ISO/IEC JTC 1 N 9436

2008-11-25

ISO/IEC JTC 1
Information Technology

Document Type: Other Document(Defined)

Document Title: SC 2 Business Plan presented to JTC 1 at Plenary in Nara, Japan

Document Source: JTC 1 Secretariat

Reference:

Document Status: For Information

Action ID: For Information

Due Date:

No. of Pages: 18

Secretariat, ISO/IEC JTC 1, American National Standards Institute, 25 West 43rd Street, New York, NY 10036; Telephone: 1 212 642 4932; Facsimile: 1 212 840 2298; Email: lrajchel@ansi.org

ISO/IEC JTC 1
SC 2 Business Plan
(Coded Character Sets)

TATSUO L. KOBAYASHI
SC 2 CHAIR
FOR JTC 1 PLENARY NARA, JAPAN 2008

SC 2 Scope

- Scope
  - Standardization of graphic character sets and their characteristics, including string ordering, associated control functions, their coded representation for information interchange and code extension techniques. Excluded: audio and picture coding.
- WGs and Standards
  - WG 2 - 16 bit and 32 bit coding
    - Ideographic Rapporteur Group (IRG)
    - Universal Coded Character Set (UCS): 10646
  - SC 2 direct
    - 7 bit and 8 bit coding: IS 646, 6937, 8859
    - Code structure and code extension technique: 2022,4873
    - Control functions: 6429
    - String ordering: 14651 (developed and maintained by OWG-SORT)
    - Registration procedure of coded character sets: 2375
      - "Procedure for Registration of Escape Sequences"

SC 2 Meetings

- Since Last JTC 1 Plenary Meeting
  - Plenary meeting:
    2008-04-25, Redmond, Washington, U.S.A.
  - WG 2 and OWG-SORT:
    2008-04-21/25, Redmond, Washington, U.S.A.
    2008-10-13/17, Hong Kong SAR, China
  - IRG :
    2007-11-12/16, San Jose, California, U.S.A.
    2008-6-9/13, Bussan, Korea

SC2 Meetings (Future)

- Future Meetings
  - SC 2 Plenary with WG 2 and OWG-SORT: 2009-10-26/30, Tokushima, Japan
  - WG 2 and OWG-SORT:
    2009-04-20/24, Dublin, Ireland
  - IRG :
    2009-06-15/19, Hong Kong SAR, China
    2009-11-??, Taipei, Taiwan Island

SC2 Members

- 29 P-members
- 19 O-members
- Internal Liaisons
  - SC 22, SC 29, SC 31*, SC 32*, SC 34, SC 35
  - ISO/TC 37/SC 2, ISO/TC 46/SC 4, ISO/TC 211

  *=to be terminated
- External Liaisons
  - IETF/ISOC, ITU-T, TCA, UC Berkeley, UNICODE, W3C, EC, UN-ECE, UNCTAD, WIPO, WMO

SC 2 projects (including latest updates)

- Total Number of Projects:
  - 58 including all subprojects
- New Subprojects
  - ISO/IEC 10646: 2003/Amd. 7
- Achievements
  - ISO/IEC 10646: 2003/Amd 3 → Published
  - ISO/IEC 10646: 2003/Amd 4 → Published
  - ISO/IEC 10646: 2003/Amd 5 → FDAM
  - ISO/IEC 10646: 2003/Amd 6 → 2nd PDAM
  - ISO/IEC 14651 2nd edition → Published
  - ISO/IEC 14651: 2007/Amd 1 → PDAM
- Active Projects:4
  - 10646: 2003/Amd 5
  - 10646: 2003/Amd 6
  - 10646: 2003/Amd 7
  - 14651: 2007/Amd 1

Supportive position of SC 2 for the Language Diversity in Cyberspace

- **RESOLUTION M15-02: Support for the 2008 International Year of Languages by the United Nations**
- SC 2 supports the designation of 2008 as the International Year of Languages by the United Nations. The International Year of Languages will draw attention to the diversity of the world's languages, the importance of preserving these languages, and the need to provide access to them in cyberspace. The key to providing access is to be sure the scripts are in the international character encoding standard 10646 (/Unicode) and sorting order standard 14651. SC 2 requests JTC 1 to support the International Year of Languages.

