Abstracts—Multilingual document processing

dvipdfmx, an eXtension of dvipdfm

Jin-Hwan Cho Korea Institute for Advanced Study chofchof@ktug.or.kr

In this presentation I would like to introduce a DVI to PDF translator, dvipdfmx (formerly dvipdfm-cjk), which is an extension of dvipdfm developed by Mark A. Wicks.

One might ask why we consider a DVI to PDF translator at this time, since we already have the powerful TEX software pdfTEX, which generates PDF results directly from TEX sources without using the DVI format. It is true for people using languages which make use of the Latin alphabet (or other 8-bit character set) that pdfTEX is usually sufficient.

However, the situation is quite different for those who use Northeast Asian languages (Chinese, Japanese and Korean; simply CJK) or Unicode using 16-bit characters. The current version of pdfTEX has no ability to handle 16-bit characters. Even though a PDF viewer shows 16-bit characters in a PDF file generated by pdfTEX, the codes are not 16-bit but 8-bit. Thus, extracting and searching those 16-bit characters is impossible. Furthermore, it is hard to generate a PDF file with pdfTEX having bookmarks or text annotations with 16-bit characters.

That is the main reason why I am introducing dvipdfmx. The DVI driver software, dvipdfmx, handles 16-bit character using CID-keyed font technology which is already included in the PDF specification. Therefore, dvipdfmx works well with almost all TEX variants including ASCII pTEX, the most popular TEX software in Japan, and Omega. In particular, it is interesting to see a PDF example containing 16-bit characters from dozens of different languages, which are extractable and searchable as a matter of course.

Recently there was revolutionary progress in developing dvipdfmx, namely when dvipdfmx began to support ConTeXt. Much of dvipdfmx was rewritten at this point. At present, dvipdfmx handles many ConTeXt documents containing complex MetaPost figures (color shading too) and interactive forms (JavaScript too). I would like to show those fantastic examples in the presentation.

There are also many features in dvipdfmx not mentioned above, PDF encryption for example. More information on dvipdfmx can be found at the project home page, http://project.ktug.or.kr/dvipdfmx. The dvipdfmx project is a combination of the dvipdfm-jpn project by Shunsaku Hirata and its modified version, dvipdfm-kor, by Jin-Hwan Cho.