The ℵ (Aleph) project

Keywords
Omega, Aleph, eTeX, eOmega

Abstract
A brief introduction to the ℵ project, a TEX extension providing most of Ω and ε-TEX features.

The path

TEX was created by Donald Ervin Knuth more than 20 years ago. At the time when it was initially developed, it was supposed to serve mainly one purpose (provide a high–quality typesetting workbench for Knuth's books), but was general–purpose and powerful enough to be quickly adopted as a more or less standard environment in the scientific community, thanks to its capability to easily typeset complex formulas.

Usage of TEX outside the scientific/technical domain has always been confined to niche applications, partly due to the absence of high–level formats like LATEX, but more geared towards nontechnical writing, and partly because TEX, in its original design, has very limited support for languages other than English.

Efforts to push the limits of TEX have been made, in at least three different directions, by different teams. This led to the creation of multiple, sometimes incompatible extensions of the original engine; we have for example

- pdfTEX, which gives TEX the capability to produce output directly in PDF form, and introduces micro–typesetting capabilities;
- ε-TEX, based on previous extensions of TEX which added right–to–left typesetting capabilities to TEX, which strives to break some of the structural limitations of TEX while maintaining maximum compatibility;
- Ω (Omega), an effort to bring the TEX world to up–to–date standards and push it towards a multicultural world.

As it was mentioned before, not all these extensions are compatible with each other; for example, while pdfTEX and ε-TEX can be merged in a single program, the changes in Ω are so extensive that they put the program in a rather isolated position.

While hopes and desires for a unified TEX extension have always been present, they have not been pressing because the most common format (LATEX) made no use of ε-TEX extensions, and other formats that did take advantage of those extensions (like ConTEXt) didn't have enough of a market to be of interest to Ω users. Things started to change recently, as the LATEX team found the original TEX more and more restrictive, and ConTEXt started spreading, its power and flexibility appealing to users outside the domain of latin scripts.

Birth of a new branch

The Ω project started with the best of intentions and reached some outstanding results; for example:

- 16–bit registers allow Ω to typeset documents too complex to be handled by TEX (or even ε-TEX);
- Ω introduces the concept of ΩTP (Ω Translation Process), a way to transcode texts, therefore allowing the input of text in any script in any encoding, in a font independent way and without the use of active characters; for example, one can write in Greek or Arabic using a plain English keyboard: ΩTP takes care of translating Latin characters (or sequences thereof) into the appropriate Arabic or Greek characters
- Ω can typeset text in many directions; ε-TEX provided some support for right–to–left typesetting, but Ω brings this capability to any combination of direction (left to right, right to left, top to bottom, bottom to top), easily combining these layouts in the same page.

But Ω presents some characteristics that can make its adoption a difficult choice with an uncertain future. These are the characteristics of an experimental project with a very broad (maybe too broad) final destination, a moving target where stability is only a secondary if not even tertiary target. While this experimental nature of the project is not intrisically negative (on the contrary, it guarantees that the project has enough dynamism to project itself into the future; in fact, it has been what brought Ω to its current status) it does hinder a widespread adoption of the tool for production use.

We therefore have a program, Ω, that has a lot of potential, and that needs to continue to develop; but this potential has to be somewhat harnessed to make the results available in production use contexts. Therefore, we chose to take the Ω code base and use it to start a new project, called ℵ (Aleph),[1] which could provide the power of Ω in an "affordable" manner.
Four main goals were set:

□ it had to be stable;
□ it had to be fast;
□ it had to be powerful;
□ it had to be readily available.

Meeting goals

Stability and speed
When the project was started, in late 2002, there were two publicly available versions of Ω which could be called "current": version 1.15 and version 1.23 (which I would call the "old" and "new" version respectively).

The reason why the old one was not taken off the distributions when the new one was published is that the new one suffered from "excessive bloat" which rendered it essentially unusable: processing even a simple document with the new version could take from five to ten times longer than processing it in the old version, memory consumption during the processing was at least twice as much and the resulting DVI was enormous. These shortcomings of the new version were a result of the introduction of a very powerful enhancement, with interesting potential but that needed to be greatly refined before it could become of common usage.