Beyond Encoding Scripts

- United Nations Declared 2008 the International Year of Languages
- What SC 2 can do for the Language Diversity in Cyberspace
- Encoding diverse scripts and characters in ISO/IEC 10646 is not good enough
- So, what can we do?

UCS's Support for African Scripts

Non-Latin/Arabic Scripts

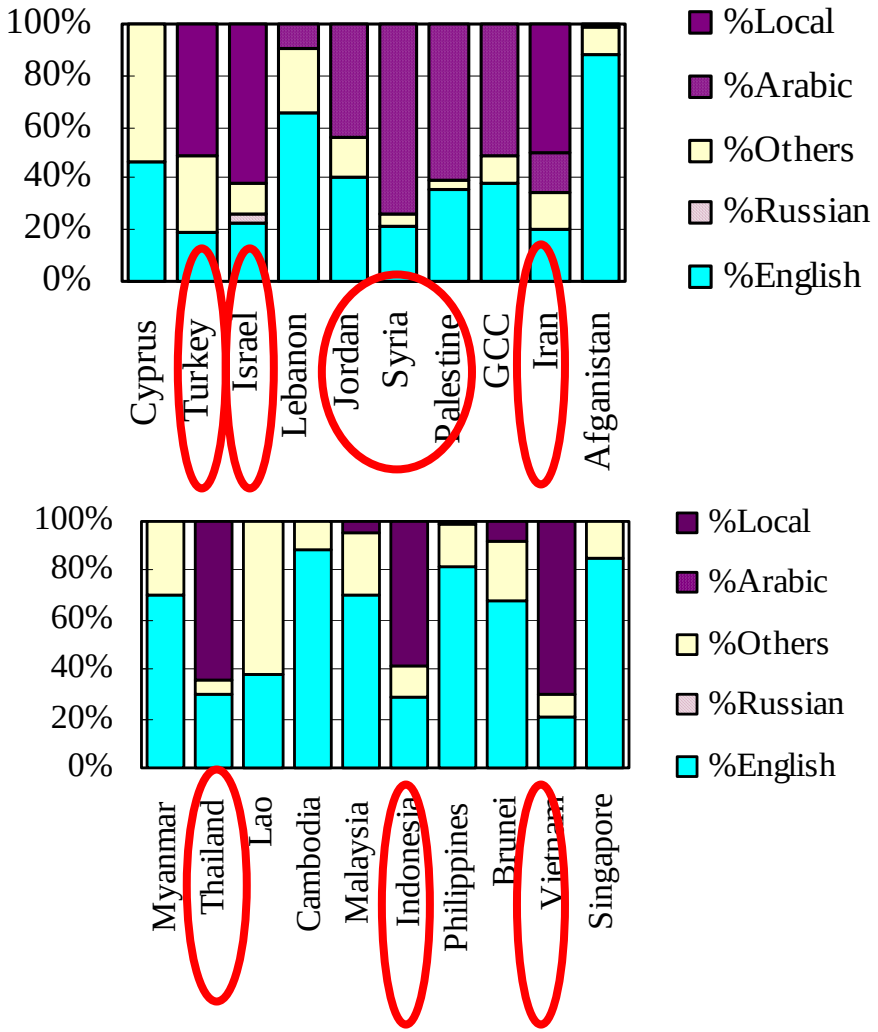
- Ethiopic
- Osmanya
- Tifinagh
- N'Ko
- Vai

