. Gating element is Vertical Extension A.

Freytag: Anything for BMP needs to be into WG2 by March.

Ksar: Dates were WG2’s choice; WG2 pushed with SC2 to get them.

Freytag: As Unicode liaison, pushed for spring and fall meetings. US meeting (Michel arranged), London an easy site, Beijing more difficult travel wise. Looking for fares, convenience of travel.

Ksar: Nominated Michel to be editor for Part 2. Denmark nominated Everson. Ksar to be acting editor until resolved. SC2 plenary discussion on re-engineering of JTC1. SC22/WG20, SC18, WG9 and WG8 – cultural adaptability workshop on structuring.

McGowan: Would like to have official Unicode representation at this workshop.

Introduction of Julie Doll Allen as assistant editor for the Unicode Standard.

Action 73-15 for Greenfield: Add Julie Allen to Roster (Unicode Office Section)

2D. Consent Docket on WG2 Resolutions [L2/97-186]

Motion [#73-M5]: That the UTC approve the consent docket as amended.
Moved by Freytag, Adams seconded.
Unanimous approval 14, 0, 0

Motion [#73-M6]: That the UTC concurs with SC2/WG2 Resolution M33.13.
Moved by Freytag, seconded by Adams.
Approved 13, 0, 1

Action 73-16 for All: Send detailed technical comments on Byzantine Greek Music proposal to “unicore” no later than next UTC, December 3rd.

3. Technical Issues

3.A Bidi Category Proposal [L2/97-191]

Bishop: No change to Bidi algorithm itself. Just to change classification of these 6 characters.

McGowan: Choice of separator with dates.

Moore: Behavior of “.” depends on the environment.

McConnell: Step P0 [“P zero”] is kind of strange. Digits change and separator becomes a neutral.

Moore: Makes good case for “@”. But had strong feedback from IBM implementers in Middle East.

Whistler proposed an e-mail ballot to get decision before December UTC.

3.B. Lightweight language tagging

a) **Plane 14 Proposal from Ad Hoc Drafting Group (Whistler) [L2/97-171]; and,**
b) **Alternative proposal (Cowan) [L2/97-198]**

Response to Internet Draft issued in May, with “unicore” discussion. Multilingual String Format – reworking of UTF-8 to hide language tagging via bit patterns for Internet protocols; e.g. ACAP = Application Capability Protocol

McGowan: New Internet protocols should specify a way to indicate language information.

Whistler: Discussion regarding MLSF. Alternative way to encode language tags within Unicode vs. alternative outside control of UTC. MLSF uses unused bit patterns from UTF-8. Could not be inter-converted to Unicode. Developers of protocols felt that the Consortium would not provide language tagging. Unicore discussion felt MLSF would be bad. Move away from pure UTF-8 undesirable. Bit-twiddling would lead to inter-operability problems. In late June, an ad hoc group (John Meyer, Chris Newman, Mark Crispin, and Unicode volunteers) met. Plane 14 proposal = consensus of that working group. Encode language tags. Meet Internet group requirements. Minimal impact on Unicode applications. Take small area of Plane 14, ASCII as “meta ASCII” control characters. RFC 1766 is source for language tags.

Ksar: Why Plane 14?

Whistler: 1. Not disrupt roadmap with some extraneous characters; and, 2. Quite different from normal content characters. Plain text, but for a markup scheme that is carried in plain text, not part of plain text context.

Ksar: How to consider regarding Part 2?

Whistler: Would be part of Part 2.

Ksar: How many characters?

Whistler: 98 in block of 128. Does not constrain remainder of Plane 14. 98 equals graphic characters of ASCII. Consistent with other tagging schemes. “Cancel” in position corresponding to “delete” in ASCII.

Ksar: Anyone wanting to use would have to use UTF-16.

Whistler: Use of BMP would cause problems with sub-setting. Better to have them completely outside BMP.

Ksar: Plane 16 is last public plane.

Whistler: Beginning of plane used to support range checking, debugging, etc.

Freytag: Is there an advantage to use a plane other than Plane 14? What extension space does not get used much past Plane 3?

Whistler: Realistic space to be encoded is Plane 0–2. Other planes likely to be used are 15 and 16 (Private Use). Since you segment the properties and the tables that look them up – not that much different.

