Hàn Thế Thành

Hàn Thế Thành is the creator of and still maintains pdfTeX.

[Interview completed 24 July 2008.]

Dave Walden, interviewer: Please tell me a bit about your personal history independent of TeX.

Hàn Thế Thành, interviewee: I was born in Vietnam in 1972 and lived there until 1990, when I got a chance to go study in the Czech Republic (at that time still named Czechoslovakia). I studied at the Masaryk University in Brno from 1991 to 2001 and got a Master's degree and later PhD degree in Computer Science. Then I went back to Vietnam and worked at the University of Pedagogy in Ho Chi Minh City (also known as Saigon). Since 2006 I have been living with my wife in Bielefeld, Germany, where my wife is studying.

DW: What was your job at the University of Pedagogy?

HTT: I was teaching introductory programming and also working as a network administrator.

DW: Please tell me when and how you first got involved with TeX.

HTT: During the first years at the university, I heard from time to time from my schoolmates that TeX is an amazing typesetting system, very powerful also but very difficult to use. But I didn't use TeX myself at all until I had to choose the subject for my Master's thesis in the fourth year. There were a number of subjects to choose among, and I picked one that sounded like "Automated typesetting systems" or something similar. The idea of my supervisor was for me to rewrite TeX using a high-level language. Later it became clear that rewriting TeX in a high-level language was a too difficult task for a student like me, so my thesis supervisor changed my topic to "TeX typesetting system and the Portable Document Format". The intention was more or less what pdfTeX does today: to change TeX so that it can produce pdf directly. So, I first had to learn about TeX — to get really involved with TeX. Before that I only heard about TeX but never used it. That was in 1994.

DW: One can argue that a key reason TeX remains so vibrant today is the existence of pdfTeX. From what you just said it sounds like you sort of stumbled into creating pdfTeX which everyone uses today — that your thesis supervisor pushed you in this direction more than that you had an initial deep desire to work in this area. Is that correct?

HTT: Yes, pdfTeX started more or less like that: my supervisor, Professor Jiří Zlatuška, used to be a very active TeX user and developer. Jiří is also a fan of logical programming.

So his original intention was for me to rewrite TeX using a declarative language like Prolog and use that for further development. I had little (if any) clue what all this meant. I picked the subject I did simply because: (1) I liked logical programming; (2) from what I had heard about TeX, the subject sounded interesting; and (3) I didn't find a more interesting subject to choose from the available topics.

After a few months of playing with rewriting TeX, it was clear to me and also to Jiří that rewriting TeX in Prolog was too difficult for me. Suddenly one day Jiří called me to his office, gave me the printed PDF specification version 1.0, and said that I might try to change TeX to produce PDF output directly (later I learnt that Jiří got that idea from some discussion with Phil Taylor and Knuth at Stanford University). I was keen about the idea and the original plan also seemed unrealistic; so we changed our plan. I started to read the PDF specification and to learn how to hack TeX with the Knuthian web system, Web2c, Kpathsea and friends. After a few months I made a "Hello, world!" PDF from TeX, and it was rather an exciting moment for us. But I didn't expect pdfTeX to be as widely used as it is today (I think Jiří didn't expect that either, but I might be wrong here).

I learnt TeX "the hard way": I started with reading *The TeXbook* since that's what Jiří gave me in the beginning. Then I started using plain TeX since I had heard that LaTeX was not as good as plain if you wanted to learn the details of TeX, to control every part of typesetting, and so on. So I used plain TeX to typeset a periodical, thesis work of my friends, and other occasional materials. But later I started using LaTeX, since doing everything in plain is rather painful. So I use LaTeX mostly, and use plain TeX only for a few very specific applications that would be better done in plain TeX.

Learning the PDF format was not very hard, since the specification of PDF version 1.0 was a very thin book (I would have given up immediately if I got, for example, version 1.3 or later). But learning web change files, Web2C and friends was rather hard for me: too many steps were involved, and when something goes wrong, it is not easy to find out where the mistake is.

Jiří wanted me to follow the literate programming paradigm and have everything done via the change file mechanism. However later it became more and more difficult to maintain things this way, so I decided to move certain things to C. The main criteria in deciding what to keep in Pascal WEB or and what to do in C is this: if it is backend-related, it should be in C; otherwise it should be in WEB. Jiří was not happy with my decision, but more or less accepted it (or at least let me do it).

DW: Did you finish an operating version of pdfTeX as your Master's thesis, or did you somehow finish your thesis and then keep working until an operational version of pdfTeX was available?

HTT: I cannot recall when pdfTeX became really "functional for use", since the development was gradual with contributions by many people in various areas. When I finished my Master's thesis, the state of pdfTeX was about as stated in Petr Sojka's article: support for embedded Type1 fonts, virtual fonts, hyperlinks, LZW compression (later LZW compression was replaced by zip compression). There was no image inclusion yet!

DW: Are you referring to the article entitled "The Joy of tex2pdf — Acrobatics with an Alternative to DVI Format" by Sojka, Jiří, and you, in *TUGboat* 17:3 (1996)?

HTT: Yes.

DW: How did the greater TeX world become aware of your pdfTeX system and come to include it in all the TeX distributions?

HTT: I got connected to other people in the TeX world for the first time when Jiří

corresponded with Sebastian Rahtz about pdfTEX (at that time still 'tex2pdf'). Sebastian was interested and started playing with pdfTEX, supporting it in various ways: he set up the pdfTEX mailing list, compiled and tested it on other platforms, introduced it to other users, etc. Sebastian gave a vital push to pdfTEX development in those early days.

