

# Intl.Segmenter

Daniel Ehrenberg



# Summary

- What?
  - Locale-tailored grapheme, word and sentence breaks ([UTS 29](#))
  - Line breaks ([UAX 14](#))
- Why?
  - Complex in-browser rich text editors, design tools, rendering
  - Want to deprecate Chrome's Intl.v8BreakIterator
- Out of scope
  - Hyphenation

# API shape

```
// Create a segmenter in your locale
let segmenter = Intl.Segmenter("fr", {type: "word"});

// Get an iterator over a string
let segmentIterator = segmenter.segment(
  "Ceci n'est pas une pipe");

// Iterate over it!
for (let {index, breakType} of segmentIterator) {
  console.log(`index: ${index} breakType: ${breakType}`);
  break;
}
// logs the following to the console:
// index: 3 breakType: letter
```

## Short-cut accelerated API:

```
let segmentIterator = segmenter.segment(
  "Ceci n'est pas une pipe");

// next() returning undefined
segmentIterator.advance()

// index of current break result
segmentIterator.index()

// breakType of current break result
segmentIterator.breakType()
```