At the request of our convener Mr. Mike Ksar, I have packaged the set of foils (modified slightly) that I had presented at the National Workshop on Unicode, New Delhi, Sept 24-26, 2003, organized by the Ministry of Information and Communication Technology, India. Some of you involved with JTC1/SC2/WG2 and the Unicode Technical Committee may find it of some use.

In particular, slide number 4 of the second presentation – on page 14 – titled 'Framework for Discussion' was also used in WG2 meeting M44 during our ad hoc on Tibetan. It is a gist of the principles to follow while proposing additions or changes to the standard.
Unicode and ISO/IEC 10646

V.S. Umamaheswaran
umavs@ca.ibm.com
IBM Toronto Lab, Canada

Topics

- Unicode and ISO/IEC 10646
- UCA and 14651
- Processes
- Guidelines for Proposals
- Organize the Expertise
### Unicode and ISO/IEC 10646

<table>
<thead>
<tr>
<th></th>
<th>Unicode</th>
<th>10646</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code Space</td>
<td>0 to x10FFFF</td>
<td>0 to x10FFFF*</td>
</tr>
<tr>
<td>Repertoire</td>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>Supp. Planes</td>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>BMP non CJKV</td>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>BMP CJKV</td>
<td>Single Col</td>
<td>CJKV Cols</td>
</tr>
<tr>
<td>Chart Creation</td>
<td>Common DB</td>
<td>Common DB</td>
</tr>
</tbody>
</table>

#### Row/Cell

<table>
<thead>
<tr>
<th>Hex code</th>
<th>C</th>
<th>J</th>
<th>K</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>078/000</td>
<td>G· Hanzi</td>
<td>-T</td>
<td>Kanji</td>
<td>Harja</td>
</tr>
<tr>
<td>4E00</td>
<td>0-523B</td>
<td>1-4421</td>
<td>0-306C</td>
<td>0-8C69</td>
</tr>
<tr>
<td></td>
<td>0-5027</td>
<td>1-3601</td>
<td>0-1576</td>
<td>0-7673</td>
</tr>
</tbody>
</table>
### Unicode and ISO/IEC 10646

<table>
<thead>
<tr>
<th></th>
<th>Unicode</th>
<th>10646</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Book Style</td>
<td>ISO Style</td>
</tr>
<tr>
<td>Conformance</td>
<td>=Level 3</td>
<td>Levels 1, 2, 3 (use 3 for Indic)</td>
</tr>
<tr>
<td>BiDi</td>
<td>Defined</td>
<td>Refers to Unicode</td>
</tr>
<tr>
<td>Normalization</td>
<td>Defined</td>
<td>Refers to Unicode</td>
</tr>
</tbody>
</table>

2003-09-25
Session 10, National Workshop on Unicode, New Delhi

---

### Unicode and ISO/IEC 10646

<table>
<thead>
<tr>
<th></th>
<th>Unicode</th>
<th>10646</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combining</td>
<td>Property + TRs+ Text</td>
<td>List + Minimal Info</td>
</tr>
<tr>
<td>Format Chars</td>
<td>Property</td>
<td>Some Listed</td>
</tr>
<tr>
<td>Script Info</td>
<td>Lot of Detail</td>
<td>Minimal</td>
</tr>
<tr>
<td>Annotations</td>
<td>Many more</td>
<td>Some in Annex</td>
</tr>
<tr>
<td>Naming Rules</td>
<td><em>uses 10646</em></td>
<td>Defined</td>
</tr>
</tbody>
</table>

2003-09-25
Session 10, National Workshop on Unicode, New Delhi
Unicode and ISO/IEC 10646

<table>
<thead>
<tr>
<th>Properties + Processing Rules</th>
<th>Defined</th>
<th>Out of scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTF-8,-16, -32/UCS4</td>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>Compressions</td>
<td>Defined</td>
<td>Not included</td>
</tr>
</tbody>
</table>

Conforming to Unicode will automatically conform to 10646 Level 3 plus lots more

Unicode Collation Algorithm and ISO/IEC 14651

- Synchronized with Each Other
- Share same Concepts for Weights Categories and Tailoring
- Tailoring Required in Both
- Default Weights and Repertoire Identical in Both - generated from the same data base
- 14651 Editions + Amds versus UCA Versions

Conforming to UCA will also conform to 14651 plus more functions
## Processes

<table>
<thead>
<tr>
<th>Stage ⇒</th>
<th>Initial proposal</th>
<th>Provisional acceptance</th>
<th>Final acceptance</th>
<th>Pipeline and Unicode Database</th>
<th>Beta Publication Web plus Book Database</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In process within UTC</td>
<td>Further progression</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WG2 Stage ⇒</th>
<th>Initial proposal</th>
<th>Provisional acceptance</th>
<th>Final Acceptance</th>
<th>WG2 Bucket</th>
<th>SC 2 Ballot</th>
<th>JTC 1 Ballot</th>
<th>ITTF Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In process within WG2</td>
<td>Further progression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2 Ballots Draft, Final 12-18 months
Processes

- UTC has additional procedures for preparing and processing Technical Reports

- See FAQ page at Unicode site

Processes

- Membership in SC2
  - National Bodies
    - Ex: INCITS in USA, SCC in Canada, BIS in India
    - Roster on SC2 site [www.dkuug.dk/JTC1/SC2](http://www.dkuug.dk/JTC1/SC2)

- Membership in UTC
  - Review by all members and experts
  - Voting by Corporate Members
    - Government of India is a Corporate Member
    - Roster on Unicode site.
Proposal Guidelines

