

Unicode – The Unknown World Standard

Prof. Johannes Bergerhausen, University of Applied Sciences, Mainz, Germany

Panel “Unicode and Academia”, moderator: Deborah Anderson, UC Berkeley, California, USA

29th Internationalization and Unicode Conference, IUC29, Burlingame, California, USA

Good morning! First of all I would like to thank Deborah Anderson of the University of California, Berkeley, for the invitation to take part in this panel.

Entitling my contribution “Unicode – the unknown world standard”, I surely exaggerate. Unicode has long been established and is known to a great many people all over the world. Outside the specialist community, however, hardly anyone has any idea about what Unicode actually is. But I can see a few good reasons why this subject might be of interest to a much larger audience.

1. Unicode – a cultural achievement

First of all, Unicode represents, of course, an important technological achievement. Especially on this conference there is a lot of talk about terms like bi-directional behaviour, collation tables, case mapping and so on. But the characters and scripts we are talking about together form a part of the key achievements of a culture. This makes Unicode and its future completion with all writing systems of the world a great cultural achievement of mankind. In 1984, Joseph D. Becker had a vision: “Nevertheless, the ultimate goal must be multilingual word processing”¹. Who would have thought that less than 20 years later his utopian dream was to become reality for a growing number of people? Also, the idea to include all characters of the world in one single encoding standard to me seems a deeply democratic thing to do: Basically, all characters are created equal. There are no competing encoding standards and thus, on this level for once, there is no war between cultures.

1.1 We shape tools and they in turn shape us

As early as in 1964, Marshall McLuhan with this phrase² pointed out that every new technology also exerts a great influence on culture. There is no denying that Gutenberg’s invention and television have deeply affected our culture and continue to do so. In this sense, the ASCII code, too, has changed the world. Without a common code, inventions like the Internet or e-mail would have been unthinkable of. Some of the cultural effects Unicode has had we already take for granted today, like the localization of cell phones or the subtitling of a customary DVD in up to 32 languages. But world-wide access to the characters of the world will most certainly change culture in ways we are unable to foresee today. Likewise, the influence of Unicode on typography has only just begun to take effect. Any home PC user today has more characters at his or her disposal than did the most advanced multilingual punch-cutters and printers of the Renaissance. Besides, all today’s scientists work with the universal tool of the computer. But they have only just begun to unearth the treasure of characters Unicode holds for them.

1.2 The cultural opportunity

These reasons offer Unicode the opportunity – and the task – to complete the repertory of characters of mankind. Where the commercial interest of some enterprises dwindles, academic projects can make substantial contributions to further this cause. Unfortunately, neither politics nor the IT community are as yet much aware of this dimension. A keyboard layout for a minority language, besides being a technical achievement, is an immensely important cultural contribution! I fail to understand why many computer firms have yet to realize the potential this dimension holds by way of sponsorships. If we regard Unicode as a kind of typographic United Nations General Assembly where every culture ought to be entitled to a seat we may be needing something like a typographic UNESCO.

1.3 Communicate more!

From my working experience as a communication designer I can only recommend the Unicode Consortium to communicate beyond just specialist circles. Unicode has yet to gain the status this incredible project might actually have in the international community. What I would like to briefly present in the second and final part of my contribution is the very small part we take in communicating Unicode.

2 The project decodeunicode

Motivated by above mentioned reasons, the project decodeunicode was launched in 2004 at the University of Applied Sciences Mainz. In my keynote to IUC27 in Berlin, I presented it in detail. I would now like to briefly outline the concept of the project.

All those who are sitting here and many, many more have done a wonderful job encoding all these characters. Laypersons and the interested public community, however, have had no comprehensible access to their meaning. In a way, all these characters now need to be decoded again.

More detailed information about our concept can, among others, be found in the English version of our project brochure³.

2.1 The BMP-poster

To get an initial overview we started by designing and having printed a rather sizeable poster with all BMP characters listed in Unicode 4.0. With all 300 poster copies already sold out, we are planning to present a new version this summer which will cover the content of Unicode 5.0.

2.2 The website www.decodeunicode.org

The Graphical User Interface of the database on all BMP characters has a clearly laid out structure. The top area in white includes Unicode data like the hexadecimal values and properties along with a reference glyph of each character in GIF format. The bottom part of the site is a wiki. This is where information about every single character can be collected. Like Ken Whistler said: "Every character has a story." Basically, we are looking for any kind of information about each character, be it linguistic, semiotic, typographic or from other disciplines such as mathematics. A small team of student administrators is in charge of and, where necessary, modifies any new entry. Ten months after going online, our database today comprises over 3,000 entries on more than 1,200 characters. The website is well received: Our statistics register a daily average of over 700 visitors. In December, we had visitors from as many as 105 countries of the world.

2.3 Book design in graphic design programme

In my stage II seminar course in book design at the University of Applied Sciences Mainz I asked the graphic design students of three consecutive semesters to look at the phenomenon Unicode through the eyes of the designer. In the course of one semester, respectively, every student worked on one chosen script or group of characters. At the end, each presented a thoroughly designed book. Some examples I would like to show you.

2.4 Interdisciplinary work

Unicode is an interface where many disciplines meet. From what I have seen, these disciplines are as yet taking too much care of themselves alone. With the universal tool of the computer, however, which is being used in all sciences without exception, a new universality of the sciences may dawn – provided we dare look beyond the end of our noses.

To us, the current state of work at decodeunicode is a foundation we can base our future work on. Our small team would gladly welcome further interdisciplinary contacts. Much remains to be done.

Thank you very much.

Cologne, Germany, January 7, 2006

English translation by Nicola Fischer, Heidelberg, Germany

¹ Joseph D. Becker: Multilingual Word Processing, Scientific American No. 251, July 1984, page 96-107, re-translated into English from a German translation of unknown origin.

² Marshall McLuhan, Understanding Media – The Extensions of Man, New York, USA, McGraw Hill, 1964

³ decodeunicode Information brochure, English Version: ISBN 3-936723-09-5, German Version: ISBN 3-936723-08-7, Fachhochschule Mainz, Mainz, Germany, 2004