Tracking issues in the Unicode Standard

Norbert Lindenberg, 2024-04-19

Proposal

I propose that:

- The UTC creates a public issue tracker to track issues in the Unicode Standard.
- Feedback received through the UTC's error reporting and PRI feedback web forms after the UTC meeting #180 (i.e., after making final decisions on the contents of Unicode 16.0) is converted to issues in this public issue tracker.
- Issues in the Unicode Standard that are currently recorded as UTC action items and whose resolution is not targeted for Unicode 16.0 are converted to issues in the issue tracker.
- Issues in the Unicode Standard that are currently tracked internally by the UTC working groups and whose resolution is not targeted for Unicode 16.0 are moved into the public issue tracker, except where the working group chairs determine that issues or their discussions need to be kept confidential.
- After UTC meeting #180, working groups use the public issue tracker to record their recommendations to the UTC on the issues tracked there.

This proposal focuses on "issues" as small pieces of feedback, as they currently typically come in through the UTC's web forms for error reporting and PRI feedback. It does not try to address major proposals, such as those for new scripts or major changes to algorithms, although these might be incorporated at a later time.

This document provides a tentative initial process for using the public issue tracker to show that using such a tracker is feasible. However, I expect that the process will be adapted by the UTC or the Release Management Group, and that it will evolve over time.

Problems to be solved

The UTC currently has no mechanism to track issues in the Unicode Standard from initial report to complete fix. Instead, a variety of tools are used at various stages: the web forms mentioned above, a document collecting feedback received through those forms, individual documents reporting and discussing issues, other documents in which the working groups discuss issues and make recommendations, minutes of the UTC, action items tasking individuals with either implementing fixes or with further research, documents proposing fixes, updates to the various documents that make up the standard, public review items asking for feedback on those updates, and eventually, if all goes well, a new version of the standard that addresses the issue.

The current process has a number of problems:

- It's very hard to trace an issue from initial report to resolution. Discussions of issues often include important information on reasons, precedents, constraints, or priorities. Such information is currently scattered over many different places.
- When an issue has no obvious fix, research often needs to be initiated. The UTC mechanism for that currently is an action item. An action item requires an owner. The UTC has no staff to which it could assign research tasks, so it has to look for a volunteer among member company employees and pro-bono contributors. If nobody volunteers, the issue is quietly dropped. In some cases, reporters are told that they should come with a solution instead of a problem.
- Action items are usually only assigned to people who participate in meetings of the UTC or its working groups. This excludes people who might be better qualified but don't have time to participate in long meetings or live in time zones far off the UTC's preferred time zone, U.S. Pacific time.
- As action items often have only a minimal description, and because they always have an owner, it's difficult for people who are not regular contributors to find out where their contributions might be needed and welcome.
- Some volunteers sign up for too many action items, or their priorities change. So action items pile up. As of 2024-01-14, the UTC had 40 open action items from UTC meeting #158 or earlier, that is, five years or older. A push to reduce that number followed, resulting in almost half of those action items being closed with no actual resolution to the underlying issues.
- Some issues don't get reported, as people realize that nothing happens if they don't present a ready-made solution or volunteer to take an action item for further research.
- As a result, the UTC has no institutional understanding of the overall quality of its main product, the Unicode Standard.

Public issue trackers

Pretty much every participant in the UTC knows that there's a better way: The companies they work for have used issue trackers for decades, many private, some public. Microsoft uses a public issue tracker for its OpenType documentation. Standards bodies such as the W3C use public issue trackers. The other technical committees of the Unicode Consortium do. Even UTC working groups do, privately.

Public issue trackers have several advantages:

- An issue can be traced from initial report to resolution. All information about the issue can be kept in one place.
- A public issue tracker becomes a place for collaboration. Many issues can't be solved by any single individual; they require the collaboration of a reporter, experts in various scripts, an expert in a particular algorithm, and a technical writer skilled in writing specifications.
- Issue trackers are available 24×7. This makes it more convenient for people to contribute who can't participate in meetings of the UTC or its working groups.
- Issues can be tagged with the names of scripts or algorithms, which makes it easier for outside experts to find issues where they can contribute.
- Issues can be created and kept around even if nobody is assigned to work on them. In fact, given the resources available to work on the standard, it's quite likely that unassigned issues will pile up. Open issues provide important information about the overall quality of the standard.
- The UTC can prioritize certain areas for upcoming releases of the standard and tag issues with priorities.
- Large numbers of open issues may indicate to potential sponsors that additional resources are needed to maintain and improve the standard.

Tentative initial process

Issue tracker platform: At this point, GitHub seems to be the right platform for the UTC issue tracker. Many potential contributors already have accounts and are familiar with it; some groups within the Unicode Consortium already use it. GitHub can't completely replace the current feedback mechanism, as GitHub isn't reliably accessible from everywhere (reportedly, it's not always

accessible from China), and Unicode can't require everyone to get a GitHub account in order to report an issue.

Issue creation: Issues can be created directly on GitHub, or indirectly via the UTC's existing web forms.

Screening: UTC staff screens new issues, rejects those that are out of scope for the Unicode Standard or are malformed (e.g., in a language that nobody in the UTC understands), and tags the remaining issues with the name of the UTC working group in whose area the issue belongs.

Initial review: The tagged UTC working group reviews new issues, closes obvious duplicates, closes requests that would require violating a Unicode policy, closes requests that have already been rejected and don't come with substantial new information. It also tags issues with the relevant scripts or algorithms.

Assignment: A working group participant or possibly a third party contributor self-assigns the issue to indicate that they will be responsible for driving the issue to resolution. This should be seen as a commitment to provide a recommendation to the UTC within the next six months.

Contributions: Working group participants and third parties comment on issues, providing their respective expertise.

Periodic review: The tagged UTC working group periodically reviews issues to make sure that assigned bugs are progressing towards a recommendation, and to try and find assignees.

Recommendation: The tagged working group adds a comment to the issue that provides a recommended resolution to the UTC, similar to those in current recommendation documents. The issue is tagged for the upcoming UTC meeting.

Decision: The UTC meeting reviews issues that have been tagged for it, accepting, rejecting, or modifying the proposed resolution. Decisions are recorded in the issue.

Spec updates: The relevant parts of the Unicode Standard are updated as directed by the UTC. Commits and issues are cross-referenced. The issue is closed.

ၰ၊ဝၰ၊