Unicode line breaking

Where a line may end inside a run of text follows the Unicode line breaking algorithm (UAX #14). Kana and ideographs break between characters; word-break: keep-all keeps them together and break-all also breaks between Latin letters.

normal

テキストキストテストスキストテ

keep-all

テキストキスト
テストスキスト

break-all

Chargoggagoggmanchaugg 2024

line-break sets how strictly characters that should not start a line are kept off it: strict and normal keep a small kana (□) with the character before it, loose lets it start a line, and anywhere allows a break after every character.

line-break: strict

テキスト　テキスト　テキスト　テキ

line-break: loose

テキスト　テキスト
　テキスト　テキ

line-break: anywhere

Chargoggagoggmanchaugg 2024

Latin text

A well-known hyphen breaks after itself; abc-123 stays whole, and !important, and/or and 23/Jan/Feb are never split around the punctuation.

Across elements the same rules apply: foo**bar** stays whole, **well-***known* breaks after its hyphen, **abc-***123* does not, and *and*/**or** is never split.

Text in (parentheses) and “quotes” does not break inside them, and 3.14, $5.00 and 1,000 are single tokens.