McGowan: Initial objection to lower plane is non-0 probability of Han spelling over to Plane 3.

Whistler: The big unknown in Plane 1 is the size of non-Han ideographic scripts.

Freytag: How about at front of Plane 1?

Becker: Putting out on Plane 14 because they are not coded characters.

Freytag: If screening is based on high surrogate, Plane 14 is unusable for anything except these characters.

Becker: Core problem that this solves is Han unification problem. Has been accepted by IETF participants.

Ksar: Implications for Bidi algorithm?

Freytag: If you know language, gives better guidance.

Whistler: Implication for Bidi is having language-tagged text. From point of view of Bidi, are neutral. Treat like private use characters.

Bishop: Characters are 32 bit ASCII characters.

Adams: Usage may be an issue in context of WG2, therefore define merely as repertoire and encoding. Leave usage/semantics to be defined elsewhere.

Ksar/Uma: May refer to RFC in context of use of “script.” Internet Society is liaison organization. Semantics and syntax are not defined by WG2.

Adams: Open issue. Source set tag identifiers to be used. Possibilities:

- MIME character set registry: somewhat random collection of sets.
- ISO registry (ECMA): referenced by other standards (e.g., SGML).
- Independent registry.

Action 73-17 for Whistler and Edberg: Write a UTR on character set identification and structure.

Adams: Should not touch language identification.

McGowan: How many important standards are there that need to be identified as an external reference via Plane 14 proposal?

Adams: We do mapping tables for specific character sets. Some burden, but we would be in control.

Whistler: Before having a complete source set registry in place (needed for mapping tables anyhow). “Grandfather” in CJKTV as identifiers initially. This would meet the needs of the IETF participants.

Becker: Register identifiers if mapping table is provided.

Moore: Start with Kobayashi-san’s set in his proposal.

Uma: Attacks a number of different problems. Language tagging; source code separation. Split into two parts: a) language tagging = ACAP proposal; and, b) source code separation.

Adams: Repertoire and codes can proceed independently of set of identifiers, including semantics of scope.

Moore: Character set source tags directed to CJK.

Whistler: CJKTV as single character identifiers solve CJK source separation.

Ksar: Don’t want this to be another registration scheme. Constrain to MLSF needs + CJKTV. Buys time, solves immediate need.

Freytag: Source set stuff is not ready for prime time. Language tagging source set separation can be done later. No agreement on structure of tag.

Kobayashi: Both language tagging and source set separation are capable of distinguishing CJKV. Language tagging is more generic. For CJK just go with language tagging as a superset of source set identifier. Our direction should not be towards opponents of Unicode, go with worldwide approach to increase Unicode acceptance; e.g., through WG2 and SC2. Source separation is not good enough. Need to have variant tag.

Hiura: Are some unified characters even in Japanese standards? Need to distinguish them. Need clear variant mechanism.

Whistler: Need for source set identifier is inter-conversion. Need a registry of what it is being converted to.

Adams: How important are transcoding “hint” characters to Apple?

McGowan: Primarily for CJKT. Want short, preferably single-character tags.

Adams: Should spend more time on a generic tagging structure. Three types: 1) different form in future; 2) reverse domain name; 3) local convention. Model is formal public identifier syntax of SGML. Names have “tag character” as prefix rather than post fix, and block-tag character. TAG LATIN CHARACTER A, also suggested.

Becker: Don’t require that tag be correct for conformance. Need a disclaimer.

Sargent: Concerned about reverse domain name, not particularly secure against collisions.

Adams: That’s why I suggested high-level 3-way division.

McGowan: Plane 14 seems like lesser evil.

Action 73-18 for Whistler and others: Prepare Internet draft for Plane 14 (minus source set tagging and generic tagging).

Action 73-19 for Adams: Draft proposal incorporating Cowan’s proposal on generic tagging plus three-way split described by Adams.

Motion[#73-M7]: That the authors should proceed with the specified work items: (a) preparation of a Unicode Technical Report regarding character set identification and structure; (b) preparation of an Internet draft containing the original Plane 14 Proposal minus source set tagging and generic tagging; (c) preparation of a proposal on generic tagging (based on John Cowan’s proposal and including the three-way split proposed by Glenn Adams).

Moved by Freytag, seconded by Moore

Approved 13, 1, 0

3.C BiDi Reference Implementation

McGowan: This shows why we need a reference implementation of the algorithm.

