Misspelled Words

The World Wide Web provides a new way of measuring how difficult it is to spell a word. The Web consists of millions of documents containing billions of words. Each Web page is usually prepared with some care, since it is expected that it will be read by strangers. It is the modern equivalent of the edited document, or at least as close as we can come to it. These Web pages represent the combined written output of hundreds of thousands of fairly well-educated people. Presumably, a word that is hard to spell will be misspelled frequently on the Web.

There are programs called "spiders" that walk the Web, visiting every page, extracting and indexing every word. The index produced by the spider is used for full-text searching of the Web. As a by-product of this process, these spiders compute word counts for every word on every page on the Web. This database is then freely accessible to everyone over the Web itself. It consists of word counts for several billion word instances.

One definition of the hardest word to spell is the word that is misspelled most often on the Web. Below we exhibit a table that consists of the word counts for several words that are well-known to be difficult to spell, provided that they occur at least 100 times on the Web. We require that they occur at least 100 times to avoid statistical anomalies. We also require that the documents they occur in be written in English.

The first column of the list is the ratio of the count of the word in the second column (the correct spelling) divided into the sum of the counts of the words in the subsequent columns (various incorrect spellings). It is, in short, the misspelling ratio. The list is sorted by increasing value of this ratio.

At the top of the list we find extremely obscure words that are never misspelled. This does not mean they are easy to spell; rather it means that anyone who knows the word either already knows how to spell it, or looks it up in the dictionary every time. We cannot tell which is the case, so we cannot accurately award the title of "hardest word to spell" to these words. To avoid this ambiguity we refine the definition of "hardest word to spell" to be "word most frequently misspelled."

Proceeding down the list, we come to the .500 ratio. Below this point are words that are misspelled half as often as they are correctly spelled. We find, however, that in almost every case, the spellings in column three are listed in many dictionaries as variant spellings. These cannot be considered as true misspellings.

However, one word stands out: "desiccate." This is the only word with a misspelling ratio greater than 1.000, which means that it is misspelled more often than it is correctly spelled. Surprisingly, no dictionary lists either of its common misspellings as a variant. These two facts combine to justify the awarding of the title "hardest word to spell" to the word "desiccate."

Ratio	Correct Spelling	Incorrect (or Variant) Spelling(s)
0.000	eisteddfodau (114)	–
0.000	kwashiorkor (963)	–
0.001	irritable (22889)	irritible (20), iritable (20)
0.005	development (1626032)	developement (8380)
0.006	rijsttafel (296)	risttafel (1), ristafel (1)
0.007	dissipate (22873)	disipate (160), disippate (3)
0.010	Orszaggyules (949)	Orszagyules (8), Orzaggyules (2)
0.010	similar (3000671)	similiar (31176)
0.011	Houyhnhnm (259)	Houhnhnm (3)
0.011	a lot (2549399)	alot (30399)
0.011	tyranny (55802)	tyrrany (626)
0.011	yield (301799)	yeild (3466)
0.017	Phacochoerus (112)	Phacocoerus (1)

