Anthony Johnson

Homework 2: ContactFinder

**Report**

1. So, I basically went through and did as many as I could through trial and error. I have not included any appends that returned false positives, but they are in my code and clearly labeled. I used Regexpal to help me with phone numbers. I did not include the provided examples in the report.

Cheriton email

[uma@cs.stanford.EDU](mailto:uma@cs.stanford.EDU)

It’s basically the same as the original example, except you capitalize ‘edu’

epatterns.append(‘([A-Za-z.]+)@([A-Za-z.]+)\.EDU’)

Dabo email

dabo @ cs.stanford.edu

Again, basically the same as the example with spaces around the @, but with an extra space.

epatterns.append('([A-Za-z.]+)\s\s@\s\s([A-Za-z.]+)\.edu')

Engler email

engler WHERE stanford DOM edu

Replaced @ with ‘WHERE’ and . with ‘DOM’ and added spaces.

epatterns.append('([A-Za-z.]+) WHERE ([A-Za-z.]+) DOM edu')

Latombe email

latombe<del>@cs.stanford.edu

Similar to the examples, with ‘<del>’ added.

epatterns.append('([A-Za-z.]+)<del>@([A-Za-z.]+)\.edu')

Levoy email

ada&#x40;graphics.stanford.edu

Just replaced the @ with ‘&#x40;’

epatterns.append('([A-Za-z.]+)&#x40;([A-Za-z.]+)\.edu')

Bgirod phone number

(650) 724-3648

I had to add in extra parentheses, so adding [(] & [)] was necessary.

ppatterns.append('(?:[(])(\d{3})(?:[)])[ ]\*(\d{3})-(\d{4})')

Nass phone number

[650] 723-5499

Similar to Bgirod, but with added brackets instead of parentheses, plus an added space.

ppatterns.append('(?:[[])(\d{3})(?:[]])\s(\d{3})-(\d{4})')

Pal phone number

650 725 9046

Simply removed hyphens.

ppatterns.append('(\d{3}) (\d{3}) (\d{4})')

Shosham phone number

650 723-3432

Removed one hyphen.

ppatterns.append('(\d{3}) (\d{3})-(\d{4})')

Tim phone number

(650)724-9147

Similar to Bgirod, except no spaces.

ppatterns.append('(?:[(])(\d{3})(?:[)])(\d{3})-(\d{4})')

**Final Results**

Before expressions were added (with provided examples):

TP:41

FP: 0

FN: 76

After I hit a wall, I ended up with (It was TP:106, FP:7, FN:11, but I started getting random errors after changing nothing, so I copy-pasted everything into a fresh file and got the result below instead. No idea what happened.):

TP: 67

FP: 7

FN: 50

Output:

True Positives (106):