Moore: Three characters have EN separator type: period, slash, figure space.

Bishop: Change, if accepted, would leave only figure space as EN. Propose e-mail discussion for a month.

Freytag: Split into two groups.

McGowan: 1) Switch of EN to common separator as result of errata = fallout from errata. Consider including figure space in this group. 2) Terminator vs. Separator semantic change for “+”, “-”; 3) Strong Left to Right → other neutral: “&” “@” are syntax characters. In original BiDi, had reference implementations. Would like to have Java applet or C source code. Request to Microsoft to create one as reference implementation.

McGowan: C source needs to be revised.

Jenkins: Trying to get a reference version. Recommend Java applet on Web.

Freytag: Look into creating reference application.

Action 73-37 for McGowan: Send C source code for last known working BiDi reference application to Lloyd Honomichi.

Action 73-38 for Honomichi: Create Java applet as reference implementation of BiDi algorithm (to be made available on the Unicode Web site). Include errata from Unicode Web site. Target date: September 1.

Action 73-39 for Bishop: Revise BiDi Property Change proposal, including “figure space” as common separator. Send to “unicore” list, and post on Web site (include Arabic text examples as GIFs in Web version).

Action 73-40 for Bishop: Schedule special meeting on BiDi. Post invitation on “unicore” list. Suggested date: just after IUC.

Texin: Add list of test cases to reference implementation.

Freytag: Needs to have reference implementation to test the changes proposed. Want to see running version at BiDi meeting; not necessarily reference version.

3.D Proposal for In-Line Notation (ruby)

Moore: Directionality change question

Freytag: Would have special tag for 90° rotation. Main formatting information.

Adams: Why not use Plane 14 tagging proposal? If BMP should be used (for efficiency), what about the corporate part of the private use zone?

Whistler: Question regarding proposed syntax? Reason for annotation before base?

Freytag: Propose 2 tags because there are reasons to do either way.

Sargent: Keep plain text at least legible. Warichu and ruby would look strange. Could not be read correctly, if not set off.

McGowan: Proposal not of interest for internal processing. Of interest for serialized output.

Suignard: Base character is part of text. Hiding an object is not a good idea.

Freytag: Consider two other factors in Japanese: 1) frequency of occurrence—pervasive, but may not be common in specific text; 2) current ways of serializing as plain text, e.g., hide in objects/query objects.

Whistler: “Ruby” method used with bopomofo in Chinese. Concur that 90° turning is markup issue. Ruby case can be made for serializing in plain text. Does not think tagging is solution. Warichu does not have registration problem of ruby – difficult to draw hard and fast rule between warichu and parenthetical annotation in general.

Freytag: Warichu style used in vertical text for pronunciation, similar to ruby.

Becker: Is warichu analogous to a math equation in plain text?

McGowan: Ruby is place where there is a clear line; warichu is very fuzzy—typographical taste issue on where and how it is done.

Kobayashi: Three different levels: 1) coded character set; 2) structure of textual content; 3) rendering level.

Ruby rendering level and structure. Do not believe these are part of coded character set. Prefers ruby to be handled at higher level. Only accept if big implementation benefit to inclusion in coded character set. Prefer to see it handled by higher level protocol, e.g., HTML.

Freytag: Similar issue with bidi text regarding HTML. Can handle aspects of bidi with HTML commands rather than bidi algorithm. HTML can carry much more detailed formatting information.

Sargent: Issue of plain text legibility. Have to do something to make it make sense, e.g. for warichu, put the warichu in parentheses. Problem with ruby – how far back does annotation go? Which characters?

Adams: Internal storage vs. serialization (internal use vs. standard internalization). Need to distinguish. Do we want to specify a formal mechanism for interchange? Generic tags provide one mechanism.

Bishop: Why do we think it is necessary? Many believe that it should be handled by a higher level protocol.

Suignard: Microsoft is interested. Private use is not a solution. Is there more interest?

Freytag: Response to answer frequent and recurring needs. Arose in IETF ad hoc group discussion. Improvement over earlier UTC ruby proposals. Link to Object Replacement Character.

Uma: See it as first element of a set of elements to enrich plain text, make it more legible. Footnotes is another case.

Freytag: Important qualification – thinking as strictly limited, rather than open ended. Addressing rather nasty form of rendering. Footnotes are less of a problem.