0.017	exceed (26813)	excede (460)
0.017	privilege (234299)	privelege (3040), priviledge (1143)
0.017	vichyssoise (857)	vichysoise (11), vishyssoise (4)
0.018	recommend (88674)	reccomend (1616)
0.018	repetition (109451)	repitition (2005)
0.020	pursue (33067)	persue (675)
0.021	subpoena (29864)	subpena (657)
0.023	seize (108891)	sieze (2593)
0.025	despair (138430)	dispair (3563)
0.028	dyspnoea (769)	dyspnoia (17), dispnoea (5)
0.029	independent (324037)	independant (9484)
0.032	occasion (36851)	ocassion (1186)
0.037	consensus (24106)	concensus (914)
0.038	receive (358934)	recieve (13860)
0.041	broccoli (52005)	brocolli (1182), broccolli (970)
0.050	eschscholtzia (236)	escholtzia (12)
0.053	desperate (13189)	desparate (702)
0.054	momento (238363)	memento (13053)
0.055	inadvertent (21369)	inadvertant (1189)
0.060	all right (229496)	alright (13879)
0.063	separate (163658)	seperate (10318)
0.071	battalion (11672)	batallion (832)
0.080	Nietzschean (1600)	Nietzchean (118), Nietschean (10)
0.083	embarrassment (84998)	embarassment (7088)
0.084	existence (49575)	existance (4213)
0.084	gauge (219212)	guage (18604)
0.085	anoint (6839)	annoint (586)
0.091	bouillabaisse (3611)	bouillabaise (323), bouilabaisse (7)
0.091	drunkenness (10445)	drunkeness (954)
0.098	liaison (35270)	liason (3470)
0.104	harass (20802)	harrass (2184)
0.116	embarrass (14239)	embarass (1652)
0.124	cemetery (38521)	cemetary (4781)
0.125	indispensable (8109)	indispensible (1014)
0.127	inoculate (3440)	innoculate (437)
0.132	weird (45715)	wierd (6055)
0.137	occurrence (155168)	occurence (21358)
0.142	insistent (25102)	insistant (3567)
0.148	definitely (37261)	definately (5534)
0.149	sacrilegious (3381)	sacreligious (505)
0.151	irresistible (50955)	irresistable (7744)
0.154	ecstasy (78335)	ecstacy (12105)
0.155	accidentally (180112)	accidently (28000)
0.190	accommodate (409589)	accomodate (77902), acomodate (100)
0.261	coolly (10893)	cooly (2849)
0.279	carillonneur (261)	carilloneur (68), carrillonneur (5)
0.340	judgment (35243)	judgement (11996)
0.363	peddler (10564)	pedlar (2737), pedler (1102)
0.499	millennium (198547)	millenium (97015), milenium (2178)
0.527	praemunire (91)	premunire (48)

0.562	liquify (988)	liquefy (556)
0.696	dumbbell (4419)	dumbell (3076)
0.705	supersede (1378)	supercede (972)
0.894	minuscule (502)	miniscule (449)
1.090	desiccate (528)	dessicate (361), desicate (215)

half islands—longest Morse code reversal pair for a phrase, measured in number of dots and dashes [superhelices] (33)—*peninsulas*

half-joking—hyphenated word containing most consecutive letters in the alphabet after letter f (7)

halfpenny—example of a word that contains silent letter f—*English coin*

halftimes—see "specials"

halve—example of homophones differing by letter l [have]

hamamelidaceae—uncommon word with lowest letter sum for length 14 (81)—longest uncommon word with letters from first half of alphabet (14)—*witch hazel*



hamamelidaceae

han—see "thanatopsides"—*people of Central Asia*

handcraftsmanships—highest consonants minus vowels for an uncommon word (10-t)

hand gallops—longest phrase isomorse pair, measured in number of dots and dashes [signature loans] (34)—longest well-mixed isomorse pair for a phrase [signature loans] (14-t)—*moderate gallops*

handle blanks—longest phrase with consecutive letters adjacent on phone keypad beginning with letter h (12-t)—*lumber for making handles*

handsome—example of a word that contains silent letter d

handstands—longest nontrivial internal head 'n' tail for a common word (9-t)

happed—longest full vowel (including y) substitution with an uncommon word [happed & hepped & hipped & hopped & hupped & hypped] (6)

har—see "cantharidates"—*thick cold sea fog*

hard—example of an antonym of two different words, beginning with letter h [soft, easy]

harlequin opal—longest shiftgram pair for a phrase [presumptively] [shift of 4] (13-t)—*reddish opal*

harns—see "hairns"—*Scottish for brains*

haroset—see "charoseth"—*Jewish symbolic food*

hashish—common word containing the most repeats of letter h (3-t)

haunched—longest one-syllable word beginning with letter h (8)—*with hips or hindquarters*

have—see "halve"

haw—example of a word that exists only in an idiom, beginning with letter h [hem and haw]

hazy—shortest common word that cannot have a letter added and transposed to form another word (4-t)

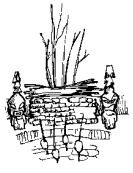
heading cabbage—phrase with lowest letter sum for length 14 (69)—*type of cabbage*

heading cabbages—phrase with lowest letter sum for length 15 (88)

hee-hee—longest hyphenated word containing one type of consonant for consonant h (6-t)—*giggle*

hegari—example of a word that ends with silent letter i—*Sudanese sorghums*

heiau—longest uncommon word containing one type of consonant for consonant h (5)—*Hawaiian temple*



heiau

height—example of a word spelled with e before i, beginning with letter h

heir—example of a word that is a confusing code for letter h [air]—see "aer"