

Bug files report for:

Summary: Word Boundaries (Hyphenation) in Indian languages (UAX#29)

Description:

I. Overview:

Word Boundaries (Hyphenation):

Word boundaries are used in a number of different contexts. The most familiar ones are selection (double-click mouse selection, or “move to next word” control-arrow keys), and “Whole Word Search” for search and replace. They are also used in database queries, to determine whether elements are within a certain number of words of one another.

Recommended solution: ABNF Valid Segmentation and hyphenation dictionary (if available)

Sentence Boundaries

Recommended solution: Some special sentence boundaries like the double poorna virama, possibly with numbers (as in Sanskrit text, shlokas etc.) A string of Unicode-encoded text often needs to be broken up into text elements programmatically. Common examples of text elements include what users think of as characters, words, lines (more precisely, where line breaks are allowed), and sentences. The precise determination of text elements may vary according to orthographic conventions for a given script or language. The goal of matching user perceptions cannot always be met exactly because the text alone does not always contain enough information to unambiguously decide boundaries. For example, the period (U+002E FULL STOP) is used ambiguously, sometimes for end-of-sentence purposes, sometimes for abbreviations, and sometimes for numbers. In most cases, however, programmatic text boundaries can match user perceptions quite closely, although sometimes the best that can be done is not to surprise the user.

Solution

Grapheme Cluster Boundaries: ABNF Valid Segmentation Based, Possible Extension for handling some cases (?)

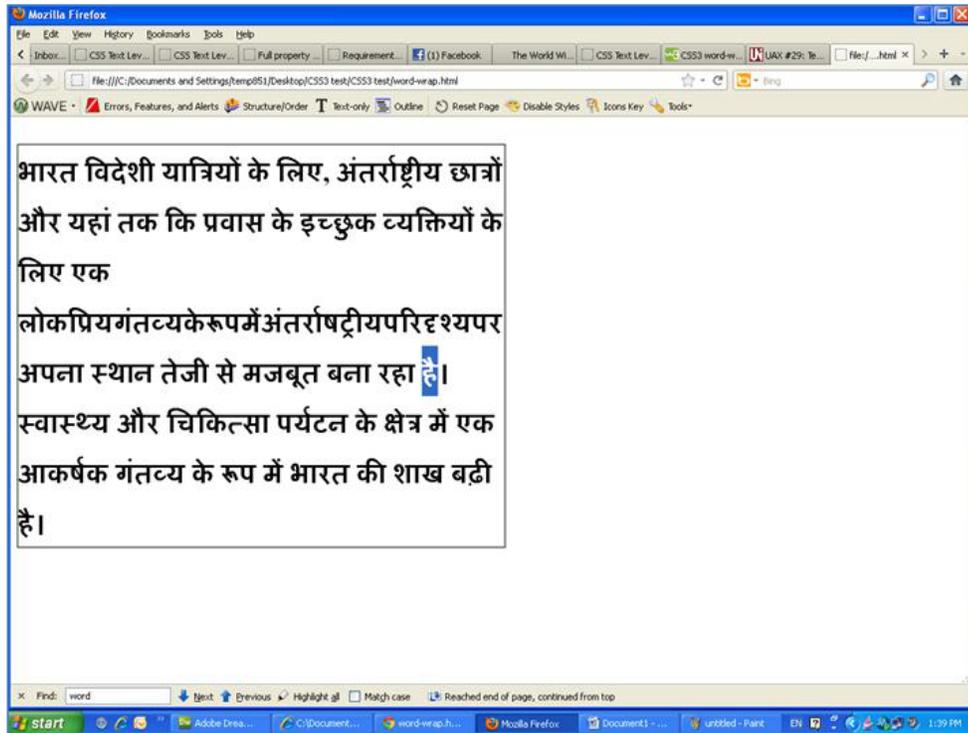
Deletion and backspace: Code point wise as well as ABNF Valid Segmentation