Whistler: Concur with Uma. If defined as “annotation,” constraint is not present. Would need to define differently.

Freytag: Stressed “in-line” annotation.

Whistler: Similar types of cascading typography.

McGowan: Very cheap for plain text exchange. Plays with existing formats. Does not affect HTML or RTF.

Becker: If ruby is sole problem, then would defer to Koybayashi-san. If general problem, should discuss in e-mail.

Suignard: Ruby is more pressing issue.

Adams: Use of HTML vs. generic tag. HTML has problems. Tagging can be stripped out easily. Would consider tagging rather than HTML. W3C is working on ruby annotations in HTML. Is concerned that we might come up with conflicting proposal.

Whistler: Issue of legible serialization vs. generalized markup. Pick some visible Unicode characters and recommend that they be used. Recognized practice in plain text.

Freytag (summarizing options): 1) Plane 14 tagging; 2) Whistler’s proposal would satisfy need except that the characters could not appear within an annotation. Applies to both private use characters or main BMP characters; 3) Special characters vs. informal protocol.

Uma: Have you accepted that this belongs in plain text?

Hiura: More information is needed to correctly render ruby correctly.

Kobayashi: Mono-ruby vs. group ruby examples.

Whistler: Break-up of semantic units of text.

Bishop: Group ruby case sometimes required.

Suignard: These are known problems.

Jenkins: Hearing that this is not adequate for ruby. But does it provide minimal legibility?

Freytag: Just enough information that a dumb implementer can filter it out and an intelligent implementer can provide default rendering.

Kobayashi: Can be used for general annotation.

Texin: Implications for searching?

Whistler: Could and should make a recommendation to use under-utilized characters.

Kobayashi: Ichitaro has two choices: 1) make only plain text, no ruby; or 2) HTML-like form, can carry ruby as tagged text. Usually use tagged form.

Hiura: Native Japanese does not expect ruby in plain text. Enriched text is always used. Are we trying to solve a hidden problem?

McGowan: Garbage laden text or text without ruby. No middle ground.

Action 73-22 for Freytag: Prepare a written proposal on in-line annotation characters for December meeting. Include discussion of existing mechanisms to support ruby, warichu, etc.

Action 73-23 for Adams: Contribute description of use of Plane 14 tagging for in-line annotation.

Action 73-24 for All: Send comments on in-line annotation to Freytag. Use “Ruby paper” as subject.

3.E Unicode Compression Algorithm

Motion[#73-M15]: That the UTC accepts the proposal on the Unicode Compression Algorithm [L2/97-199] by consensus and authorizes its publication as a Unicode Technical Report.

Moved by Whistler, seconded by Freytag.
Motion approved, by consensus.

Action 73-41 for Freytag: Publish Compression Algorithm as UTR Unicode Technical Report.

4. Specific Scripts

4.A Whole Scripts

4.A.1 Revised Proposals

▪ *Gothic*

Jenkins: Substantially the same, with exceptions. Assigns provisional code points; spells out names of characters; adds “eis with diaeresis” – new character can be represented as combination; numerals 90-900 into main body; has cloned punctuation.

McGowan: Gothic of extremely limited use. Not worth haggling over “eis with diaeresis.”

Whistler: Object to “eis with diaeresis.” Have implications for database and canonical equivalence. Hidden cost for implementers.

Jenkins: Recommend revision with Everson’s names and ordering. Everson is willing to drop punctuation. Position on “eis with diaeresis”?

Motion [#73-M10]: That the UTC accepts the proposal for Gothic script [L2/97-123], with the following deletions: the punctuation marks and eis with diaeresis. Further, the Chair of the Rare Scripts Working Group is authorized to negotiate with Michael Everson about the eis with diaeresis; the UTC provisionally accepts this character, should compromise fail.

Moved by Jenkins, seconded by Freytag.

Unanimous approval 14, 0, 0

Action 73-25 for Jenkins: Negotiate with Michael Everson re joint proposal for Gothic.

Action 73-26 for Jenkins: Prepare revised proposal for Gothic based on result of negotiations.

▪ *Etruscan*

Differences – letter names spelled out.

Motion [#73-M11]: That the UTC proposal for Etruscan script be synchronized with Michael Everson’s proposal for Etruscan script.

