

```

function [C,V]=tryit(urlp)

urlp=char(urlp)
try
html=webread(urlp);

% Use regular expressions to remove undesired markup.
txt = regexp(html, '<script.*?/script>', '');
txt = regexp(txt, '<style.*?/style>', '');
txt = regexp(txt, '<.*?>', '');

file = fopen('trrrryy.txt', 'w');
fprintf(file, '%s', txt);

Str = fileread('trrrryy.txt');
CStr = strsplit(Str, '\n');

% split each string by delimiters and add to C
C = [];
for i=1:length(CStr)
    if ~isempty(CStr{i})
        C = [C strsplit(CStr{i}, {'/', ' '})];
    end
end

% remove the empty cells
try
    C = C(~cellfun('isempty', C));

%disp(C)
for i=1:numel(C)
    p=char(C(i));
    New_String{i} = p(find(~isspace(p)));
end
%Cc=strjoin(C);
C=New_String;
C = C(~cellfun('isempty', C));

%disp(C)

C = regexp(C, '&nbsp;', '');
C = regexp(C, '&', '');
C = regexp(C, '[???%="$*.;^.,+?!?#&???( )-]', '');
C = regexp(C, '[1234567890]', '');
C = regexp(C, '[????????????????????????????????????]', '');
C = regexp(C, '[????????????????????????????????????_]', '');
C= regexp(C, '|', '');
C= regexp(C, '-->', '');

C = C(~cellfun('isempty', C));
C=remove_stopwords2(C);
C=remove_stopwords3(C);
C=remove_stopwords(C);
C=remove_stopwords1(C);
C=remove_words(C);
C=remove_words1(C);

%disp(C)

i=1;
while i<numel(C)

```

```

    k = strfind(txt,C(i));
    kk(i)=numel(k);
    C = dedupeCell(C);
    %disp('*****')
    %disp(C)
    i=i+1;
end

ky = strfind(txt,C(numel(C)));
kky=numel(ky);

%disp(C)
%disp('*****')

kk(numel(kk)+1)=kky;

for i=1:numel(kk)
    if kk(i)==0
        kk(i)=1;
    end
end
V=kk;
catch
end
catch
    C=0;
    V=0;
end
end

```