Understanding Internet Protocol Design Decisions

Michael Welzl, Carsten Griwodz, Safiqul Islam

Most RFCs should be efficient to use for people implementing a protocol or mechanism, and hence they usually do not contain much information on the design rationale.

This creates a problem for education: using RFCs, we can teach students _how_ a protocol works, but not necessarily _why_ it works the way it does. It also creates a problem for protocol design itself – both in research and engineering. For example, occasionally, in the IETF, newcomers present old (yet, to them, new) ideas, only to be told that the same idea was previously dismissed. This is a frustrating experience for these newcomers, and a waste of time for everyone involved.

On the other hand, Internet protocol design decisions are commonly documented in text, as they are taken as a result of email discussions, online discussions in github, or discussions at meetings which are also transcribed. The IETF email, meeting and github archives therefore constitute a large body of very valuable textual data, often containing detailed historic information about the decisions that were made.

Through the application of Natural Language Processing (NLP) techniques (i.e. machine reading), this information could automatically be brought into a form that is suitable for e.g. faceted semantic search, question answering, automated reasoning, including detection of contradictions, and more. Applications of these technologies would benefit both computer networks education and research on designing future Internet standards.

The first step towards such work is to download the textual archives of the IETF and bring them into a suitable form for NLP. We did this, focusing on mailing lists, and inserted the data into an Apache SOLR database. Without even applying NLP, this database already offers interesting insights into the IETF via simple queries on the text archive. Some results of such queries are documented in [1]. Code and step-by-step instructions to download the data and create such a database are available from [2].

[1] Michael Welzl, Stephan Oepen, Cezary Jaskula, Carsten Griwodz, Safiqul Islam: "Collaboration in the IETF: An Initial Analysis of Two Decades in Email Discussions", editorial note (not peer reviewed), ACM Computer Communication Review (CCR) 51(3), July 2021.

[2] https://github.com/CezaryRJ/Master_code/tree/main/DIY_IETF_Solr_%20database