Moved by Jenkins, seconded by Whistler.

Unanimous approval 14,0,0

Action 73-27 for Jenkins: Prepare proposal for Etruscan.

4.A.2 New Proposals

▪ *Athenian Acrophonics*

Adams: Should Greek epigraphic project characters be considered?

Whistler and Jenkins: Likely scope is several hundred.

Motion [#73-M12]: The UTC directs the Rare Scripts Working Group to work with Bruce Robertson and other interested parties on a proposal for Athenian acrophonics, and to proceed as follows: 1) aggressively unify what is in the set; 2) aggressively unify with existing Unicode symbols and punctuation (refraining only if unification would require modification

of the appearance of the existing character); and, 3) subset numerosity so that characters could be added to Greek in the BMP.

Moved by Freytag, seconded by Becker.

Unanimous approval 14,0, 0

Jenkins: It may be impossible to encode in Greek block of BMP.

Action 73-28 for Jenkins: Prepare proposal for Athenian acrophonics.

▪ ***Khmer [L2/194]***

Motion [#73-M13]: That the UTC accepts the proposal for Khmer script, with the following requirements: a) use of revised WG2 summary form; b) addition of justification for the gaps in the table layout; c) definition of character properties.

Moved by Jenkins, seconded by Freytag.

Unanimous approval 14,0, 0

Action 73-29 for McGowan: Prepare revised proposal for Khmer, incorporating revised summary form, justification for “holes” in encoding, list of character properties.

Action 73-30 for Adams: Advise McGowan regarding editorial changes for acceptance by WG2.

Action 73-31 for McGowan: In Khmer proposal, replace “14” by “XX”.

Action 73-32 for McGowan: Secure a copy of the font mentioned in the summary form for the Consortium and store it on the internal side of the Unicode FTP site.

Action 73-33 for McGowan: In Khmer, strip out annotations in names.

Action 73-34 for Adams/McGowan: Create Khmer proposal (as close to pDAM text as possible) for submission to WG2.

-

▪ ***Maldividan (Thaana) [L2/197]***

Motion [#73-M14]: The UTC accepts the proposal for Thana script and instructs the Living Scripts Working Group to work with Michael Everson and submit a complete proposal for UTC in December 1997.

Moved by Jenkins, seconded by Freytag.

Unanimous approval

Action 73-35 for McGowan (for Living Scripts WG): Work with Everson to complete proposal for Thaana, with target to return to UTC in December.

Action 73-36 for Adams: Contact Mr. Manaku for input on the Thaana proposal.

▪ ***Roadmap Proposal for Plane 1***

Motion [#73-M16]: That the UTC authorizes the group that prepared the BMP roadmap to prepare a roadmap proposal for Plane 1.

Moved by Freytag, seconded by Jenkins.

Unanimous.

Action 73-42 for Jenkins and others: Prepare UTC version of Plane 1 roadmap. (Jenkins, Becker, Everson, Ksar, McGowan, Whistler)

Action 73-43 for Jenkins: Invite Everson to participate in roadmap work.

Action 73-44 for Jenkins: Post explicit instructions for uploading documents to Web site.

4B. Ideographic additions

4.B.1 Hangzhou Numerals [L2/97-030, page 26]

Motion [#73-M9]: That the UTC concurs with SC2/WG2 Resolution M32.11, and accepts the 3 Hangzhou numeral characters.

Moved by Adams, seconded by Jenkins.

Unanimous.

4.B.2 Ideographic Structure Characters [L2/97-026]

Freytag: Composition method is opposite to combining mark method in Unicode.

Adams & McGowan: Works as operator with operand.

Whistler: Unlike combining marks, 2 operands come first. Have to collect more than one form first.

Whistler: Professor Shieh from Taiwan has proposed only 3 operators. Does this method have to cover all encoded characters, or should it only be for unencoded remainder?

Adams: Only for unencoded items. Of the Vertical Extension, over 95% could have been encoded by composition using three structure characters. In China, much support. Most strong opposition is from Japan. Japan has done most careful work lately on variations in JIS. Important to China for rare characters.

McGowan: Comes down to need to put in Chinese classics, etc.

Adams: Not intended to provide for input methods. Not for display methods and font decomposition.

