IAB Workshop on Analysing IETF Data (AID)

Wrap-Up

Corinne Cath and Mirja Kühlewind - Dec 2, 2021

Session 1: Tools, data, and methods

Data
- **Data used in papers:** Mailing list archives, RFC Index, Datatracker (user account information; meeting proceedings and minutes)

- **Other sources:**
  - ietf.org webpage visitor stats (https://www.ietf.org/policies/web-analytics/)
  - GitHub and interop testings
  - Survey data
  - Download stats of RFC?

- Also combine with data from other SDOs/sources

Tools
- **The ietfdata Library**: datatracker, RFC index, email achieve
- **BigBang project**: email subscriptions and user account data from datatracker
- **The RFC Prolog Database**: RFC content and metadata
- **AUTHORSTATS** (https://www.arkko.com/tools/allstats/) and **IETF document statistics** (https://www.arkko.com/tools/docstats.html)

➔ Interest in deriving name and gender data
  - Open questions about privacy & legitimate use

Session 1: Tools, data, and methods

Data
- **Dat**...
  RFC...
  infor...
  minu...

- **Othe**...
  - i...
    (...
    ...
  - G...
  - S...
  - D...

- Also c...
  SDOs...

Getting started

- Explore the development environment and the v1 API

- Ask questions:
  - I'm available at email:rjsparks@nostrum.com and on the IETF Slack
  - Consider subscribing to tools-discuss@ietf.org

Also see https://notes.ietf.org/iab-aid-datatracker-database-overview

Instructions to set up local dev environment:
https://notes.ietf.org/iab-aid-data-resources

Session 2: Observations on affiliation and industry control

Questions from papers
- Impacts of corporate interest vs. individual contributions?
- Competition on leadership position?

Discussion & Hacking
- What are the affiliation trends? – Sebastian
- Are there indirect affiliation or fundings?
- How to categorize stakeholder groups / categories?
- How to model characteristics of companies/organizations? – Niels et al.
    - E.g. Growth/change in corporate participation

-> Combine data from different SDOs to improve results

Session 3: Community and diversity

Questions from papers
- Who are we now? Who is not there who should be?
- What is the existing diversity within the IETF?
- How can we increase that diversity?
- What would be different if the diversity had been stronger in the past?

Discussion
- What can we measure using existing datasets?
  - Measure diversity for what? 1) Better protocols = diversity as competency; 2) To do better work, e.g. make things relevant for operator in Africa
- Can we understand organizational structure and norms?
  - E.g. do certain mailing lists or groups drive people/newcomers away?
- How to interact with other groups like policy makers? Compare with other SDOs (which may have more resources)

Hacking
- Evaluate gender identification tooling across author name data -- Stephen
- Studying country/region participation (vs other organizations maybe too) - Wes
- Participation per-country and tenure in the IAB - Robin
- Privacy and PPI data like gender – Jari

Session 4: Publications, process, and decision-making

Questions from papers
- How does the IETF make decisions?
- Can we detect and study those decisions? Can we help improve the way with IETF makes decisions?
- What makes a successful RFC?

Discussion & Hacking
- Can we find WG last calls in the emails? – Michael
- Who writes RFCs? Newcomers vs. old timers? Affiliation? Nationality? – Carsten
- Data driven mining of homophilic users - Cluster in mailing list activity – Priyanka
- How does the process of seeking consensus differ for in-person vs. online meetings?
- Requirement for NLP to understand decision making process
- How to measure deployment?
  - Mesure download from code libraries & Analyse code similarity to assess deployment successes
  - Look at reference to RFCs to measure sucess
- How to measure the efficiency of the process? Where are delay in the process?

Session 5: Environmental Sustainability

1. How green is the IETF (output)?
    - Only few keyword matches in RFCs -> context needs to be analysed
    - More working on going to analyse mail list and draft
    - Awareness is important to start discussion in the IETF

2. How sustainable are IETF meeting?
    - Proposal to reduce to one in-person per year to reduce CO2 emission
    - Further discussion in shmoo or maybe GAIA

Final remarks and follow up work

Thanks for your attention

MEETINGS MEETINGS EVERYWHERE