

Algorithm for Stemmer

Definitions:

CONSONANT: a letter other than A, E, I, O, U, and Y preceded by consonant

VOWEL: any other letter

With this definition, all words are of the form:

(C)(VC)m(V)

C=string of one or more consonants (con+)

V=string of one or more vowels

E.g.,

Troubles

C V CVC

The rules are of the form:

(condition) S1 -> S2

Where S1 and S2 are suffixes

Conditions:

m : The measure of the stem

*S : The stem ends with S

v: The stem contains a vowel

*d : The stem ends with a double consonant

Examples :-

SSES -> SS

caresses -> caress

IES -> I

ponies -> pony

ties -> ty

SS -> SS

caress -> caress

S

cats -> cat

Past tense, progressive :

(m>1) EED -> EE

Condition verified: agreed -> agree

Condition not verified: feed -> feed

(*V*) ED

Condition verified: plastered -> plaster

Condition not verified: bled -> bled

(*V*) ING

Condition verified: motoring -> motor

Condition not verified: sing -> sing