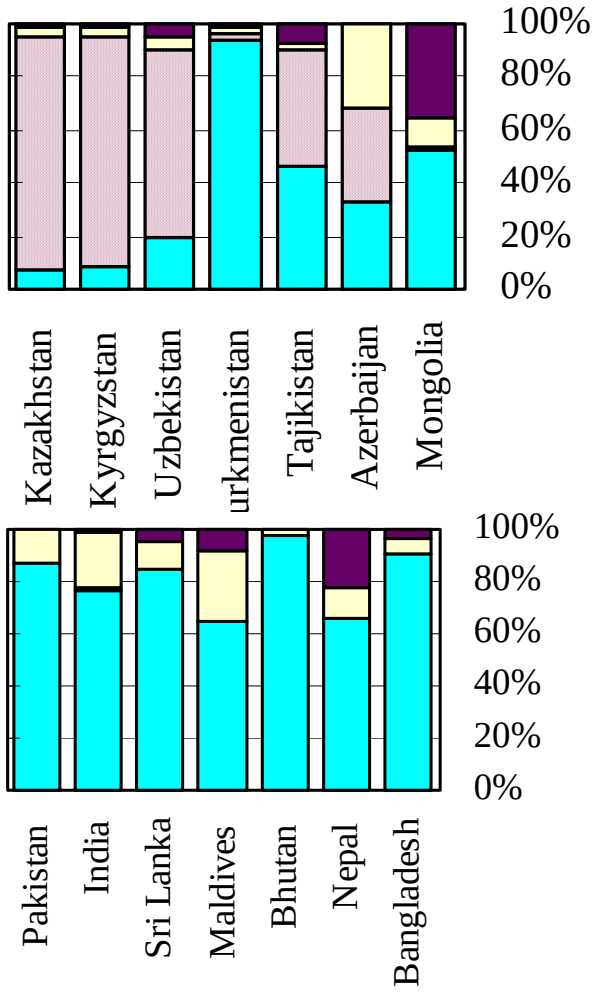
African language support in ICT

Issues with N’Ko

- Right-to-left script
- Not in Uniscribe
- Alternative: SIL’s Graphite; were some issues with Graphite r-to-l; now issues with Graphite + OpenOffice
- Most users run Windows

