

To: WG2 and Unicode Technical Committee
From: Debbie Anderson, SEI, UC Berkeley
Title: Mongolian Ad Hoc Report (Hohhot, Inner Mongolia)
Date: 29 September 2017

收件: WG2 和 Unicode 技术委员会
发件: Debbie Anderson, 伯克利加州大学 SEI
主题: 蒙古文特别会议报告 (内蒙古呼和浩特)
日期: 2017 年 9 月 29 日

This is an ad hoc report of the WG2 ad hoc meeting on Mongolian that took place in Hohhot, Inner Mongolia, China, 23–25 September 2017.

本文是 2017 年 9 月 23 至 25 日在中国内蒙古呼和浩特举行的 WG2 蒙古文特别会议的特别报告。

In attendance:

Chen Zhuang (China Electronics Standardization Institute), Prof. Quejingzhabu (Inner Mongolia University), Prof. Nashunwuritu (Inner Mongolia University), Prof. Menghejiya (Inner Mongolia University), Prof. Bao Haishan (Inner Mongolia University of Finance and Economics), Bai Shuangcheng (Inner Mongolia Academy of Social Science), Husile (Inner Mongolia Academy of Social Science), Prof. Wu Jian (Institute of Software, Chinese Academy of Sciences), Liang Jinbao (Ethnic Affairs Commission of the Inner Mongolia Autonomous Region), Liushisan (Inner Mongolian Publishing Group), Lü Jianchun (Jade Bird Huaguang), Zhang Guorong (Founder Group), Manduhu (Inner Mongolia University), Haila (Inner Mongolia University), Peng Daruhan (Inner Mongolia University), Peter Constable (Unicode Consortium Liaison to WG2; Microsoft), Michel Suignard (Editor of ISO 10646; WG2 Convenor), Debbie Anderson (UC Berkeley), Lisa Moore (Unicode Technical Committee Chair), Andrew Glass (Microsoft), Greg Eck (Greyson Translation Services), Liang Hai (individual expert), Zheng Weizhe (individual expert), Shen Yilei (individual expert), Badral Sanlig and Jamyansuren (Bolorsoft, Ulaanbaatar), B. Undraa (Mongolian Agency for Standardization and Metrology, Ulaanbaatar), Prof. D. Tumurtogoo (Mongolian Academy of Sciences, Ulaanbaatar), Byambaa Tsogt (Greyson Translation Services, Ulaanbaatar), Andrew West (individual expert), Jirimutu (Almas, Japan), Michael Everson (Everttype).

出席名单:

陈壮 (中国电子技术标准化研究院)、确精扎布教授 (内蒙古大学)、那顺乌日图教授 (内蒙古大学)、孟和吉雅教授 (内蒙古大学)、包海山教授 (内蒙古财经大学)、白双成 (内蒙古社会科学院)、呼斯勒 (内蒙古社会科学院)、吴健教授 (中国科学院软件研究所)、梁金宝 (内蒙古自治区民族事务委员会)、六十三 (内蒙古出版集团)、吕建春 (青鸟华光)、张国荣 (方正集团)、曼杜呼 (内蒙古大学)、海拉 (内蒙古大学)、彭鞑茹罕 (内蒙古大学)、Peter Constable (Unicode 联盟在 WG2 的联络人; 微软)、Michel Suignard (ISO 10646 编辑; WG2 召集人)、Debbie Anderson (伯克利加州大学)、Lisa Moore (Unicode 技术委员会主席)、Andrew Glass (微软)、Greg Eck

(Greyson 翻译服务)、梁海(个人专家)、郑维喆(个人专家)、沈逸磊(个人专家)、Badral Sanlig 和 Jamyansuren (Bolorsoft, 乌兰巴托)、B. Undraa (蒙古国标准化与计量局, 乌兰巴托)、D. Tumurtogoo 教授(蒙古国科学院, 乌兰巴托)、Byambaa Tsogt (Greyson 翻译服务, 乌兰巴托)、Andrew West (个人专家)、吉日本图 (Almas, 日本)、Michael Everson (Evertime)。

We met and the following participants made presentations:

- Prof. Quejingzhabu on N4880, one of the improved Phonetic models.
- Jirimutu on the vendor experience of Almas.
- Badral Sanlig on the vendor experience of Bolorsoft.
- Lü Jianchun on the vendor experience of Jade Bird Huaguang.
- Zhang Guorong on the vendor experience of Founder Group.
- Shen Yilei on the positional mismatches.
- Andrew Glass and Liang Hai on the Graphetic model and text processing layers.
- Liang Jinbao on suggested improvements to the Phonetic model.

