### Task 4 – Assignment 1 item 1: Stemming

Search for existing open source stemmers (Porter or other) for English, select one and run it on a text file of your choice (at least 500 words). Identify false positives and false negatives.

## Stemmer:

<http://snowball.tartarus.org/demo.php>

## Sample Text:

Microsoft's Kinect launched back in November and, according to Microsoft, the hands-free device has already hit the 10 million units sold worldwide plateau.

In fact, the Kinect has sold so quickly that Guinness has awarded it with a new world record as the fastest selling consumer entertainment device. The record will show up in the 2011 gamer's edition of the book, and relates to the first 60 days the Kinect was on the market, as it managed to sell eight million units in that time frame. Software didn't fare quite as well, however, with a total of 10 million games sold, making for an attach rate of just one game per Kinect.

Microsoft also announced some new Kinect-specific features that will be coming in the near future.

## Output:

**False Positive  
False Negative**

microsoft -> microsoft

s -> s

kinect -> kinect

launched -> launch

back -> back

in -> in

**november -> novemb**

and -> and

**according -> accord**

to -> to

microsoft -> microsoft

the -> the

**handsfree -> handsfre**

**device -> devic**

has -> has

**already -> alreadi**

hit -> hit

the -> the

million -> million

units -> unit

sold -> sold

**worldwide -> worldwid**

plateau -> plateau

in -> in

fact -> fact

the -> the

kinect -> kinect

has -> has

sold -> sold

so -> so

quickly -> quick

that -> that

**guinness -> guin**

has -> has

awarded -> award

it -> it

with -> with

a -> a

new -> new

world -> world

record -> record

as -> as

the -> the

fastest -> fastest

selling -> sell

consumer -> consum

entertainment -> entertain

**device -> devic**

the -> the

record -> record

will -> will

show -> show

up -> up

in -> in

the -> the

gamer -> gamer

s -> s

edition -> edit

of -> of

the -> the

book -> book

and -> and

relates -> relat

to -> to

the -> the

first -> first

days -> day

the -> the

kinect -> kinect

was -> was

on -> on

the -> the

market -> market

as -> as

it -> it

managed -> manag

to -> to

sell -> sell

eight -> eight

million -> million

units -> unit

in -> in

that -> that

time -> time

**frame -> frame**

**software -> softwar**

didn -> didn

t -> t

fare -> fare

quite -> quit

as -> as

well -> well

**however -> howev**

with -> with

a -> a

total -> total

of -> of

million -> million

games -> game

sold -> sold

making -> make

for -> for

an -> an

attach -> attach

rate -> rate

of -> of

just -> just

one -> one

game -> game

per -> per

kinect -> kinect

microsoft -> microsoft

also -> also

announced -> announc

some -> some

new -> new

kinectspecific -> kinectspecif

features -> featur

that -> that

will -> will

be -> be

coming -> come

in -> in

the -> the

near -> near

future -> futur