4C. Additional Non-Ideographic Characters

4.C.1 Banjeom Characters (old Hangul tone marks) L2/97-149

Adams: The issue is: Are these combining characters or not?

Action 73-46 for All: Provide input on Banjeom character proposal (L2/97-149) to Adams before December meeting.

Action 73-47 for Adams: Prepare recommendations on Banjeom character proposal for December meeting.

Action 73-48 for Adams: Consult with Lee Collins re Banjeom characters.

Action 73-49 for Aliprand: Include Banjeom character proposal on agenda for December meeting.

4.C.2. Keyboard Symbols

Motion [#73-M8]: To prepare a proposal for two new characters: keycap cartouche and screen cartouche.

Moved by Umamaheswaran, seconded by Suignard.

Approved 12, 2, 1

Action 73-21 for Freytag: Prepare written proposal for Keycap cartouche, screen cartouche (to support composable keyboard symbols).

5. W3C issues

5.A. Registration of (a) MIME charset(s) corresponding to Unicode 2.0

Roberts: "Unicode" without qualification.

Becker: In MIME context, customers believe Unicode = UTF-8.

Whistler: Battle of MIMEs – most successful way is to distinguish UTF-8 & UTF-16 as different forms of Unicode.

Freytag: If we go with MIME registration, how does this affect definition in the book?

Whistler: Two different ways to highlight UTF-16 as a term.

Motion [#73-M17]: That the UTC approves MIME char set designations for the Unicode Standard, Version 2.0 as follows: a) “UTF-16” is the primary designation; b) “UNICODE UTF-16” is to be registered as an alias for “UTF-16”. The exact syntax re hyphens vs. spaces is to be determined by MIME registration conventions. In addition, “UNICODE UTF-8” is to be registered as an alias for “UTF-8”.

Moved by Freytag, seconded by Jenkins.

Approved unanimously

Action 73-51 for Goldsmith: Register “UTF-16” as designation for the Unicode Standard, Version 2.0.

Action 73-52 for Goldsmith: Register “UNICODE UTF-16” as alias for “UTF-16”.

Action 73-53 for Goldsmith: Register “UNICODE UTF-8” as alias for “UTF-8”.

Action 73-54 for Goldsmith: Withdraw registration for “UNICODE-1-1”.

Motion [#73-M18]: For Version 3.0, treat UTF-8 as true part of Unicode. Document UTF-16 on the same level.

Moved by Freytag, seconded by Becker.

Unanimous.

Action 73-55 for Editorial: For Version 3.0, the editorial group is directed to a) treat UTF-8 as true part of Unicode; and, b) document UTF-16 on the same level

6. Other business

6A. SGML Base Set

Action 73-56 for Adams: Respond to Wolf’s mail on SGML base set with proposal from Bruce Paterson’s re collection 300.

6B. ISO/IEC CD 14651 (Default Sorting)

Davis and LaBonté discussed CD at length. Includes definition of collation that is a normative part of the standard. Does not include Asian and Indic collation.

Winkler: Input from Japan (from Sato-san) will be in next version.

Freytag: Is it the view of this committee that URO order is desired default order for Han? With extension work, is combined URO order desirable? Are two sub-orderings acceptable for Han?

Whistler: Part of the problem with the CD is that it is specified like POSIX character set. Impossible to understand without implementing.

Freytag: Wrong approach for a standard. Standard needs to be human readable.

Whistler: Will vendors who implement Unicode publish default collation in their own hook? As specified, CD is not consistent with Unicode canonical reordering – very serious fault. Default treatment for Indic scripts is to treat as “ignorables” – absolutely wrong. Worse than if you specified Unicode value as primary.

McGowan: Irresponsible to treat whole scripts as “ignorable.”

Whistler: Should we try to fix, or should we define a Unicode ordering?

Freytag: Need to give feedback. Do a strawman proposal and present that as an alternative.

Winkler: Editors are waiting for input from NSBs on specific scripts.

Whistler: US has requirement for default collation for all scripts. Basic principles: Three basic levels of Latin: case, accents; and separate scripts. Implementation in the document is inadequate and lacking in detail.

Whistler: A person used to a special sort may also be familiar with how certain characters tend to be mis-sorted in other sorts.

McGowan: Aiming for reasonable comprehension.

Whistler, Bishop, Roberts and Texin expressed interest when Freytag asked for volunteers; McGowan recommended Edberg.