{('ashishg', 'e', 'ashishg@stanford.edu'),

('ashishg', 'e', 'rozm@stanford.edu'),

('ashishg', 'p', '650-723-1614'),

('ashishg', 'p', '650-723-4173'),

('ashishg', 'p', '650-814-1478'),

('balaji', 'e', 'balaji@stanford.edu'),

('bgirod', 'p', '650-723-4539'),

('bgirod', 'p', '650-724-3648'),

('bgirod', 'p', '650-724-6354'),

('cheriton', 'e', 'cheriton@cs.stanford.edu'),

('cheriton', 'e', 'uma@cs.stanford.edu'),

('cheriton', 'p', '650-723-1131'),

('cheriton', 'p', '650-725-3726'),

('dabo', 'e', 'dabo@cs.stanford.edu'),

('dabo', 'p', '650-725-3897'),

('dabo', 'p', '650-725-4671'),

('engler', 'e', 'engler@lcs.mit.edu'),

('engler', 'e', 'engler@stanford.edu'),

('eroberts', 'e', 'eroberts@cs.stanford.edu'),

('eroberts', 'p', '650-723-3642'),

('eroberts', 'p', '650-723-6092'),

('fedkiw', 'e', 'fedkiw@cs.stanford.edu'),

('hager', 'p', '410-516-5521'),

('hager', 'p', '410-516-5553'),

('hager', 'p', '410-516-8000'),

('hanrahan', 'e', 'hanrahan@cs.stanford.edu'),

('hanrahan', 'p', '650-723-0033'),

('hanrahan', 'p', '650-723-8530'),

('horowitz', 'p', '650-725-3707'),

('horowitz', 'p', '650-725-6949'),

('jurafsky', 'p', '650-723-5666'),

('kosecka', 'e', 'kosecka@cs.gmu.edu'),

('kosecka', 'p', '703-993-1710'),

('kosecka', 'p', '703-993-1876'),

('kunle', 'e', 'darlene@csl.stanford.edu'),

('kunle', 'e', 'kunle@ogun.stanford.edu'),

('kunle', 'p', '650-723-1430'),

('kunle', 'p', '650-725-3713'),

('kunle', 'p', '650-725-6949'),

('lam', 'e', 'lam@cs.stanford.edu'),

('lam', 'p', '650-725-3714'),

('lam', 'p', '650-725-6949'),

('latombe', 'e', 'asandra@cs.stanford.edu'),

('latombe', 'e', 'latombe@cs.stanford.edu'),

('latombe', 'e', 'liliana@cs.stanford.edu'),

('latombe', 'p', '650-721-6625'),

('latombe', 'p', '650-723-0350'),

('latombe', 'p', '650-723-4137'),

('latombe', 'p', '650-725-1449'),

('levoy', 'e', 'ada@graphics.stanford.edu'),

('levoy', 'e', 'melissa@graphics.stanford.edu'),

('levoy', 'p', '650-723-0033'),

('levoy', 'p', '650-724-6865'),

('levoy', 'p', '650-725-3724'),

('levoy', 'p', '650-725-4089'),

('manning', 'e', 'dbarros@cs.stanford.edu'),

('manning', 'e', 'manning@cs.stanford.edu'),

('manning', 'p', '650-723-7683'),

('manning', 'p', '650-725-1449'),

('manning', 'p', '650-725-3358'),

('nass', 'e', 'nass@stanford.edu'),

('nass', 'p', '650-723-5499'),

('nass', 'p', '650-725-2472'),

('nick', 'e', 'nick.parlante@cs.stanford.edu'),

('nick', 'p', '650-725-4727'),

('ok', 'p', '650-723-9753'),

('ok', 'p', '650-725-1449'),

('pal', 'p', '650-725-9046'),

('psyoung', 'e', 'patrick.young@stanford.edu'),

('rajeev', 'p', '650-723-4377'),

('rajeev', 'p', '650-723-6045'),

('rajeev', 'p', '650-725-4671'),

('rinard', 'e', 'rinard@lcs.mit.edu'),

('rinard', 'p', '617-253-1221'),

('rinard', 'p', '617-258-6922'),

('serafim', 'p', '650-723-3334'),

('serafim', 'p', '650-725-1449'),

('shoham', 'e', 'shoham@stanford.edu'),

('shoham', 'p', '650-723-3432'),

('shoham', 'p', '650-725-1449'),

('subh', 'e', 'subh@stanford.edu'),

('subh', 'p', '650-724-1915'),

('subh', 'p', '650-725-3726'),

('subh', 'p', '650-725-6949'),

('thm', 'e', 'pkrokel@stanford.edu'),

('thm', 'p', '650-725-3383'),

('thm', 'p', '650-725-3636'),

('thm', 'p', '650-725-3938'),

('tim', 'p', '650-724-9147'),

('tim', 'p', '650-725-2340'),

('tim', 'p', '650-725-4671'),

('ullman', 'e', 'ullman@cs.stanford.edu'),

('ullman', 'p', '650-494-8016'),

('ullman', 'p', '650-725-2588'),

('ullman', 'p', '650-725-4802'),

('widom', 'e', 'siroker@cs.stanford.edu'),

('widom', 'e', 'widom@cs.stanford.edu'),

('widom', 'p', '650-723-0872'),

('widom', 'p', '650-723-7690'),

('widom', 'p', '650-725-2588'),

('zelenski', 'e', 'zelenski@cs.stanford.edu'),

('zelenski', 'p', '650-723-6092'),

('zelenski', 'p', '650-725-8596'),

('zm', 'e', 'manna@cs.stanford.edu'),

('zm', 'p', '650-723-4364'),

('zm', 'p', '650-725-4671')}

False Positives (7):

('plotkin', 'e', 'server@infolab.stanford.edu')

('subh', 'e', 'uma@cs dot stanford.edu')

gold: ('subh', 'e', 'subh@stanford.edu')

gold: ('subh', 'e', 'uma@cs.stanford.edu')

gold: ('subh', 'p', '650-724-1915')

gold: ('subh', 'p', '650-725-3726')

gold: ('subh', 'p', '650-725-6949')

