

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]

Extract the alphabetic content of the string `first_part`

To extract the alphabetic content (i.e. alphabetic characters) from the string `first_part` we pass it to the function `extract_text`. Using the following line of code we assign the variable `alphabetic_first_part` to the resulting string.

```
alphabetic_first_part = extract_text(first_part)
```

And now, when we inspect the string returned by `extract_text` by, for example, running a cell containing only the line...

alphabetic first part

we see — provided `extract_text` is correctly defined — that this string is as follows:

'TheProjectGutenbergBookofPrideandprejudicebyJaneAustenThisBookisfortheuseofanyoneanywher
ntheUnitedStatesandmostotherpartsoftheworldatnocostandwithalmostnorestrictionswhatsoeverYoua
ycopyitgiveitawayorreuseitunderthetermsoftheProjectGutenbergLicenseincludedwiththisBookor
ineatwww.gutenberg.orgIfyouarenotlocatedintheUnitedStatesyouwillhavetochekthelawsofthecountr
yhereyouarelocatedbeforeusingthisBookTitlePrideandprejudiceAuthorJaneAustenReleaseDateNovemb
reBookMostrecentlyupdatedMarchLanguageEnglishProducedbyChuckGreifandtheOnlineDistributedProo
freadingTeamathttp://www.pgdp.netThisfilewasproducedfromimagesavailableattheInternetArchiveSTARTOFT
HEPROJECTGUTENBERGBOOKPRIDEANDPREJUDICEIllustrationGEORGEALLENPUBLISHERCHARINGCROSSROADLOND
NRUSKINHOUSEIllustrationReadingJanesLettersChapPRIDEandPREJUDICEByJaneAustenwithaPrefacebyGe
rgeSaintsburyandIllustrationsbyHughThomsonIllustrationRuskinCharingHouseCrossRoadLondonGeorge
AllenCHISWICKPRESSCHARLESWHITTINGHAMANDCOTOKSCOURTCHANCERYLANELONDONIllustrationToJComynsCar
rinacknowledgmentofallIowetothisfriendshipandadviceheseillustrationsaregratefullyinscribedHug
hThomsonPREFACEIllustrationWaltWhitmanhassomewhereafineandjustdistinctionbetweenlovingbyallow
anceandlovingwithpersonalloveThisdistinctionappliestobooksaswellastomenandwomenandinthecaseof
thenotverynumerousauthorswhoaretheobjectsofthepersonalaffectionitbringsacuriousconsequencewit
hitThereismuchmoredifferenceastotheirbestworkthaninthecaseofthoseotherswhoarelovedbyallowance
byconventionandbecauseitisfelttobetherightandproperthingtolovethemAndinthesectfairlylargeand
etunusuallychoiceofAusteniansorJanites therewouldprobablybefoundpartisansofthecleitoprimaryof
almosteveryoneofthenovelsTosomethedelightfulfreshnessandhumourofNorthangerAbbeyitscompleteness
finishandentertainobscuretheundoubtedcriticalfactsthatitsscaleissmallanditsschemeafterallthato
fburlesqueorparodyakininwhichthefirstrankisreachedwithdifficultyPersuasionrelativelyfaintint
oneandnotinthrallingininteresthasdevoteeswhoexaltabovealltheothersitsexquisite delicacyandkeep
ingThecatastropheofMansfieldParkisadmittedlytheatricaltheheroandheroineareinsipidandtheauthor
hasalmostwickedlydestroyedallromanticinterestbyexpresslyadmittingthatEdmundonlytookFannybecau
se'

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:

```
'theprojectgutenbergbookofprideandprejudicebyjaneaustenthisebookisfortheuseofanyoneanywherei
ntheunitedstatesandmostotherpartsoftheworldatnocostandwithalmostnorestrictionswhatsoeveryouma
ycopyitgiveitawayorreuseitunderthetermsoftheprojectgutenberglicenseincludedwiththisbookoronl
ineatwwwgutenbergorgifyouarenotlocatedintheunitedstatesyouwillhavetocheckthelawsofthecountryw
hereyouarelocatedbeforeusingthisbooktitleprideandprejudiceauthorjaneaustenreleasedatenovembe
rebookmostrecentlyupdatedmarchlanguageenglishproducedbychuckgreifandtheonlinedistributedproof
readingteamathhttpwwwpgdpnetthisfilewasproducedfromimagesavailableattheinternetarchivestartoft
heprojectgutenbergbookprideandprejudiceillustrationgeorgeallenpublishercharingcrossroadlondo
nruskinhouseillustrationreadingjanesletterschapprideandprejudicebyjaneaustenwithaprefacebygeo
rgesaintsburyandillustrationsbyhughthomsonillustrationruskincharinghousecrossroadlondongeorge
allenchiswickpresscharleswhittinghamandcotookscourtchancerylanelondonillustrationtojcomynscar
rinacknowledgmentofalliowetohisfriendshipandadviceheseillustrationsaregratefullyinscribedhug
hthomsonprefaceillustrationwaltwhitmanhassomewhereafineandjustdistinctionbetweenlovingbyallow
anceandlovingwithpersonalllovethisdistinctionappliestobooksaswellastomenandwomenandinthecaseof
thenotverynumerousauthorswhoaretheobjectsofthepersonalaffectionitbringsacuriousconsequencewit
hitthereismuchmoredifferenceastotheirbestworkthaninthecaseofthoseotherswhoarelovedbyallowanc
ebyconventionandbecauseitisfelttobetherightandproperthingtolovethemandinthesectfairlylargeandy
etunusuallychoiceofausteniansorjanitestherewouldprobablybefoundpartisansofthecclaimtoprimacyof
almosteveryoneofthenovelstosomethedelightfulfreshnessandhumourofnorthangerabbeyitscompletenes
sfinishandentrainobscuretheundoubtedcriticalfactsthatitsscaleissmallanditsschemeafterallthato
fburlesqueorparodyakindinwhichthefirstrankisreachedwithdifficultypersuasionrelativelyfaintint
oneandnotenthrallingininteresthasdevoteeswhoexaltabovealltheothersitsexquisite delicacyandkeep
ingthecatastropheofmansfieldparkisadmittedlytheatricaltheheroandheroineareinsipidandtheauthor
hasalmostwickedlydestroyedallromanticinterestbyexpresslyadmittingthatedmundonlytookfannybecau
se'
```

