

```

function [oc]=occ(urlp)

urlp=char(urlp)
urlp=['https://' urlp];
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('exp.txt', 'w');
fprintf(file, txt);

fid=fopen('exp.txt');

A = {'buy'; 'sell'; 'sold'; 'services'; 'stock';
'price'; 'achat'; 'vendu'; 'offre'; 'liquidation'; 'paiement'; 'catalogue'; 'promotion'; 'magas
in'; 'prix'; 'discount'; 'solde'; 'product'; 'sale'; 'credit'};

A1 = {'Buy'; 'Sell'; 'Sold'; 'Services'; 'Stock';
'Price'; 'Achat'; 'Vendu'; 'Offre'; 'Liquidation'; 'Paiement'; 'Catalogue'; 'Promotion'; 'Magas
in'; 'Prix'; 'Discount'; 'Solde'; 'Product'; 'Sale'; 'Credit'};
str=textscan(fid, '%s');
fclose(file);
text=str{:};

for k=1:numel(A)
    freq(k)=sum(ismember(text, A{k}));
    freq1(k)=sum(ismember(text, A1{k}));
end

```