Later pdfTEX was seen by Knuth during his visit to Masaryk University and it received positive comments by Knuth, which was very encouraging for us (me and Jiří). The first article about pdfTEX was the one I just mentioned, by Petr Sojka. The pdfTEX mailing list was an extremely useful place to discuss pdfTEX development in the beginning. And one day Hans Hagen showed up on that list and started experimenting with pdfTEX, reporting problems, discussing new ideas and features, etc., which was another great impact on pdfTEX. The fact that such well known and active members of the TEX community (Sebastian, Hans, etc.) liked pdfTEX was the key to pdfTEX becoming more "known". Then it got included in teTEX by Thomas Esser. Once something is in teTEX, usually it will be accepted consequently by other TEX systems.

DW: Please clarify for me the distinction, if any, between pdfTEX and the microtypographic extensions to TEX described in your PhD thesis that was reproduced in a special issue of *TUGboat* (volume 21, number 4), "Microtypographic extensions to the TEX typesetting system".

HTT: The goal of my Master's thesis was to make PDF output directly from TEX. When I started my PhD study (also under Jiří), we only knew I would do something related to typesetting, but we did not know exactly what. During the first year or so of my PhD studies, I was still developing pdfTEX and also looking for an idea for the PhD thesis. I came up with a few, then Jiří told me to stick with the micro-typographic extensions, which was again a very wise decision in my opinion.

DW: Please clarify for me the timing of your Master's studies and your PhD studies. You said you were at Masaryk University from 1991 to 2001. When did your Master's degree finish and your PhD studies begin?

HTT: I finished my Master's studies in summer 1996. A few months afterward I started my PhD studies.

DW: Please elaborate a bit more on your approach to learning about micro-typography, to making the necessary extensions to pdfTEX, and the research component necessary for a successful PhD thesis.

HTT: I don't feel very qualified to talk about how to do a successful PhD thesis, since I was struggling with mine to get it done at all.

DW: Sorry; I didn't mean to remind you of a stressful time. Mainly I am interested in how you learned what you needed to know about micro-typography?

HTT: I cannot recall exactly how I learned about micro-typography — it was a gradual process, as for most people, I suppose. I started by reading some books and articles, and then I searched for more relevant resources. The most useful resources I can remember were the paper by Hermann Zapf "About microtypography and the hz-program" and the brochure about the hz-program by URW, the German type foundry. I also experimented a lot with Adobe InDesign, which claims to have some modules from the hz-program integrated. It's interesting to see that some ideas of the hz-program were inspired by TEX itself originally.

DW: I know that pdfTEX was quite operational by the time you finished your PhD thesis.

Did you keep developing pdfTEX after you returned to Vietnam?

HTT: When I returned to Vietnam, in the beginning I had a long break in pdfTEX development due to difficulty with network access and various things. Then occasionally I found time to make small extensions to pdfTEX, but it was no longer active development as before. Of course other people have been contributing to pdfTEX, too. The most significant contributions to pdfTEX in the recent years were done by a very quiet person named Hartmut Henkel. His patches greatly improved pdfTEX in many aspects: speed, stability, cleaner code, better functionality, etc.

DW: More recently you seem to have gotten more involved in developing pdfTEX again.

HTT: Yes, I have been more involved with pdfTEX since I moved to Germany with my wife. In Germany I work at home as a consultant for River Valley Technologies — Kaveh and Radhakrishnan's company. I support network administration, automating some editing tasks, and also pdfTEX deployment.

DW: I gather that pdfTEX development works in some fashion as an "open source" development effort. And you told me in a message last week that you had to take a few days away from our interview because Karl Berry wanted you to fix something about pdfTEX immediately for the upcoming TEX Live release. Please tell me about the on-going organization and coordination of pdfTEX development.

HTT: pdfTEX development has evolved over time, and presently works more or less like this. There is a project page for pdfTEX hosted at `sarovar.org` where people submit bug reports, feature requests or patches. There is also a mailing list for people interested in pdfTEX development. And there is a core team (Hans Hagen, Taco Hoekwater, Hartmut Henkel, Martin Schröder and me) where we discuss the decisions made on pdfTEX.

DW: I understand that you have spent (a lot of) effort adding Vietnamese support to a number of fonts. How did you get involved with this and how did you go about it?

HTT: As I was learning TEX, I was interested in using it for Vietnamese too. At that time there was a package called `vcmr` by Werner Lemberg, which already provides quite good support for typesetting Vietnamese. However, I was not happy with the shapes of Vietnamese letters, so I decided to add Vietnamese letters to the CM fonts by myself. It was only a hobby activity, and I didn't have any artistic background. I learnt mostly by looking at existing fonts and reading materials that I could find, as well as from comments I received from experienced people. I added the Vietnamese letters to CM fonts using Metafont. To convert those fonts to Type1 format, I used a combination of several tools: Metafog, FMP (by Y&Y), a2ac, and some of my own Perl scripts. To add Vietnamese letters to existing Type 1 fonts, I used more or less the same route, although I drew the accents using FontLab.

There are more people involved in `vntex`: Werner Lemberg and Vladimir Volovich for LATEX support, and Reinhard Kotucha for testing/maintaining the package and making everything neatly conform to TDS and providing what is required by TEX Live and CTAN such as having a `README` file, copyright notices, etc. There is no active development on `vntex` anymore, since `vntex` has quite a large number of fonts already. There is even a Vietnamese translation of the math font survey for TEX by Stephen Hartke, which means that most of the text fonts mentioned in the survey have a Vietnamese version, too.

DW: Thank you, Thành, for taking the time to participate in this interview. It has been an honor for me to communicate with someone who has had such a major impact on the continuing use of TEX.