

Bibliography in footnotes

Note: *This text demonstrates a form of bibliography: footnotes for the immediate citations and a collection of the annotated references at the end:*

If you are interested in the subject of this text - National Language Support - see www.daube.ch/share/nla00.html

Aim of this paper

It was clearly stated in ¹⁾ that an architected approach solving the many problems in the area of National Language Support is required, rather than a piecemeal solution. This paper presents SHARE Europe's view of the areas that must be covered by a National Language Architecture:

- Language (the natural language spoken and written)
- Country customs (culture-sensitive functions)
- Data encoding
- Presentation services

So one Sunday in mid 1988 Klaus Daube sat down to lay out this paper, which he also presented to other user IBM groups ²⁾. And having awakened the sleeping lion he became the chairman of the newly established Special Technical Working Group.

It was the time of the great more or less peaceful revolution in Eastern Europe, when Klaus was editing the paper. When the Berlin wall was torn down, Klaus expected the implementation of a National Language Architecture to be relatively easy compared to these developments. The next step of editing coincided with the turnover in Rumania. Klaus remembered his statement years ago, that Europe not only needs the characters for the Western languages, but also - and in the same environment - those of the Eastern European languages.

Identification of Language

The identification of the language must be universal. Text files are not only processed by special applications like formatters or text processing systems, but also with very general purpose instruments like program editors. Hence the human reader of such a file must be able to recognize the specification. This could be achieved by a notation using only «printable characters» (see syntactic character set) as shown in the following example from a formatter input specification ³⁾.

```
.assign LANGUAGE, SWGERMAN
```

to specify the German language as used in Switzerland. The identification must include both the language itself and the country.

- 1 Gardner-1: Gardner Peter (ed.), *SEAS National Character Task Force: White Paper on national character, language and keyboard problems*. Geneva, SHARE European Association, September 1985.
- 2 Daube-2: Daube, Klaus, 'Text and Code - A Dragons Pond.' *Proceedings of the 30. G.U.I.D.E. conference in Basel*. G.U.I.D.E. Headquarters, 1989.
- 3 Daube-3: Daube, Klaus, *National Language Support in SUSI*. OBRZ AG, Handbuch 410 (Dokumentverarbeitung mit SUSI), Kapitel 410.20.35. Zürich 1990.

Sort rule

Both predefined (e.g. Swedish, German, Swiss-German) and installation-specified order must be possible. See also Hierarchy of defaults.

Please note the subtle rules for French, where accents are not taken into account except for homographs like the the above example in which discrimination starts from the last character and moves left (to the first character) with priority depending on accents. Each language uses different rules.⁴⁾

A key construction algorithm (like that of Alain LaBonté⁵⁾) is not the same thing as a sort algorithm (like those found in⁶⁾). The merit of the key construction method is that it computes a single key (constructed from parts), which allow for single pass sorting, not multiple passes as used by others.

Compiler and Runtime System

Although the character set and code issue is often not specified in the standards for a programming language⁷⁾ certain criteria must be met by a modern compiler:

- The vocabulary of the programming language (syntactic elements, key words, reserved words) are independent of translations into national languages.
- Source code must be allowed to use all code pages available on that system. Compiler directives must make use of the NLA functions needed to support different code pages.
- A compiler must not use an arbitrary otherwise undefined code page).
- Comments must be allowed to use all national characters available in the actual code page.
- String constants must be allowed to use all national characters available in the actual code page. The string constant must be tagged in the same way as requested for files.
- Labels must both allow lengthy names and use of all national alphabet characters available in the actual code page
- Input parameters must be specified in the same order on the call statement.
- Output parameters, where applicable, must be in the same order on the call statement.

Both the compiler and the runtime system must make use of the same base functions.

It must be possible to use the NLA to write a program with French comments and Turkish string constants.

4 Wingen-2: Wingen, Johan W. van, *Coded Character Sets and Programming Languages*. ISO/IEC JTC1/SC2 N1961R and ISO/IEC JTC1/SC22 N587R, Revised April, 1989.

5 LaBonté-3: LaBonté, Alain, 'A New Data Type for National Language?'. *Proceedings of SEAS Anniversary Meeting 1989*, vol II, p 1519 - 1523. Geneva, SHARE Europe (SEAS), 1989.

