

Alphabetic only texts/strings

For the core parts of the Hybrid Cryptography project you are expected to work with text strings that only contain alphabetic characters, i.e. no spaces, no numbers, no punctuation, just the letters of the alphabet.

You will therefore need to write a function that extracts precisely the alphabetic content of a string. Let's suppose that you have done this and that you have defined the following function

```
def extract_text(text):  
    # Body of the function which takes the string  
    # text and returns a string containing only the  
    # alphabetic content of text. For you to define...
```

Now as an example we will define the string `first_part` to be the string containing the first 4000 characters of the string `austen_text` whose content is the book *Pride and Prejudice* by Jane Austen. (4000 characters chosen so that we have started the actual book or, at least, the preface!) To do this we simply repeat the work done at the beginning of the notebook `get_online_texts.ipynb` where we noted that we could find the *utf-8* encoded version of *Pride and Prejudice* by Jane Austen here:

<https://www.gutenberg.org/files/1342/1342-0.txt>

and where we downloaded this as a string using the function `url_to_text_utf8` defined below.

```
In [4]: import requests, os  
  
In [5]: def url_to_text_utf8(url):  
        '''  
        Given a url for a text that is  
        'utf-8' encoded this function  
        returns that text.  
        '''  
        response = requests.get(url)  
        response.encoding = 'utf-8-sig'  
        return response.text
```

So now let's get the book *Pride and Prejudice* (as a string) once again.

```
In [6]: austen_text = url_to_text_utf8("https://www.gutenberg.org/files/1342/1342-0.txt")
```

And continue by defining `first_part`.

```
In [7]: first_part = austen_text[:4000]
```

Now let's have a look at this string. Notice all the space and other junk...

```
In [8]: first_part
```

[illegible]

The Point

You can convert all of the string `austen_text` in this way to obtain a purely alphabetic string then use it for encryption and decryption etc. (and/or save it in a file for use later on). You can (and should) of course follow the same process for any of the text (.txt) files on the Hybrid Cryptography GitHub page, or for any other text files that you choose to use.

Note on cases

You can design your encryption and decryption functions so that they preserve case. However you can also decide to work only in upper case or lower case (although preservation of case would be a nice feature to have). To do this you would simply have to convert your string to lower or upper case.

For example executing the line

```
alphabetic_first_part.lower()
```

returns the string:

```
'theprojectgutenbergebookofprideandprejudicebyjaneaustenthisebookisfortheuseofanyoneanywherei
ntheunitedstatesandmostotherpartsoftheworldatnocostandwithalmostnorestrictionswhatsoeveryouma
ycopyitgiveitawayorreuseitunderthetermsoftheprojectgutenberglicenseincludedwiththisbookoronl
ineatwwwgutenbergorgifyouarenotlocatedintheunitedstatesyouwillhavetocheckthelawsofthecountryw
hereyouarelocatedbeforeusingthisbooktitleprideandprejudiceauthorjaneaustenreleasedatenovembe
rebookmostrecentlyupdatedmarchlanguageenglishproducedbychuckgreifandtheonlinedistributedproof
readingteamathhttpwwwpgdpnetthisfilewasproducedfromimagesavailableattheinternetarchivestartoft
heprojectgutenbergebookprideandprejudiceillustrationgeorgeallenpublishercharingcrossroadlondo
nruskinhouseillustrationreadingjanesletterschapprideandprejudicebyjaneaustenwithaprefacebygeo
rgesaintsburyandillustrationsbyhughthomsonillustrationruskincharinghousecrossroadlondongeorge
allenchiswickpresscharleswhittinghamandcotookscourtchancerylanelondonillustrationtojcomynscar
rinacknowledgmentofalliowetohisfriendshipandadvicetheseillustrationsaregratefullyinscribedhug
hthomsonprefaceillustrationwaltwhitmanhassomewhereafineandjustdistinctionbetweenlovingbyallow
anceandlovingwithpersonallovethisdistinctionappliestobooksaswellastomenandwomenandinthecaseof
thenotverynumerousauthorswhoaretheobjectsofthepersonalaffectionitbringsacuriousconsequencewit
hitthereismuchmoredifferenceastotheirbestworkthaninthecaseofthoseotherswhoarelovedbyallowance
byconventionandbecauseitisfelttobetherightandproperthingtolovethemandinthesectfairlylargeandy
etunusuallychoiceofausteniansorjanitestherewouldprobablybefoundpartisansoftheclaimtoprimacyof
almosteveryoneofthenovelistosomethedelightfulfreshnessandhumourofnorthangerabbeyitscompletenes
sfinishandentrainobscuretheundoubtedcriticalfactsthatitsscaleissmallanditsschemeafterallthato
fburlesqueorparodyakindinwhichthefirstrankisreachedwithdifficultypersuasionrelativelyfaintint
oneandnotenthrallingininteresthasdevoteeswhoexaltabovealltheothersitsexquisite delicacyandkeep
ingthecatastropheofmansfieldparkisadmittedlytheatricaltheheroandheroineareinsipidandtheauthor
hasalmostwickedlydestroyedallromanticinterestbyexpresslyadmittingthatedmundonlytookfannybecau
se'
```