The attendees of this meeting agreed upon the following points:

- The Mongolian script is an important script used today that has a rich and long history.
- There are a number of current problems, including typing errors, display errors, security concerns due to visual ambiguity, positional mismatches, missing glyphs, and underdocumented specifications.
- There are alternative approaches to solving these problems: a new Graphetic model or improved Phonetic models. The new Graphetic model, presented by Andrew Glass and Liang Hai, is an adaptation of the original Graphetic encoding standard that incorporates cursive joining behavior based on the UCS character/glyph model. One improved Phonetic model, pre-

我们举行了会议, 有下列与会者做展示:

- 确精扎布教授, 关于改良语音模型之一的 N4880 文档。
- 吉日本图, 关于 Almas 的厂商经验。
- Badral Sanlig, 关于 Bolorsoft 的厂商经验。
- 吕建春, 关于青鸟华光的厂商经验。
- 张国荣, 关于方正集团的厂商经验。
- 沈逸磊, 关于位置失配。
- Andrew Glass 和梁海, 关于字形模型以及文本处理的各种层面。
- 梁金宝, 关于对语音模型所建议的改进。

本次会议的出席者在下列要点上达成一致:

- 蒙古文是当今使用的重要文字, 有着丰厚且悠久的历史。
- 存在若干当前问题, 包括打字错误、显示错误、由视觉歧义引起的安全问题、位置失配、图形缺失、规范文档不足。
- 存在多个解决这些问题的替代方法: 新版字形模型或多个改良语音模型。由 Andrew Glass 和梁海展示的新版字形模型是基于 UCS 字符/图形模型并结合连写行为对原先字形编码标准的改编。梁金宝展示的改良语音模型之一完善

sented by Liang Jinbao, refines the current encoding model to improve consistency. Liang Jinbao's presentation also identified pros and cons to both approaches. Prof. Quejingzhabu presented a second improved Phonetic model.

- The Mongolian Working Group stated their commitment to the Phonetic model. The representatives from Mongolia also were committed to the Phonetic model. Representatives of the Unicode Technical Committee did not state a preference for either model, but the representatives expressed concerns with the current state of Mongolian text processing.
 - Whichever approach is taken to solve the issues, the long-term goal is to make Mongolian text interchangeable and accessible for future generations.
 - There are different layers of text processing implementation. We noted where the different encoding models distributed the inherent complexity of Mongolian script. In short, the Graphetic model imposes complexity on sides that involve phonetic information (such as the implementation of input and collation) and removes complexity from the encoding level and display side, whereas the Phonetic models shift complexity to the encoding and fonts. The two approaches have different advantages and disadvantages.
 - It is important to provide clear specifications for implementers and font providers, especially for those working outside China and Mongolia.
 - Those working on the Graphetic model will produce a test implementation of the Graphetic model and apply it to a corpus of Mongolian data to see if it works for all Mongolian text.
 - We agreed to share information (and knowledge) and collaborate in helping to resolve the problems.
- 了当前编码模型以改善一致性。梁金宝的展示还识别了两个方法的利弊。确精扎布教授展示了另一个改良语音模型。
- 蒙古文工作组声明对语音模型的坚持。来自蒙古国的代表也坚持语音模型。Unicode 技术委员会的代表没有声明对任何模型的偏好，但代表们表达了对蒙古文文本处理当前状况的担忧。
 - 无论选定哪个方法来解决这个问题，长期目标是让蒙古文的文本对后代们可交换且可获得可使用。
 - 文本处理的实现有各种层面。我们注意到了编码模型各自把蒙古文固有的复杂度分配在哪里。简单说，字形模型把复杂度施加到了涉及语音信息的方面上（比如对输入和排序的实现）并将复杂度从编码层和显示方面移除，而语音模型把复杂度转移到编码和字体。两个方法有不同的优势和劣势。
 - 为实现方和字体提供方提供清晰的规范很重要，尤其对于在中国和蒙古国外部工作的那些来说。
 - 做字形模型工作的几位将会制作字形模型的测试版实现，并将其应用到蒙古文的语料数据中，看是否适用于所有蒙古文文本。
 - 我们同意要分享信息（及知识）并合作以帮助解决问题。

Thanks are to due the Inner Mongolia University as well as the Henry Luce Foundation and Unicode Consortium, for providing support that made the meeting possible.

感谢内蒙古大学以及 Henry Luce 基金会和 Unicode 联盟提供支持使本次会议成为可能。



Group photo 合影