Mouse Selection: At ABNF Valid Segmentation and code point level

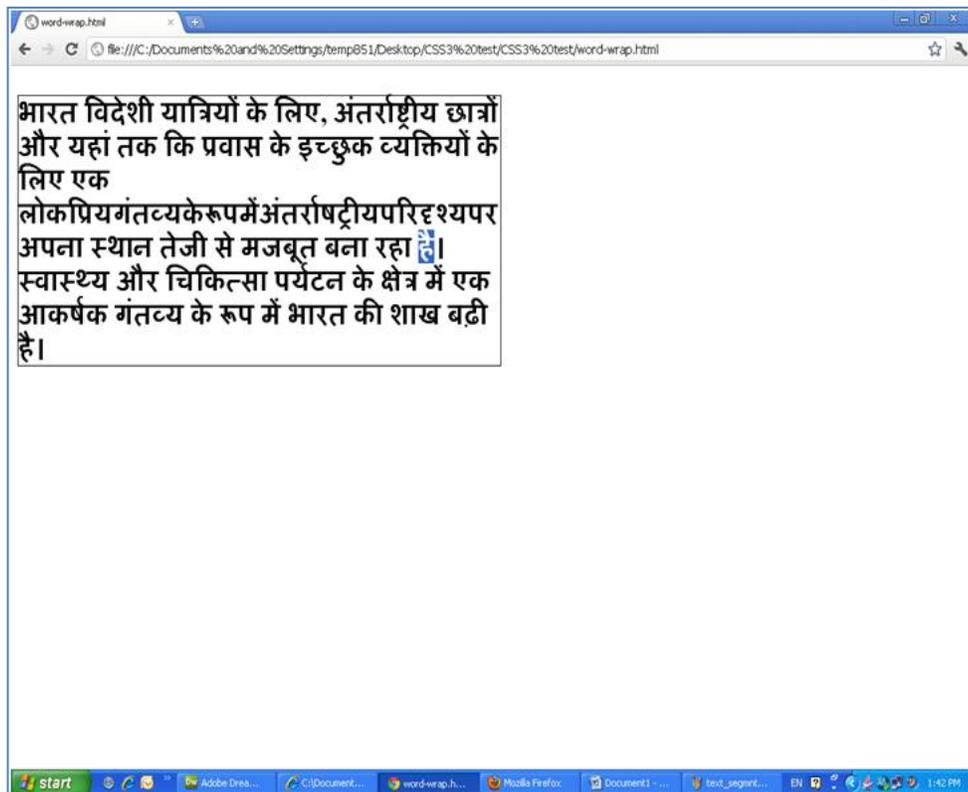
View Report: <http://w3cindia.in/ABNFValidSegmentationdocument.html#uax29>

II. Actual Result of above test on following web browser

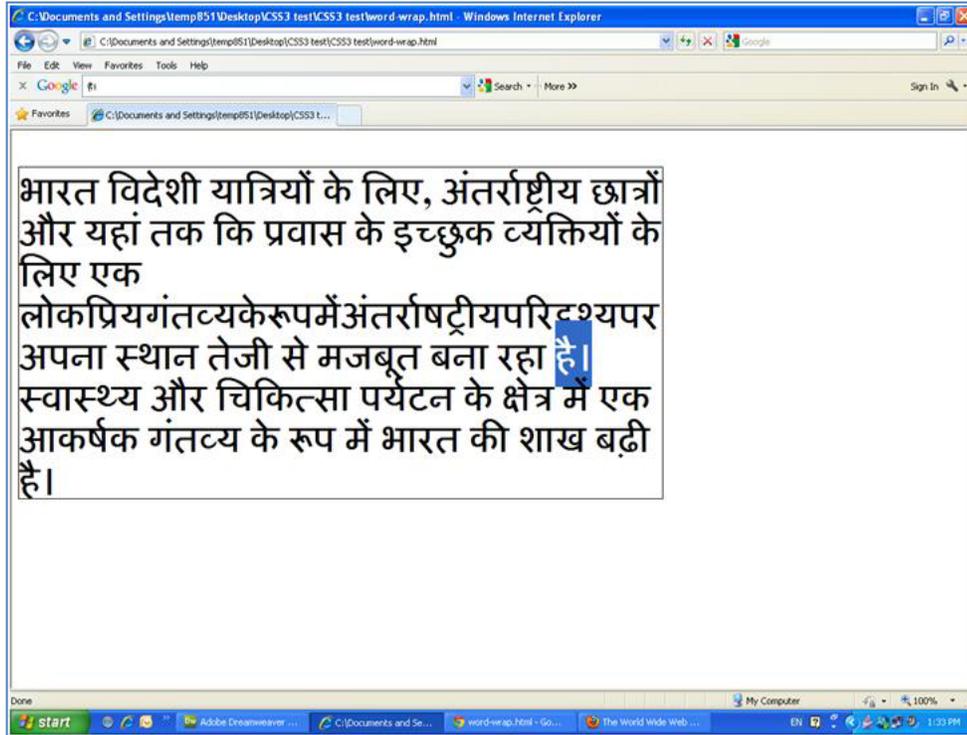
1. Mozilla Firefox(version 16.0.1)