Because of this, most Ω people kept using the old version, which was almost as fast as TEX (although obviously not as slim). This version, on the other hand, had some extremely serious bugs which caused it to crash whenever overfull boxes were present.

Finding which version to choose to base ℵ on was not easy. Indeed, the first release of ℵ (at the time ε-Ω, Release Candidate 0) which was just a proof of concept that ε-TEX could be merged with Ω, was available both in a 1.15–based version and in a 1.23–based version, although the officially supported one was the former, therefore with an implied preference for speed over stability, on the assumption that fixing bugs would have been easier than solving the speed/bloat problem. This has later proved to be indeed the best choice.

Power
The immediate outcome of the third goal was that ℵ should have provided both ε-TEX and Ω features; since of course Ω already provides some of the extensions provided by ε-TEX, we could limit ourselves to the programming enhancement (extra marks, protected macros, `\scantokens`, etc).

Availability
This was probably the most important goal, since it would have been the one that "made the difference" with, e.g., Ω 2: ℵ had to be available in a usable status as quickly as possible; this led, among other things, to the choice of stripping from the Ω base the code that dealt with SGML and XML, since it conflicted with the code that implemented the `\middle` primitive in ε-TEX. Priorities led to this decision.

History of releases

The first version of ε-Ω was released in December 2002; while the "official" version merged ε-TEX on Ω 1.15, a parallel release based on Ω 1.23 was also made available. That version was no more stable than any of its components, and had an extra few bugs that crept in during the adaptation of the ε-TEX changefiles to the Ω structure. In particular, it had all the bugs present in Ω 1.15, which made it scarcely usable due to the major problem with overfull boxes.

The second step was trying to fix the most outstanding bugs coming from the Ω 1.15 codebase. This led to the first version of ε-Ω testable in production–use environments, Release Candidate 1, around June 2003. This was the version presented at TUG2004.

Subsequent versions finally officially switched to the ℵ name; the last public release (Release Candidate 2) also fixed some other significant bugs and started introducing some minor new features (the most important being the `\boxdir` primitive to retrieve/change the direction of a box, a feature backported from Ω 1.23).

Status

ℵ is actively developed on the TEXlive repository. A mailing list for discussions concerning the present and future of the project, including both the core program itself and any ancillary tool, is available (`aleph@ntg.nl`).

Development, as always, is focused mainly on the discovering (and fixing) of bugs, but discussions on possible future features are welcome. Currently, issues and bugs in ΩTPs and their interaction with some ε-TEX features (namely protected macros and the `\scantokens` primitive) are the most prominent target.

Progress

The main focus of ℵ will always be one of stability. This means, among other things, that each new release is supposed to be at least as stable as the previous one.

A test suite analogous to the TRIP test for TEX is being discussed. Indeed, TRIP proved itself a trusted friend in the discovery and resolution of the most notable bugs coming from the Ω 1.15 code base, namely the ones dealing with overfull boxes and leaders.

Given the stability of ℵ, it is important to remark that this does not imply a static, frozen behavior (à la TEX); on the contrary, ℵ should be considered a foundation on which to build: experimental projects to test the implementation of new features and ideas are welcome, provided they are developed separately; once they reach enough stability to be available for production uses, they might be candidates for the introduction in future versions of ℵ.

Acknowledgements

I wish to thank

- Donald Ervin Knuth, for providing us all with TEX
- John Plaice and Yannis Haralambous, for giving us Ω
- Peter Breitenlohner and the NTS team, for giving us ε-TEX
- Idris S Hamid, Alan Hoenig and Hans Hagen for pushing me into attempting the merge and supporting me all time long
- all the distribution maintainers for their constant feedback, help and support, with a particular thank to Christian Schenk and Fabrice Popineau for their essential help in getting me started with the coding
- everybody in the TEX world for making it the great community it is

Notes

1. The project was originally named ε-Ω, since it provided both ε-TEX and Ω features.

Giuseppe Bilotta
Dipartimento di Matematica e Informatica
Università di Catania
viale A. Doria, 6
95125 Catania
Italy
gip.bilotta@iol.it