Intl.Segmenter

2021 October
Richard Gibson

https://github.com/tc39/proposal-intl-segmenter

Background

Breaking down strings with Intl.Segmenter

Taking guidance from <u>UAX 29</u>, we expose higher-level string breakdown to ECMAScript code by constructing locale-dependent iterators.

- **code unit**: `string.matchAll(/./sg)`, `[].slice.call(string)`
- **code point**: `Array.from(string)`
- **grapheme (cluster)**: `Array.from(new Intl.Segmenter(locale, {granularity: "grapheme"})).segment(string), data => data.segment)`
- **word**: `Array.from(new Intl.Segmenter(locale, {granularity: "word"})).segment(string), data => data.segment)`
- **sentence**: `Array.from(new Intl.Segmenter(locale, {granularity: "sentence"})).segment(string), data => data.segment)`

Current Status

What's new?

- Shipping in [V8](#) since Chrome 88 and [JSC](#) since [Safari Technology Preview Release 114](#)
- Don't create an `isWordLike` property for non-word segmentation [#128](#)
- Editorial tweaks [#126](#), [#130](#), [#135](#), [#143](#), [#144](#), [#145](#)
- Approved for Stage 4 by TC39-TG2 (ECMA-402 Internationalization API Task Group) on [2021-10-07](#)

Open issues

- None blocking v1

Pending

- MDN updates: https://github.com/mdn/content/pull/8402
- ecma402 PR: https://github.com/tc39/ecma402/pull/553

Next steps

Questions?

Feedback?

Stage 4?