Do your homework
? Check if Already encoded ?
(see http://www.unicode.org/standard/where/)

- Check Charts in Unicode V4
- Also charts in TRs –
  - TR15 Normalization charts
  - TR10 Collation charts
  - TR21 Case map charts
  - TR24 Script charts
- or for legacy sets ICU Charmaps or equivalents

May be in a block with recognized name ..
- Search Nameslist file in Unicode Database
  Name could be in Annotations
- Shape in standard can be a variant
  (see handout page 2)
- Is it a Glyph (from a Font for example?)
  http://www.unicode.org/reports/tr17/#Characters vs. Glyphs
  and TR 15285 – Character Glyph Model
  http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm??Redirect=1
Proposal Guidelines

- Character may be under consideration
- Look in Unicode Pipeline
  http://www.unicode.org/alloc/Pipeline.html

- Check if previously considered and rejected -
  http://www.unicode.org/alloc/rejected.html

- Also for any accepted pending scripts:
  http://www.unicode.org/pending/pending.html

Proposal Guidelines

Do your homework

For entire script - check out the ROADMAPS:

http://www.unicode.org/roadmaps
http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html

Already encoded- **Bold text** in Roadmap
proposal accepted

- **(Bold text between parentheses)**
  under consideration (Text between parentheses)
exploratory ¿Text between question marks?
possible future – no suggestions ???
hot links for latest proposal included
Proposal Guidelines

Do Your Homework

? Can the character be represented as sequences?

*Remember no Duplicate Representation*

- Indic conjuncts fall into this category
- Check out Chapter 9 of Unicode 4.0
  (Examples in handout last 3 pages)
- [http://www.unicode.org/standard/where/](http://www.unicode.org/standard/where/), and
- [http://www.unicode.org/faq/char_combmark.html](http://www.unicode.org/faq/char_combmark.html)
Proposal Guidelines

Other proposals may exist elsewhere in draft form especially with archaic / minority scripts

Ex: Kharoshthi, Brahmi, Surashtrian .. proposals

Ask / network on the public discussion lists http://www.unicode.org/consortium/distlist.html

indic@unicode.org is set up for Indic
WHEN YOU ARE CERTAIN A NEW PROPOSAL IS WARRANTED

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPertoire OF ISO/IEC 10646

Please fill all the sections A, B and C below.
(Please read Principles and Procedures Document for guidelines and details before filling this form.)
See http://www.dkuug.dk/JTC1/WG2/docs/summaryform.htm for latest Form.
See http://www.dkuug.dk/JTC1/WG2/docs/roadmaps.htm for latest roadmaps.

Prepare the Proposal Summary Form
www.dkuug.dk/JTC1/SC2/WG2/summaryform.htm

Proposal Guidelines
Proposal Summary Form

- Contains several questions to be answered
- See Submitter’s Responsibilities in Form
- Most related to the previous checking steps
- Additional Information to assist in evaluation by UTC and WG2
  - Unicode Properties, Evidence of use, References
  - Information about submitters & others consulted
  - Preferred location, Glyphs/Font for publications

Facilitates evaluation by UTC, WG2 and other experts worldwide
Organize the Experts

Some Observations / Suggestions

- Workshops are Educational
- Formal review and Consensus Process helps in consolidated national positions
- Participation by Regulators (Governments), User Communities and Industry – is important
- Possibly re-activate BIS working group
- Be present at UTC and ISO committees with some Continuity of Participation
- Maximize use of e-discussion lists – free dialog
- Continue to Prepare and disseminate Resources and Education material
Unicode Issues
Dravidian Group
*Kannada, Malayalam, Tamil & Telugu*

V.S. Umamaheswaran
(umavs@ca.ibm.com)
IBM Toronto Lab, Canada

Characters added in V4.0
(in response to latest request from India)

0CBC KANNADA SIGN NUKTA
0CBD KANNADA SIGN AVAGRAHA

(from TNG Keyboard Layout)

0BF3 TAMIL DAY SIGN (Naal)
0BF4 TAMIL MONTH SIGN (Maatham)
0BF5 TAMIL YEAR SIGN (Varudam)
0BF6 TAMIL DEBIT SIGN (Patru)
0BF7 TAMIL CREDIT SIGN (Varavu)
0BF8 TAMIL AS ABOVE SIGN (Merpadi)
0BF9 TAMIL RUPEE SIGN (Rupai)
0BFA TAMIL NUMBER SIGN (Enn)
Additions in V4.0

Additions to text of Chapter 9 to address several of the requests in latest input from Gov of India and from other inputs.

Some examples:

- Added text - where users are to look for the DANDA and DOUBLE DANDA characters (in the Devanagari block).
- 0 CCD  KANNADA SIGN VIRAMA
  * preferred name is halant

See handout charts and names list for Annotations added.

Framework for discussion

- Respect Stability Policy
  - No removal of existing character
  - No relocation / reordering of existing code positions
  - No name changes
  - No changes to existing canonical equivalences / normalization
  - No new multiple spellings
  - No new encoding model
  - If sequences satisfy the requirement no new character needed (Ch 9)
- Suggestions that can be entertained
  - Text for FAQ, Tech Note, Standard - for better understanding
  - Possible new sequences
  - Annotations where appropriate
  - New characters only with evidence
  - Deprecation only with strong justification
Packaging Results of Discussion

For each Dravidian Script categorize issues as:

- Proposal for FAQ material
- Proposal for Unicode Technical Note
- Proposal for Explanatory text
- Proposal for Annotation
- Proposal for Deprecation
- Proposal for New Character

Assign an Owner for Each