ISO/IEC JTC1/SC2/WG2 IRG N2623
Date: 2023-10-13

Source:	China
Meeting:	IRG#61
Title:	Activity Report
Status:	Member's submission
Actions required:	FYI
Distribution:	IRG
Medium:	Electronic
Pages:	2

1. Submitting plan for WS2024

China plans to submit around 1000 characters to WS2024, including 700 Chinese characters from the second edition of Hanyu Dazidian（《汉语大字典》）and other sources, and 300 Zhuang characters sorted by Guangxi University.

2. Development of GB 18030-2022/Amd 1 *Information technology -- Chinese coded character set--Amendment 1*

From 2023-04-06 to 2023-05-06, the draft of GB 18030-2022/Amd 1 has been opened for comment for the second time. Based on received feedbacks, China has made the following modifications to the draft and plans to publicly solicit opinions for the third time soon.

1) Repertoire and code points of "Supplementary Characters of Public Security Population Information Special Font"
In the light of strong need of citizen real-name certification in China, "Supplementary Characters of Public Security Population Information Special Font" (公安人口信息专用字库补充汉字), was added in GB 18030-2022/Amd 1. Through several rounds of review, the final repertoire consists of 622 characters. Since all of these characters, with the great help of IRG members and Unicode Consortium experts, have already been encoded in the CJK Unified Ideographs Extension I block (U+2EBF0~U+2EE5D), GB 18030-2022/Amd 1 moved them into the new code points to synchronize with UCS.

2) Addition of five new Ideographic Description Characters (IDCs)
GB 18030-2022/Amd 1 has also added five new IDCs (code points: U+2FFC~ U+2FFF, U+31EF) which are adopted in UCS.

3.Proposal for expanding G-source references

In the future, China plans to invite more influential Chinese academic institutions and companies to join in character collection work. Therefore, China believes it is necessary to establish some new G-source reference identifiers to cope with the growing number of character sources.

The list of new G-source references and their suggested acronyms are shown below.

1) GZK - Chinese Characters Repertoire Project (中华字库)
The project has collect a set of nearly 300,000 characters. The remaining uncoded characters will be submitted in the future according to the rules of IRG PnP.

2) GZH - Zhonghua Book Company (中华书局)
Zhonghua Book Company has participated in the early encoding work of UCS. In recent years, they established a database of Chinese ancient books. The total word count of the database has hit 3 billions and about 20,000 characters in it are not encoded.

3) GGT - National Library of China (中国国家图书馆)
The National Library of China has participated in the early encoding work of UCS. In recent years, it has established a database of ancient book bibliography and now is developing the database of full text. In the future, it will sorted and submitted uncoded characters to IRG.

4) GCW - Center for Toponym Research SISU (四川外国语大学地名研究中心)
The center collects place name related data from a variety of sources, including National Geomatics Center of China, national surveying and mapping data, new & old version of standard place names included in the local chronicles, new & old addresses in the registered residence management system of the Ministry of Public Security of China, and scattered place names being used by public. At present, the number of uncoded characters is approximately 800-1000.

5) GCCPP - Chinese Character Protection Project by Tencent (腾讯“汉字守护计划”)

The project is based on Tencent's Wechat app, and aims to collect characters which ordinary people used in their daily lives but cannot be processed by any electronic devices. In the first stage, over 300 uncoded characters have been found, among which 50 important characters related to place names, dialects, and technology are planned to be submitted in IRG WS2024.

6) GERS - Erudition Digital Technology Research Center (爱如生数字化技术研究中心)

As a Chinese company, Erudition (also called Eruson) is a pioneer in the construction of ancient book databases and has accumulated rich data of uncoded characters.

Besides, the previous G-source references for professional characters, such as GKJ (technology), GXM (personal name), GDM (place name), etc., will continue to be used in the future. It will be up to the submitter to decide whether to identify the characters through professional character reference series or submitter reference series.

END.