Document Type	Working Group Document
Title	Proposed additions to WG2 principles and procedures and Proposal Summary Form
Source:	V.S. Umamaheswaran, IBM Canada
Status	Expert contribution
Action	For consideration by WG2 and UTC
Distribution	ISO/IEC JTC1/SC2/WG 2 and UTC

This document proposes three additions to WG 2's P&P document and the Proposal Summary Form.

A. Contiguous encoding of decimal digits

When script-specific decimal digits are encoded in UCS, the decimal digits will be encoded contiguously and in order, with room left for missing digits so that, if digits are later used as part of a place-value notation (i.e. a decimal radix notation) they can be used in that manner. Exceptions are to be made only where (like numeric ideographs) the digits also serve as letters, or otherwise their use in decimal-radix notation can be safely excluded.

B. Information about use of standardized characters as part of a script's repertoire.

There is interest in gathering information about characters, such as punctuations, that are used in multiple scripts. The intent is to narrow down the particular scripts that such characters are most commonly used with. It is proposed to add a question to the Proposal Summary form along the following lines:

> "If the proposal is for a new script, identify already encoded characters that are used directly, or unified with characters (such as punctuation marks) that are part of the script's repertoire.

C. Information about confusable characters

Because UCS contains such a large number of characters and incorporates the varied writing systems of the world, incorrect usage can expose programs or systems to possible security attacks. In order to address this problem the Unicode Consortium has prepared two technical reports:

> - UTR #36: Unicode Security Considerations (http://www.unicode.org/reports/tr36/) describes some of the security considerations that should be take into account, and provides specific recommendations to reduce the risk of problems.
> - UTS #39 - Unicode Security Mechanisms (http://www.unicode.org/reports/tr39/ specifies mechanisms that can be used in detecting possible security problems.

Visually confusable characters lead to non-uniqueness in identifier strings and leads to problems such as spoofing on the internet – see section 2 on Visual Security Issues in UTR #36. UTS#39 describes the different kinds of confusables. It also includes data files on characters that can be confused with each other that was prepared by examining UCS for visually confusable characters within and across scripts. As new proposals come on board, it would be useful to have some information about confusability of newly proposed characters with standardized ones.

The proposers already have to ensure that a proposed character does not already exist, or they can be unified with etc. per the set of questions that are in the current proposed summary form. As an extension of that exercise it would be useful to gather information towards enhancing the data tables for visually confusable characters from the proposers. The information could be as simple as "similar to characters in script xxx or block yyy'. It could also be a listing of at least one character that is confusable citing already encoded character a proposed character could be confused with. The level of detail is at the discretion of the proposer.

Towards this goal, it is proposed to add an additional question along the following lines in the proposal summary form:

> "Include any information you can provide about characters in the standard that could be potentially confused with already standardized characters."