L2/13-079

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 (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** A proposal for encoding twelve localizable sentence markup bubble characters, together with a request for action to set up an encoding infrastructure for the encoding in a separate standardization document of complete localizable sentence markup bubbles using those twelve characters
2. Requester's name William Overington
3. Requester type (Member body/Liaison/Individual contribution) Individual contribution
4. Submission date: 27 April 2013
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: This document is a complete proposal.
(or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):
   Proposed name of script: Localizable sentence markup bubble characters
   b. The proposal is for addition of character(s) to an existing block:
   Name of the existing block: —-
2. Number of characters in proposal: 12
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary
   A-Contemporary B.1-Specialized (small collection) B.2-Specialized (large collection)
   C-Major extinct D-Attested extinct E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided? Yes
   a. If YES, are the names in accordance with the “character naming guidelines”
   in Annex L of P&P document? Unsure as the document is not available to the author at the time of writing this proposal
   b. Are the character shapes attached in a legible form suitable for review? Yes
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   William Overington
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   William Overington
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? No
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
   of proposed characters attached? No.
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes. The proposal addresses localization.

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

The information as known at present is in the pages of this document that follow this form.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? No
   If YES explain
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Not in relation to this specific proposal. There has been some discussion previously about the localizable sentence technology and its applications in email discussions, in a forum and in a mailing list. Some of the mailing list discussions were at a much earlier stage of research and development, before the present application possibilities were discovered. The forum discussions took place before the latest research developments.

   If YES, with whom?
   If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

This is a new technology for the future. There is not presently any user community.
   Reference:
4. The context of use for the proposed characters (type of use; common or rare)

If the invention becomes widely used, then use could be common.
   Reference:
5. Are the proposed characters in current use by the user community? No
   If YES, where? Reference:
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? No
   If YES, is a rationale provided?
   If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? Yes, preferably
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? No
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? Yes
    If YES, is a rationale for its inclusion provided? The ten digit characters look similar to ordinary digits, but they have a different function as they are control characters.
    If YES, reference: This document
11. Does the proposal include use of combining characters and/or use of composite sequences? No
    If YES, is a rationale for such use provided?
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? Yes
    If YES, describe in detail (include attachment if necessary)
    The special properties are described in the pages of the present document.

13. Does the proposal contain any Ideographic compatibility characters? No
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference:

This proposal is for the encoding of twelve characters to assist communication through the language barrier.

In addition an important part of this proposal is to request action by the committees to set up an encoding infrastructure for the encoding in a separate standardization document of complete localizable sentence markup bubbles using those twelve characters.

The characters are as follows..

LOCALIZABLE SENTENCE MARKUP BUBBLE OPENING BRACKET

LOCALIZABLE SENTENCE MARKUP BUBBLE CLOSING BRACKET

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT ZERO

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT ONE

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT TWO

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT THREE

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT FOUR

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT FIVE

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT SIX

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT SEVEN

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT EIGHT

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT NINE

Here are the glyphs, in the same order.

A localizable sentence would be encoded for interchange by a sequence commencing with a LOCALIZABLE SENTENCE MARKUP BUBBLE OPENING BRACKET, followed by one or more LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT characters, followed a LOCALIZABLE SENTENCE MARKUP BUBBLE CLOSING BRACKET.

For example, using an example from my research, from the collection of localizable sentences for enquiries after a disaster.

The enquirer is the brother of the first person that was named.

If localizable sentences technology comes into widespread use, then the typical display unit would not display the sequence of codes. The typical display unit would display the sentence that the sequence represents localized into the local language.

Typically, an end user composing a message would assemble a message from a cascading menu system where the choice of sentences available for use would be displayed in the language of the end user.

For an end user receiving a message, the message would first pass through an automated localization facility software module where any localizable sentences in the message would be localized and would then proceed to a display rendering software module so as to produce a display.

However, if there were no automated localization facility on that particular receiving system, or if the automated localization facility were turned off, then the sequence of LOCALIZABLE SENTENCE MARKUP BUBBLE characters would be displayed on the display screen, provided that there were font support for the LOCALIZABLE SENTENCE MARKUP BUBBLE characters.

In addition, if the font were an appropriate advanced format font and glyph substitution were turned on, then if an unmapped glyph for that sentence were available in the font, then the unmapped glyph would be displayed.

The system would thus be capable of providing three types of display, namely as follows.