whereas executing the line

```
alphabetic_first_part.upper()
```

returns the string

```
'THEPROJECTGUTENBERGBOOKOFPRIDEANDPREJUDICEBYJANEAUSTENTHISEBOOKISFORTHEUSEOFANYONEANYWHEREI
NTHEUNITEDSTATESANDMOSTOTHERPARTSOFTHEWORLDATNOCOSTANDWITHALMOSTNORESTRICTIONSWHATSOEVERYOUA
YCOPYITGIVEITAWAYORREUSEITUNDERTHE TERMS OF THE PROJECT GUTENBERG LICENSE INCLUDED WITH THIS BOOK OR ONL
INE AT WWW GUTENBERG ORG IF YOU ARE NOT LOCATED IN THE UNITED STATES YOU WILL HAVE TO CHECK THE LAWS OF THE COUNTRY W
HERE YOU ARE LOCATED BEFORE USING THIS BOOK TITLE PRIDE AND PREJUDICE AUTHOR JANE AUSTEN RELEASED ATENOVEMBE
RE BOOK MOST RECENTLY UPDATED MARCH LANGUAGE ENGLISH PRODUCED BY CHUCK GREIF AND THE ONLINE DISTRIBUTED PROOF
READING TEAM AT HTTP WWW PGDP NET THIS FILE WAS PRODUCED FROM IMAGES AVAILABLE AT THE INTERNET ARCHIVE START OF T
HE PROJECT GUTENBERG BOOK PRIDE AND PREJUDICE ILLUSTRATION GEORGE ALLEN PUBLISHER CHARING CROSS ROAD LOND O
NRUSKIN HOUSE ILLUSTRATION READING JANE S LETTERS CHAP PRIDE AND PREJUDICE BY JANE AUSTEN WITH A PREFACE BY GEO
RGESAINTSBURY AND ILLUSTRATIONS BY HUGH THOMSON ILLUSTRATION RUSKIN CHARING HOUSE CROSS ROAD LONDON GEORGE
ALLEN CHISWICK PRESS CHARLES WHITTINGHAM AND COTOOKS COURT CHANCERY LANE LONDON ILLUSTRATION TO J COMYN SCAR
RIN ACKNOWLEDGMENT OF ALLIOWETO HIS FRIENDSHIP AND ADVICE THESE ILLUSTRATIONS ARE GRATEFULLY INSCRIBED HUG
H THOMSON PREFACE ILLUSTRATION WALT WHITMAN HAS SOMEWHERE A FINE AND JUST DISTINCTION BETWEEN LOVING BY ALLOW
ANCE AND LOVING WITH PERSONAL LOVE THIS DISTINCTION APPLIES TO BOOKS AS WELL AS TO MEN AND WOMEN AND IN THE CASE OF
THE NOT VERY NUMEROUS AUTHORS WHO ARE THE OBJECTS OF THE PERSONAL AFFECTION IT BRINGS A CURIOUS CONSEQUENCE WIT
H IT THERE IS MUCH MORE DIFFERENCE AS TO THEIR BEST WORK THAN IN THE CASE OF THOSE OTHERS WHO ARE LOVED BY ALLOWANC
E BY CONVENTION AND BECAUSE IT IS FELT TO BE THE RIGHT AND PROPER THING TO LOVE THEM AND IN THE SECT FAIRLY LARGE AND Y
ET UNUSUALLY CHOICE OF AUSTENIANS OR JANITES THERE WOULD PROBABLY BE FOUND PARTISANS OF THE CLAIM TO PRIMACY OF
ALMOST EVERY ONE OF THE NOVELS TO SOME THE DELIGHTFUL FRESHNESS AND HUMOUR OF NORTHANGER ABBEY ITS COMPLETENES
S FINISH AND ENTRAIN OBSCURE THE UNDOUBTED CRITICAL FACTS THAT ITS SCALE IS SMALL AND ITS SCHEME AFTER ALL THATO
FBURLESQUE OR PARODY A KIND IN WHICH THE FIRST RANK IS REACHED WITH DIFFICULTY PERSUASION RELATIVELY FAINT INT
ONE AND NOT ENTHRALLING IN INTEREST HAS DEVOTEES WHO EXALT ABOVE ALL THE OTHERS ITSEXQUISITE DELICACY AND KEEP
ING THE CATASTROPHE OF MANSFIELD PARK IS ADMITTEDLY THEATRICAL THE HERO AND HEROINE ARE INSIPID AND THE AUTHOR
HAS ALMOST WICKEDLY DESTROYED ALL ROMANTIC INTEREST BY EXPRESSLY ADMITTING THAT EDMUND ONLY TOOK FANNY BECAU
SE '
```