

2007-01-08

Draft contribution by Erkki I. Kolehmainen

Design principles for extended local keyboard layouts for use in the multilingual Europe

Most of the keyboard layouts that are presently in common use in Europe have been designed at the time when the number of different characters that could be entered was severely limited by the then prevalent character encoding schemes; in addition, some limitations of the original mechanical typewriters have been carried over into their design. These keyboards have been designed for each local language environment, although often with some capability to enter “foreign” letters.

The expanding European single market, however, with free movement of goods and people, has drastically changed the environment. Increasingly, the names of people, places, and companies and their products should be communicated in their proper form, for both politeness (often also as a personal choice in appreciation of cultural diversity, not just for business reasons) and accuracy, possibly also for legal reasons. The International Commission on Civil Status has signed already on 13 September 1973 their Convention No. 14 “On the recording of surnames and forenames in civil status registers” (see: <http://perso.orange.fr/ciec-sg/Conventions/Conv14Angl.pdf>).

The advent of the Universal Character Set (ISO/IEC 10646 and Unicode) has fortunately removed the limitations imposed by the ICT systems on the number of different characters to be handled, i.e., processed and rendered, at once. Consequently, one could assume that the ICCS convention will eventually be enforced; within the European Union the current plans are for the year 2009. Some countries are already committed to using the proper names in their formally recognized minority languages as a result of the European Charter on Regional or Minority Languages (see: <http://conventions.coe.int/treaty/en/Treaties/Html/148.htm>).

What remains to be done is to provide for an easy, intuitively comprehensible way to enter the data. However, there is no real possibility to start from scratch, or switch the base to formal international standards, such as ISO/IEC 9995, which have had little if any influence on the design of the present keyboard layouts in many countries.

Common background elements

Most users are familiar with one keyboard layout only, and they wouldn't welcome any disturbing changes to it. Thus, any expansion should be carefully based on their current layouts.

Many European users of other than the UK keyboards are familiar with the dead key method for keying “foreign” letters with diacritics. In this method, the key for the required diacritical mark is pressed first, followed by the base letter with which the diacritic is to be combined. This method has been in wide use since the early mechanical typewriters, and there is considerable benefit in continuing its use and expanding it to cover essentially all letters with single diacritics. However, most users consider as base characters those letters of their own alphabet that have diacritics, and thus cannot be asked to utilize any other method than a single keystroke to enter them. Such letters, e.g., å, ä and ö in Swedish, ñ in Spanish, etc. will thus have to continue to appear directly on their respective keyboards, although they will be treated as combinations on most other keyboards.

Although one could envisage the dead letter key method to be extended to entering letters with multiple diacritics, this is not generally supported by the current table driven methods that the industry employs to create keyboard drivers (e.g., MSKLC by Microsoft). Since several of the characters with multiple diacritics (e.g., many of the Lithuanian accented characters) cannot be represented as pre-composed characters, it would appear feasible to enter practically all the characters with multiple diacritics as decomposed characters, i.e., the base character followed by the combining diacritical marks.

It would also make sense to implement the two modes of operation utilizing the same keyboard layout as technically two separate keyboard drivers. This would allow for a highly flexible solution, where the majority of the users who don't need the extended capabilities provided by the decomposed mode would not even have to know of the facility. In fact, the underlying principle for all of the design approach described herein is not to impose any disturbing change to those users who are content with the current level of implementations. Those who ever need to utilize the decomposed mode, can either easily switch between the two keyboard drivers as appropriate or also decide to produce all characters in this mode only. It should be noted that for ease of processing, a facility to normalize the decomposed input characters will be needed.

The dead key method should also be extended to letters with diacritic-like additions, specifically to characters with stroke (although they are not decomposable), in order to avoid having to assign key positions to each of the various letters, such as the Polish L WITH STROKE [ł] or the Sámi T WITH STROKE [t̥]. In addition to keeping key positions available for other use, this will also meet the criteria of the method and the positioning of the characters being intuitively comprehensible.

For ease of use, the diacritical marks should be engraved on the key tops, since they can hardly be placed following any intuitively recognizable scheme. If properly placed, the other additional characters don't necessarily need to be engraved.

Sample design and implementation

The details pertaining to the design principles for what is described above, especially those on the decomposed mode, are the result of the discussions at the BOF session on the Design Principles for a Regional, Multilingual Keyboard at the 29th Internationalization and Unicode Conference in San Francisco on 6 March 2006 (see: http://www.unicodeconference.org/iuc29/program-d.htm#BOF_3).

A sample specification (for particularly Finnish and Swedish and the recognized minority languages in the Nordic countries plus the official languages of the European Union using the Latin script) is available at <http://kotoistus.fi/nappaimisto.htm>. In addition, this layout also provides for specific punctuation used in a number of European languages. At the same address there are links for downloading currently three different evaluation or trial versions implementing the basic mode of operation, two for the Windows environment (of which one is a formal evaluation version by Microsoft) and one for the Linux environment.

It should be noted that the requirement to cover the characters of the local minority languages, which these Finnish specifications respond to, is of course dependent on the region. Also, these specifications are not "pure" in the sense that certain compromises have been made in order to facilitate an agreement on common engravings with the Swedish keyboard layout that has been designed for governmental use with a high level of additional character assignments to the various key positions (most of which, however, will not be engraved).

End.