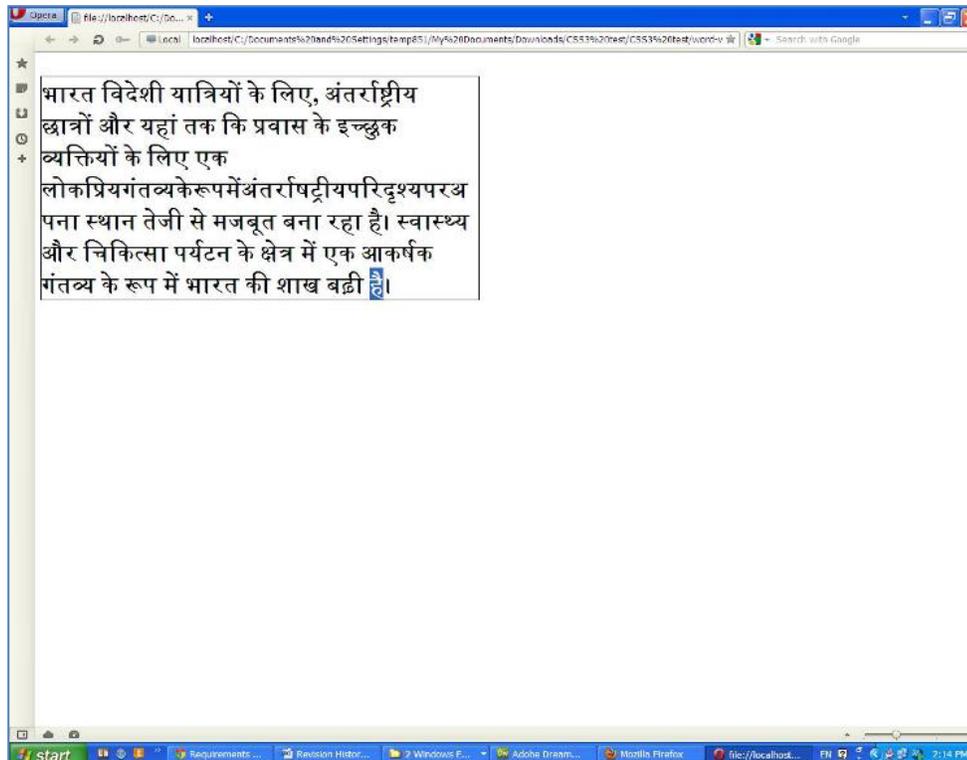
2. Google Chrome(version 22.0.1229.94m)



3. Internet Explorer(version 8.0.6001.18702)



4. Opera (version 12.02)



III. **Expected Result** of above test on web browser

भारत विदेशी यात्रियों के लिए, अंतर्राष्ट्रीय छात्रों और यहां तक कि प्रवास के इच्छुक व्यक्तियों के लिए एक लोकप्रिय गंतव्य के रूप में अंतर्राष्ट्रीय परिदृश्य पर अपना स्थान तेजी से मजबूत बना रहा है। स्वास्थ्य और चिकित्सा पर्यटन के क्षेत्र में एक आकर्षक गंतव्य के रूप में भारत की शाख बढ़ी है।

Note: - Best Actual Result output- Mozilla Firefox, Google Chrome and Opera.

IV. **Build Date & Platform:**

Build 2012-11-19 on Windows XP professional (Service Pack 3)