('dlwh', 'e', 'd-l-w-h@s-t-a-n-f-o-r-d-.edu')

gold: ('dlwh', 'e', 'dlwh@stanford.edu')

('serafim', 'e', 'serafim@cs dot stanford.edu')

gold: ('serafim', 'e', 'serafim@cs.stanford.edu')

gold: ('serafim', 'p', '650-723-3334')

gold: ('serafim', 'p', '650-725-1449')

('hager', 'e', 'hager@cs dot jhu.edu')

gold: ('hager', 'e', 'hager@cs.jhu.edu')

gold: ('hager', 'p', '410-516-5521')

gold: ('hager', 'p', '410-516-5553')

gold: ('hager', 'p', '410-516-8000')

('jks', 'e', 'jks@robotics;stanford;.edu')

gold: ('jks', 'e', 'jks@robotics.stanford.edu')

('jure', 'e', 'server@cs.stanford.edu')

False Negatives (11):

{('dlwh', 'e', 'dlwh@stanford.edu'),

('hager', 'e', 'hager@cs.jhu.edu'),

('jks', 'e', 'jks@robotics.stanford.edu'),

('jurafsky', 'e', 'jurafsky@stanford.edu'),

('ouster', 'e', 'ouster@cs.stanford.edu'),

('ouster', 'e', 'teresa.lynn@stanford.edu'),

('pal', 'e', 'pal@cs.stanford.edu'),

('serafim', 'e', 'serafim@cs.stanford.edu'),

('subh', 'e', 'uma@cs.stanford.edu'),

('ullman', 'e', 'support@gradiance.com'),

('vladlen', 'e', 'vladlen@stanford.edu')}

Summary: tp=106, fp=7, fn=11

1. a. I found 4 false positives while going through the assignment.

False Positives (7):

('jure', 'e', 'server@cs.stanford.edu')

('serafim', 'e', 'serafim@cs dot stanford.edu')

gold: ('serafim', 'e', 'serafim@cs.stanford.edu')

gold: ('serafim', 'p', '650-723-3334')

gold: ('serafim', 'p', '650-725-1449')

('hager', 'e', 'hager@cs dot jhu.edu')

gold: ('hager', 'e', 'hager@cs.jhu.edu')

gold: ('hager', 'p', '410-516-5521')

gold: ('hager', 'p', '410-516-5553')

gold: ('hager', 'p', '410-516-8000')

('dlwh', 'e', 'd-l-w-h@s-t-a-n-f-o-r-d-.edu')

gold: ('dlwh', 'e', 'dlwh@stanford.edu')

('jks', 'e', 'jks@robotics;stanford;.edu')

gold: ('jks', 'e', 'jks@robotics.stanford.edu')

('plotkin', 'e', 'server@infolab.stanford.edu')

('subh', 'e', 'uma@cs dot stanford.edu')

gold: ('subh', 'e', 'subh@stanford.edu')

gold: ('subh', 'e', 'uma@cs.stanford.edu')

gold: ('subh', 'p', '650-724-1915')

gold: ('subh', 'p', '650-725-3726')

gold: ('subh', 'p', '650-725-6949')

JKS: jks at robotics;stanford;edu – I’m not sure why it happened, but I’m assuming it’s something to do with the semicolons.

Jure/Plotkin: Server at cs.stanford.edu – The extra period.

Hager/Serafim/Pal: hager at cs dot jhu dot edu - Too many extra strings/spaces.

Dlwh: [d-l-w-h-@-s-t-a-n-f-o-r-d-.-e-d-u](mailto:d-l-w-h-@-s-t-a-n-f-o-r-d-.-e-d-u) - Too many hyphens.

b. The simplest way to obfuscate an email would probably be to add a lot of extra characters, or encode it with HTML. Better yet, if this were provided on a website, you could use a contact form that requires captcha to use.

I found an HTML encoder that changes your email to HTML entities:

Obfuscate an email: [ajohn152@syr.edu](mailto:ajohn152@syr.edu) 🡪 &#097;&#106;&#111;&#104;&#110;&#049;&#053;&#050;&#064;&#115;&#121;&#114;&#046;&#101;&#100;&#117;

Certainly not fool-proof, but I’m sure it would work against basic spambots.

Alternatively, you could URL encode with HTML: <a href="mailto:%61%6a%6f%68%6e%31%35%32%40%73%79%72%2e%65%64%75">email</a>

Source: <https://spencermortensen.com/articles/email-obfuscation/#link-url>