Whistler: Instead of POSIX approach, start with Unicode data file, 38K characters. Can post and ask for input on subordering.

Action 73-57 for Whistler: Ask Jenkins to allocate “UTC private” section of FTP site to support work on default ordering of Unicode characters. .

Action 73-58 for Whistler: Make work on default orderings available for UTC comment.

Freytag: Default ordering is a line item in a larger standard. Line item needs to be self-consistent.

McGowan: Issue of one pass or multiple pass algorithm.

Whistler: First thing to do is effective list of expected output, unambiguously defined. Should also specify kind of detail handled in this CD for accents, case and some other features.

Long: Combining characters – do you need multiple passes?

Whistler: Yes, weight combining characters with respect to each other. Precomposed and canonical equivalents should not sort differently.

Action 73-59 for Winkler: Invite Alain LaBonté to special meeting on default ordering to be held on afternoon of Friday, December 5.

Freytag: We are interested in a solution and applying it to set of 38K characters. Unlike the work with WG2, the Consortium is not going to be involved deeply in the work of WG20.

Uma: What do we want in the end? Default ordering. Input into WG20 work.

Freytag: Need for L2 representation at WG20. US participation would increase if WG20 meetings held as precursor to WG2.

Hiura: Does not want 2 different standards. Not many UNIX vendors are involved.

Whistler: If we can specify the default collation in detail, can specify POSIX order.

Texin: Frame conversation in a particular context. Limit to specific scripts.

Winkler: WG20 has a timetable. By mid-October, produce new document for balloting in JTC1. If document goes out for ballot, in all probability will be accepted.

Action 73-60 for Winkler: Talk to Alain LaBonté to postpone next phase of 14561 in order to incorporate common understanding.

Freytag: Need international input from other UNIX vendors.

Action 73-61 for Whistler: Provide work-to-date document to Winkler for WG20 meeting (November 16-20 in Cairo).

Action 170-62 for Winkler/Aliprand: Add item to L2 agenda: Output of working group on default ordering as US contribution to WG20.

6C. Publicly Available Specification (PAS) Submission

Whistler: CC type should be coordinated with Unicode character properties (cf. as Java does too).

Uma: POSIX model does not have placeholder for all properties.

Winkler: WG20 has asked Unicode to be able to use property tables.

Freytag: Unicode Consortium answer: Preference is to reference, rather than reprint.

Winkler: Unicode Consortium is not PAS submitter, so ISO standards cannot reference the Unicode Standard.

Action 73-63 for Unicode Officers: Investigate establishing the Unicode Consortium as PAS.

Freytag: Officers did not have concrete case where this would be useful.

Action 170-64 for Winkler: On behalf of WG20, reply to letter from Unicode Consortium (dated January 22nd 1997) re WG20 request to use Unicode properties. In particular, clarify mutual understanding of “verbatim”.

Action 73-65 for Whistler: Send information about PAS application procedures to Freytag and Aliprand (for Unicode Officers).

6D. Cultural Workshop

JTC1 is rethinking its work relative to market forces. Unclear to JTC1 how work of SC2, SC18/WG20 and SC20/WG8, and WG9 should be reorganized. Winkler and Ksar have both volunteered to be on organization committee. Proximity in time with WG2 would not be possible.

Freytag: As liaison organization to SC2 and which has interacted with the other 2 groups, Unicode Consortium would be interested in participation. British Columbia is preferred site. Prefer SC2 as organizer.

Action 73-66 for Aliprand: As Chair of UTC, write to Chairs of SC2, SC18 and SC20 re Cultural Workshop. Mention Unicode Consortium’s relationship/involvement with all groups; stress that we are representing large companies which are active implementers of standards from all the SCs. Express preference for British Columbia as site; SC2 as organizer. Copy to NCTIS.

Action 170-67 for Winkler/Aliprand: Include discussion of recommendation for comparable letter from L2 Chair on agenda of L2 meeting (L2 is TAG for two of the SCs and worked with SC18 on the character-glyph model).

6E. NP for a Technical Report on a character set model [L2/174]

Freytag: To prioritize our effort, would recommend that this group not participate.

Uma: Lack of problem definition; lack of clearly defined requirements. Need for Unicode Consortium to be PAS.

Action 73-68 for Adams: Send copy of document characters and character set to Umamaheswaran.

