Well-Formed Unicode Strings Update

Michael Ficarra · TC39 · September 2022

Last Presented: August 2021 by Guy Bedford and Bradley Farias

String.prototype.isWellFormed

```
basically,
    function isWellFormed() {
      return !/\p{Surrogate}/u.test(this);
    }
but more efficient (sometimes)
```

Motivation

- interfacing with web APIs that use the USVString WebIDL type
- FFI for compile-to-JS languages whose String type is well-formed Unicode
- creating data in any format that expects well-formed Unicode text
- engines can implement it more efficiently than end users (sometimes)
- clearer than doing this with a RegExp

String.prototype.toWellFormed (#13)

```
basically,
   function toWellFormed() {
    return this.replaceAll(/\p{Surrogate}/gu, '\uFFFD');
}
```

but more readable and memorable

Motivation

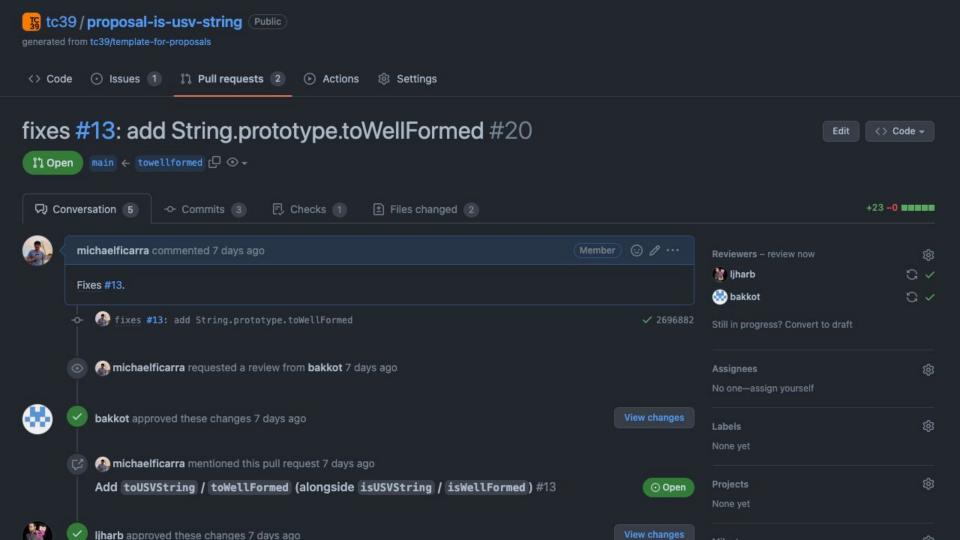
same as for isWellFormed

Potential Argument Against

Less motivated in the presence of isWellFormed

```
function toWellFormed(str) {
   return str.isWellFormed()
     ? str
     : str.replaceAll(/\p{Surrogate}/gu, '\uFFFD');
}
```

- takes advantage of performance of isWellFormed
- still hard to get right and still hard to read



Stage 2? Reviewers?