6 Knuth: Knuth, Donald E., *The Art of Computer Programming, Volume 3 / Sorting and Searching*. Addison-Wesley Publishing Company, Reading, Mass. 1973. ISBN 0-201-03803-X.

7 Wingen-1: Wingen, Johan W. van, *Coded Character Sets and Programming Languages*. ISO/IEC JTC1/SC2 N1961R and ISO/IEC JTC1/SC22 N587R, Revised April, 1989.

Annotated bibliography

- Abderhalden Abderhalden Ulrich, *National Language Support Cookbooks*. IBM Switzerland - Field Support centre, Zürich 1988.
Hints and tips for conversion of applications and data to the international code page 500. 327x terminal equipment, GDDM, ISPF/PDF, compilers, QMF, SDSF, DFSORT, TSO/E. List of machine readable material.
- Arielli d'Arielli, L., *A message management system for personal computers*. IBM Systems Technical Journal, vol 28, no 3, 1989.
A method using message skeletons is described.
- Bouch Bouch Debb, 'Adding National Language Support to Applications'. *Proceedings of SEAS Anniversary Meeting 1987*, vol II, p 995-1015. Geneva, SHARE European Association, 1987.
Summary of issues involved in NLS explained by the work done on a text processor running under VM/SP5.
- Daube-2 Daube, Klaus, 'Text and Code - A Dragons Pond.' *Proceedings of the 30. G.U.I.D.E. conference in Basel*. G.U.I.D.E. Headquarters, 1989.
Comparing the human habits of gesture and miming with codes in data processing shows a severe gap in understanding coding mechanisms. This is the source of many problems related to national language support in data processing applications.
This text also is available in French, kindly translated by Ministère des Communications du Québec, St. Foy, Canada.
- Daube-3 Daube, Klaus, *National Language Support in SUSI*. OBRZ AG, Handbuch 410 (Dokumentverarbeitung mit SUSI), Kapitel 410.20.35. Zürich 1990.
The text formatter SUSI supports code page switching (also within a file. A piece of text also may have the attribute language, to which hyphenation is bound. National keyboards can be used. The code page used internally is based on code page 037, the character set supports most Western and Eastern European languages.
- Gardner-1 Gardner Peter (ed.), *SEAS National Character Task Force: White Paper on national character, language and keyboard problems*. Geneva, SHARE European Association, September 1985.
Collection of the problems encountered in Europe with IBM hardware and software for processing textual data.
- Hart Hart, Edwin (ed.), *ASCII and EBCDIC Character Set and Code Issues in Systems Application Architecture*, Chicago, SHARE Inc., SSD #366, 1989.
This paper states that IBM can not create a consistent SAA on top of the existing character set and code inconsistencies. Although the paper describes the North American symptoms of the problem, the requirements are for the international problems.
- Knuth Knuth, Donald E., *The Art of Computer Programming, Volume 3 / Sorting and Searching*. Addison-Wesley Publishing Company, Reading, Mass. 1973. ISBN 0-201-03803-X.

-
- LaBonté-3 LaBonté, Alain, 'A New Data Type for National Language?'. *Proceedings of SEAS Anniversary Meeting 1989*, vol II, p 1519 - 1523. Geneva, SHARE Europe (SEAS), 1989.
- When comparing textual data with numeric data similarities can be seen in the necessary structure of the data. Sign, exponent and mantissa can be mirrored to base characters, accents, case and special symbols. A reduction technique is presented to lower storage requirements of this new structuring without the need to expand data for processing.
- Wingen-1 Wingen, Johan W. van, *Coded Character Sets and Programming Languages*. ISO/IEC JTC1/SC2 N1961R and ISO/IEC JTC1/SC22 N587R, Revised April, 1989.
- Summarizing the requirements for compilers for support of national language: character set, symbol classes, comments, identifiers, string-constants.
- Wingen-2 Wingen, Johan W. van, *Coded Character Sets and Programming Languages*. ISO/IEC JTC1/SC2 N1961R and ISO/IEC JTC1/SC22 N587R, Revised April, 1989.
- Summarizing the requirements for compilers for support of national language: character set, symbol classes, comments, identifiers, string-constants