1. Localized text.

2. Glyphs of the LOCALIZABLE SENTENCE MARKUP BUBBLE characters.

3. Unmapped glyphs of those LOCALIZABLE SENTENCE MARKUP BUBBLE character sequences for which there were an unmapped glyph both defined in a standardization document and available in the font, together with glyphs of the LOCALIZABLE SENTENCE MARKUP BUBBLE characters for other LOCALIZABLE SENTENCE MARKUP BUBBLE character sequences.

As examples of each of the three possibilities, for the example sentence from earlier, localized into English.

The enquirer is the brother of the first person that was named.

It happens that there is a glyph for that sentence as it is one of the sentences for which I had produced a glyph at an earlier stage of the research when I was, at that time, thinking in terms of one character code point for each sentence, which is no longer the case.

An important part of this proposal is to request action by the committees to set up an encoding infrastructure for the encoding in a separate standardization document of complete localizable sentence markup bubbles using those twelve characters.

In my research, for research purposes, I have produced some sentence and symbol pairs and encoded them in a Private Use Area.

I authored most of the sentences and phrases myself.. One phrase was suggested by a gentleman who kindly provided translations of some of the initial list of sentences into Swedish. That phrase is as follows.

Best regards,

I saw one of the phrases on a sign in a Google street view image of the foyer of the Museum of Modern Art in New York. It is as follows.

Thank you for visiting

I have designed symbols to represent the various sentences and phrases and have produced various versions of a font so that I could use the symbols and their code points in experiments.

Yet the list is incomplete. For example, for the simulations of the invention being used in the seeking and the providing of information about relatives and friends after a disaster, the only outcome encoded is that the person is safe. For a real world system, other sentences, such as for the person being injured and for the person being in hospital and for other possibilities would need to be encoded. I have chosen to leave what sentences to author and encode for the decision of experts in disaster management at a later time if the invention is ever implemented. I have sought to add sentences to explore the possibilities.

Also, there is a sentence as follows.

Where is a pharmacy please?

Other sentences, such as for seeking directions to a railway station and so on would also be desirable if the invention is ever implemented.

The system also includes the following three sentences, which show how detailed requests can be included in the system.

Where can I buy a meal with no gluten-containing ingredients in it please?

Where can I buy a vegetarian meal with no gluten-containing ingredients in it please?

Where can I buy a vegan meal with no gluten-containing ingredients in it please?

Such sentences are important to the people choosing and needing various diets. Other sentences about diets would also be needed if the invention is ever implemented. Such

sentences might only ever be used by a small proportion of the population and many people might never encounter them and some people, such as people working in hotels, restaurants and cafés might only encounter them rarely. However, as localization would, in most cases if the invention is widely used, be automatic, the situation of someone encountering such a sentence only rarely should not be a problem.

As mentioned earlier, I saw one of the phrases on a sign in a Google street view image of the foyer of the Museum of Modern Art in New York. It is as follows.

Thank you for visiting

The sign has localizations in at least four languages. I thought of the idea that if a passive radio frequency identification device tag were fitted, perhaps out of sight, in the lower left corner of such a sign, then a mobile device that has a facility to read from a passive radio frequency identification device could read a character code sequence for a phrase from the passive radio frequency identification device and could then use that character code sequence to localize the phrase that is displayed upon the sign.

The final page of this document shows an indication of the way that a character code sequence would be analysed by software in an automated localization facility.

Computing an index value to a database from a sequence.

The characters can be regarded as having the following effects.

LOCALIZABLE SENTENCE MARKUP BUBBLE OPENING BRACKET

```
temporary:=0;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE CLOSING BRACKET

```
index_value:=temporary; get_localized_sentence(index_value);
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT ZERO

```
temporary:=10*temporary;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT ONE

```
temporary:=10*temporary + 1;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT TWO

```
temporary:=10*temporary + 2;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT THREE

```
temporary:=10*temporary + 3;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT FOUR

```
temporary:=10*temporary + 4;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT FIVE

```
temporary:=10*temporary + 5;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT SIX

```
temporary:=10*temporary + 6;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT SEVEN

```
temporary:=10*temporary + 7;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT EIGHT

```
temporary:=10*temporary + 8;
```

LOCALIZABLE SENTENCE MARKUP BUBBLE DIGIT NINE

```
temporary:=10*temporary + 9;
```