Action 73-69 for Umamaheswaran: Gather information on characters and character sets to formulate response.

Action 73-70 for Aliprand: Include formulation of UTC positions for IRG December meeting. Place on agenda.

6F. Annual Meeting

Strong candidates: Industry vision and Board of Directors experience are criteria. Act as individuals.

Action 73-71 for All Corporate Member Representatives: Seek suitable candidates for Board election in December

7. Review of motions to L2, and action items

7A. Mongolian

China to set up ad hoc meeting on Mongolian in order to add ten (10) variant selector characters. We did not receive information in adequate time to send to anyone. China is still having difficulty working with Mongolia. Oliver Corff was at WG2 meeting. [There is] still

no consensus on the presentation forms. Corff said number of selection operators could be reduced.

7B. Khmer

Khmer could go to pDAM by March if we get it in.

7C. Burmese

Action 73-72 for Bishop: Prepare complete proposal for Burmese for December meeting.

Whistler: Current assumption is that Burmese and Khmer will not be in Version 3.0.

Action 73-73 for McGowan: Send all information on Burmese to Bishop.

Whistler: Upper case forms may be spurious.

7D. Revision of Plane 14 proposal

Action 73-74 for Whistler: Notify IETF participants in Plane 14 ad hoc meeting about outcome of this meeting: send them revised Plane 14 proposal.

Action 73-75 for Adams: Make revised Plane 14 proposal into Internet Draft, with UTC and IETF representatives as co-authors. Arrange with Chris Newman for formatting.

Action 73-76 for Adams: Make revised Plane 14 proposal into joint Unicode/IETF proposal to WG2.

Action 170-77 for Aliprand/Winkler: Review SC2 resolutions and include them on the December L2 meeting agenda (see M07.28 as an example).

Appendix : List of Attendees

Name	Company	L2 8/4/97	UTC 8/4/97	E-mail
John Jenkins	Apple	P	CM	jenkins@apple.com
Rick McGowan	Apple	x	CM	rmcgowan@apple.com
Waiman Long	Digital	P	CM	longman@zk3.dec.com
Don Carroll	Hewlett- Packard	A	CM	mike_ksar@hp.com
Mike Ksar	Hewlett- Packard	P	CM	dcarroll@sea.hp.com
Lisa Moore	IBM	A	VP, CM	lisam@us.ibm.com
V.S.	IBM	P	CM	umavs@ca.ibm.com
Umamaheswaran				
Brendan Murray	IBM-Lotus	x	AM	brendan_murray@lotus.com
Tatsuo Kobayashi	Justsystem		CM	tatsuo_kobayashi@justsystem.co.jp
John McConnell	Microsoft	x	CM	johnmcco@microsoft.com
Mike Kernaghan	Microsoft	x	VP, CM	mikekern@microsoft.com
Murray Sargent	Microsoft	P	CM	murrays@microsoft.com
Gary Roberts	NCR	x	CM	gary.roberts@sandiegoca.ncr.com
Lloyd Honomichl	Novell	P	CM	lloyd_honomichl@novell.com
Tex Texin	Progress	P	AM	texin@progress.com
Joan Aliprand	RLG	P	CM	br.jma@rlg.org
Cleo O'Brien	SGI	O	CM	cleo@sgi.com
Marc von Holzen	SGI	x	CM	holzen@engr.sgi.com
Glenn Adams	Spyglass	P	CM	glenn@spyglass.com
Bing-Meng Hong	Sun	x	CM	bing.hong@eng.sun.com
	Microsystems			
Chang Jung	Sun	x	CM	chjung@eng.sun.com
	Microsystems			
Hideki Hiura	Sun	x	CM	hiura@eng.sun.com
	Microsystems			
Ken Whistler	Sybase	P	CM	kenw@sybase.com
Asmus Freytag	Unicode	P	VP, AM	asmusf@ix.netcom.com
Julie Allen	Unicode	x	O	julie@unicode.org
Arnold Winkler	Unisys	P	CM	arnold.winkler@unisys.com
Joe Becker	Xerox	P	CM	becker.osbu_north@xerox.com

Legend: P - primary, A - alternate, O - observer, L - liaison, X - ex-officio, x - present
 CM - corporate member, AM - associate member, VP - vice president, Unicode, Inc.