Asian Usage of Mother Language in Web Pages





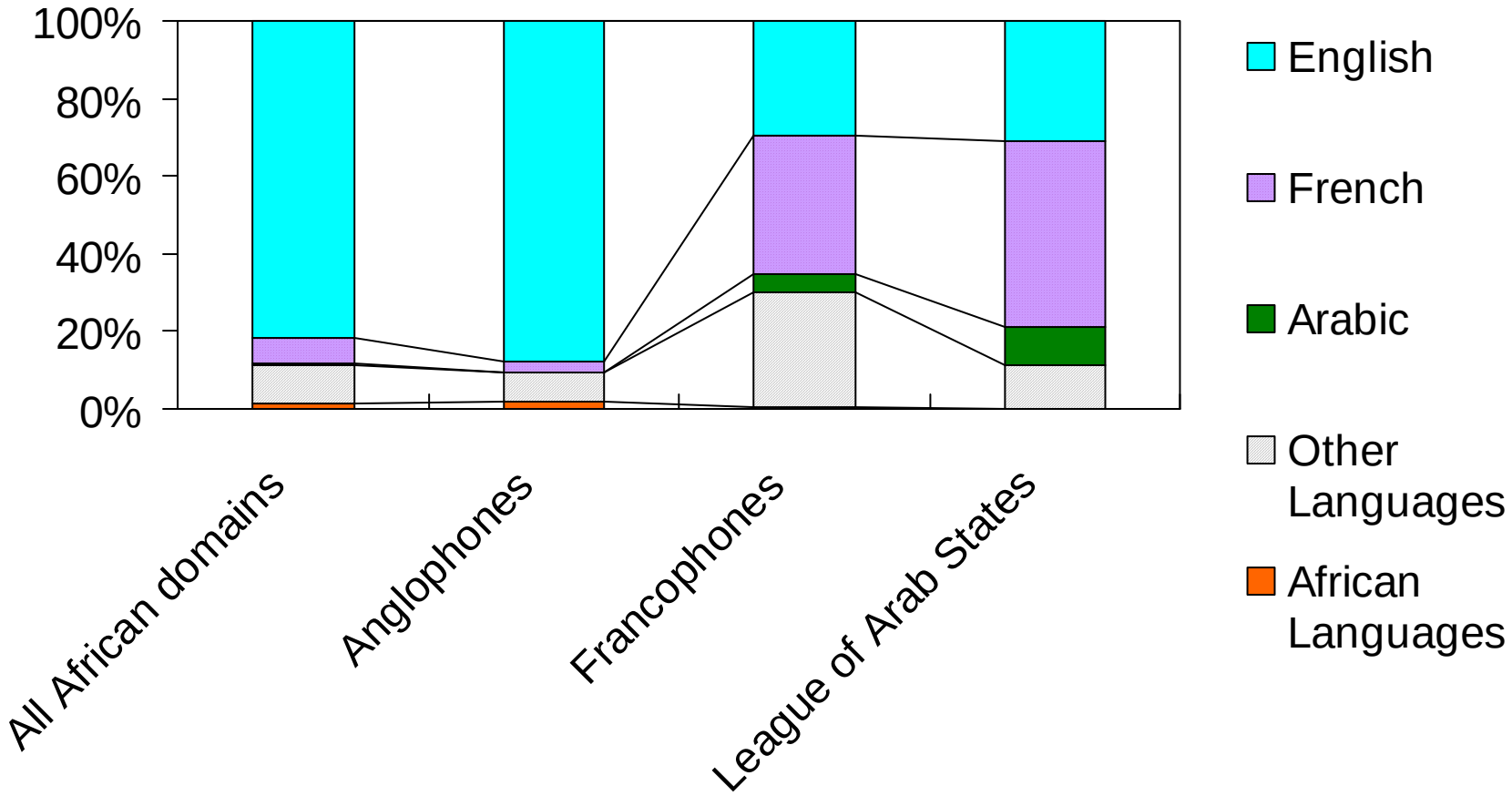
as of June 2006

Estimated number of pages
Top 10 Asian languages

Language	Script	Speaker population	pages
Hebrew	Hebrew	4,612,000	11,957,314
Thai	Thai	21,000,000	7,752,785
Turkish	Latin	59,000,000	3,959,328
Vietnamese	Latin	66,897,000	2,006,469
Arabic	Arabic	280,000,000	1,671,122
Tatar	Latin	7,000,000	1,575,442
Farsi	Latin	33,000,000	1,293,880
Javanese	Latin	75,000,000	1,267,981
Indonesian	Latin	140,000,000	866,238
Malay	Latin	17,600,000	432,784

Note: Chinese, Korean & Japanese domains are excluded. As of October 2006 as of June 2006

African Usage of Mother Language in Web Pages



as of October 2006

Estimated number of pages
Top 10 African languages

language	script	speaking region	pages
Malagasy	Latin	Madagascar	5,382
Swahili	Latin	Tanzania	5,170
Afrikaans	Latin	South Africa, Namibia	1,775
Krio	Latin	Gambia, Sierra Leone	1,575
Kinyarwanda	Latin	Rwanda	1,059
Shona	Latin	Zimbabwe, Mozambique	538
Somali	Latin	Somalia	396
Siswati	Latin	Swaziland	335
Oshiwanbo	Latin	Namibia, Angola	264
Rundi	Latin	Burundi	252

Note: South Africa is excluded. As of October 2006

Encoding Chaos leads to delay of localization

Language	Standard encoding and its share	Examples of other encodings found [footnote]
Turkish	ISO 8859 (99.5%)	
Hebrew	ISO 8859 (87.7%)	
Vietnamese	UTF-8 (96.4%)	TCVN, VIQR, VPS
Thai	TIS 620 (97.3%)	
Mongolian	UTF-8 (95.5%)	Latin-Cyrillic
Sinhala	UTF-8 (44.5%)	Metta, Kaputa, etc.
Telugu	UTF-8 (16.6%)	Shree, TLH, etc.
Tamil	UTF-8 (14.9%)	Amudham, Kumudam, Shree, Vikatan, etc.
Burmese	UTF-8 (0.7%)	WinResearcher, etc.

note: Local proprietary encodings are shown in this table by names of font (families). as of June 2006

My tentative hypothesis

- The promotion of Universal Coded Character Set serves the acceleration of usage of multiple languages in Internet
- What we should do next is to reach out to the actual community

What we can do to accommodate the Language Diversity

- Effort to encode more minority and historic scripts in ISO/IEC 10646
- Effort to promote development of fonts for each script
- Effort to promote implementations of more scripts in ISO/IEC 10646
- Effort to reach out to actual user communities for daily use