whereas executing the line

```
alphabetic_first_part.upper()
```

returns the string

```
'THEPROJECTGUTENBERGEBOOKOFPRIDEANDPREJUDICEBYJANEAUSTENTHISEBOOKISFORTHEUSEOFANYONEANYWHEREI
NTHEUNITEDSTATESANDMOSTOTHERPARTSOFTHEWORLDATNOCOSTANDWITHALMOSTNORESTRICTIONSWHATSOEVERYOUA
YCOPYITGIVEITAWAYORREUSEITUNDERTHETERMSTHEPROJECTGUTENBERGLICENSEINCLUDEDWITHTHISBOOKORONL
INEATWWWGUTENBERGORGIFYOARENOTLOCATEDINTHEUNITEDSTATESYOUWILLHAVETOCHECKTHELAWSOFTHECOUNTRYW
HEREYOUARELOCATEDBEFOREUSINGTHISBOOKTITLEPRIDEANDPREJUDICEAUTHORJANEAUSTENRELEASEDATENOVEMBE
REBOOKMOSTRECENTLYUPDATEDMARCHLANGUAGEENGLISHPRODUCEDBYCHUCKGREIFANDTHEONLINEDISTRIBUTEDPROOF
READINGTEAMATHHTTPWWWPGDPNETTHISFILEWASPRODUCEDFROMIMAGESAVAILABLEATTHEINTERNETARCHIVESTARTOFT
HEPROJECTGUTENBERGEBOOKPRIDEANDPREJUDICEILLUSTRATIONGEORGEALLENPUBLISHERCHARINGCROSSROADLONDO
NRUSKINHOUSEILLUSTRATIONREADINGJANESLETTERSCHAPPRIEANDPREJUDICEBYJANEAUSTENWITHAPREFACEBYGEO
RGESAINTSBURYANDILLUSTRATIONSBYHUGHTHOMSONILLUSTRATIONRUSKINCHARINGHOUSECROSSROADLONDONGEORGE
ALLENCHISWICKPRESSCHARLESWHITTINGHAMANDCOTOOKSCOURTCHANCERYLANELONDONILLUSTRATIONTOJCOMYNSCAR
RINACKNOWLEDGMENTOFALLIOWETOHISFRIENDSHIPANDADVICETHSEILLUSTRATIONSAREGRATEFULLYINSCRIBEDHUG
HTHOMSONPREFACEILLUSTRATIONWALTWHITMANHASSOMEWHERAEAFINEANDJUSTDISTINCTIONBETWEENLOVINGBYALLOW
ANCEANDLOVINGWITHPERSONALLOVETHISDISTINCTIONAPPLIESTOBOOKSASWELLASTOMENANDWOMENANDINTHECASEOF
THENOTVERYNUMEROUSAUTHORSWHOARETHEOBJECTSOFTHEPERSONALAFFECTIONITBRINGSACURIOUSCONSEQUENCEWIT
HITHEREISMUCHMOREDIFFERENCEASTOTHEIRBESTWORKTHANINTHECASEOFTHOSEOTHERSWHOARELOVEDBYALLOWANCE
BYCONVENTIONANDBECAUSEITISFELTTOBETHERIGHTANDPROPERTHINGTOLOVETHEMANDINTHESECTFAIRLYLARGEANDY
ETUNUSUALLYCHOICEOFAUSTENIANSORJANITESTHEREWOULDPROBABLYBEFOUNDPARTISANSOFTHECLAIMTOPRIMACYOF
ALMOSTEVERYONEOFTHENOVELSTOSOMETHEDELIGHTFULFRESHNESSANDHUMOUROFNORTHANGERABBEYITSCOMPLETENES
SFINISHANDENTRAINOBSCURETHEUNDOUBTEDCRITICALFACTSTHATITSSCALEISSMALLANDITSSCHEMEAFTERALLTHATO
FBURLESQUEORPARODYAKINDINWHICHTHEFIRSTRANKISREACHEDWITHDIFFICULTYPERSUASIONRELATIVELYFAINTINT
ONEANDNOTENTHRALLINGININTERESTHASDEVOTEESWHOEXALTABOVEALLTHEOTHERSITSEXQUISITEDELICACYANDKEEP
INGTHECATASTROPHEOFMANSFIELDPARKISADMITTEDLYTHEATRICALTHEHEROANDHEROINEAREINSIPIDANDTHEAUTHOR
HASALMOSTWICKEDLYDESTROYEDALLROMANTICINTERESTBYEXPRESSLYADMITTINGTHATEDMUNDONLYTOOKFANNYBECAU
SE'
```