This Jupyter Notebook is proof of the concept of using AI in document summarization.

It is an early-stage example of how AI can be used to summaries a parsed section of an RPPR.

The example used has its title page attached. The entire contents of the B1 section were summarized down to one sentence that accurately expressed the section's content, which initially took 659 words to express.

This method provides a means of substantially increasing the productivity of the analysis of documents and can allow POs to focus their efforts elsewhere on perhaps more value-creating endeavors for DAIDS, such as business management efforts in conjunction with our researcher to help them get across the Valley of Death which hinders the translation of early-stage technology.

Build an AI Text Summarizer in Python (see attached)

A. OVERALL COVER PAGE

Project Title: Griffithsin-based Rectal Microbicides for PREvention of Viral ENTry (PREVENT)	
Grant Number: 5U19AI113182-05	**Project/Grant Period**: 07/01/2014 - 06/30/2019
Reporting Period: 07/01/2017 - 06/30/2018	**Requested Budget Period**: 07/01/2018 - 06/30/2019
Report Term Frequency: Annual	**Date Submitted**: 05/01/2018
Program Director/Principal Investigator Information: KENNETH E PALMER , BOTH PHD **Phone number**: 270-691-5960 **Email**: kepalm02@louisville.edu	**Recipient Organization**: UNIVERSITY OF LOUISVILLE UNIVERSITY OF LOUISVILLE Sponsored Programs Administration LOUISVILLE, KY 402920001 **DUNS**: 057588857 **EIN**: 1611029626A1 **RECIPIENT ID**: OGMB140588
Change of Contact PD/PI: N/A	
Administrative Official: RYANN M GILMORE 300 E. Market Street Suite 300 Louisville, KY 40202 **Phone number**: 502-852-3788 **Email**: rmgilm01@louisville.edu	**Signing Official**: BARBARA SELLS Sponsored Prog. Grants Admin Stevenson Hall Room 503 University of Louisville LOUISVILLE, KY 40292 **Phone number**: 5028528367 **Email**: bfsell01@louisville.edu

In [1]:

```
!pip install spacy
!python -m spacy download en_core_web_sm
```

```
c:\users\houzeta\.conda\envs\nlpcours
e\lib\site-packages (from spacy) (1.
8.2)
Requirement already satisfied: srsly<
3.0.0,>=2.4.3 in c:\users\houzeta\.co
nda\envs\nlpcourse\lib\site-packages
(from spacy) (2.4.3)
Requirement already satisfied: typing
-extensions<4.0.0.0,>=3.7.4 in c:\use
rs\houzeta\.conda\envs\nlpcourse\lib
\site-packages (from spacy) (3.10.0.
2)
Requirement already satisfied: thinc<
8.1.0,>=8.0.14 in c:\users\houzeta\.c
onda\envs\nlpcourse\lib\site-packages
(from spacy) (8.0.15)
Requirement already satisfied: blis<
0.8.0,>=0.4.0 in c:\users\houzeta\.co
nda\envs\nlpcourse\lib\site-packages
(from spacy) (0.7.7)
```

In [2]:

```
import spacy

nlp = spacy.load("en_core_web_sm")
```

```
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorflow\python\frame
work\dtypes.py:516: FutureWarning: Passi
ng (type, 1) or '1type' as a synonym of
type is deprecated; in a future version
of numpy, it will be understood as (typ
e, (1,)) / '(1,)type'.
  _np_qint8 = np.dtype([("qint8", np.int
8, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorflow\python\frame
work\dtypes.py:517: FutureWarning: Passi
ng (type, 1) or '1type' as a synonym of
type is deprecated; in a future version
of numpy, it will be understood as (typ
e, (1,)) / '(1,)type'.
  _np_quint8 = np.dtype([("quint8", np.u
int8, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorflow\python\frame
work\dtypes.py:518: FutureWarning: Passi
ng (type, 1) or '1type' as a synonym of
type is deprecated; in a future version
of numpy, it will be understood as (typ
e, (1,)) / '(1,)type'.
  _np_qint16 = np.dtype([("qint16", np.i
```

```
nt16, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorflow\python\frame
work\dtypes.py:519: FutureWarning: Passi
ng (type, 1) or '1type' as a synonym of
type is deprecated; in a future version
of numpy, it will be understood as (typ
e, (1,)) / '(1,)type'.
  _np_quint16 = np.dtype([("quint16", n
p.uint16, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorflow\python\frame
work\dtypes.py:520: FutureWarning: Passi
ng (type, 1) or '1type' as a synonym of
type is deprecated; in a future version
of numpy, it will be understood as (typ
e, (1,)) / '(1,)type'.
  _np_qint32 = np.dtype([("qint32", np.i
nt32, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorflow\python\frame
work\dtypes.py:525: FutureWarning: Passi
ng (type, 1) or '1type' as a synonym of
type is deprecated; in a future version
of numpy, it will be understood as (typ
e, (1,)) / '(1,)type'.
  np_resource = np.dtype([("resource", n
p.ubyte, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorboard\compat\tens
orflow_stub\dtypes.py:541: FutureWarnin
g: Passing (type, 1) or '1type' as a syn
onym of type is deprecated; in a future
```

```
version of numpy, it will be understood
as (type, (1,)) / '(1,)type'.
  _np_qint8 = np.dtype([("qint8", np.int
8, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorboard\compat\tens
orflow_stub\dtypes.py:542: FutureWarnin
g: Passing (type, 1) or '1type' as a syn
onym of type is deprecated; in a future
version of numpy, it will be understood
as (type, (1,)) / '(1,)type'.
  _np_quint8 = np.dtype([("quint8", np.u
int8, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorboard\compat\tens
orflow_stub\dtypes.py:543: FutureWarnin
g: Passing (type, 1) or '1type' as a syn
onym of type is deprecated; in a future
version of numpy, it will be understood
as (type, (1,)) / '(1,)type'.
  _np_qint16 = np.dtype([("qint16", np.i
nt16, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorboard\compat\tens
orflow_stub\dtypes.py:544: FutureWarnin
g: Passing (type, 1) or '1type' as a syn
onym of type is deprecated; in a future
version of numpy, it will be understood
as (type, (1,)) / '(1,)type'.
  _np_quint16 = np.dtype([("quint16", n
p.uint16, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorboard\compat\tens
```

```
orflow_stub\dtypes.py:545: FutureWarnin
g: Passing (type, 1) or '1type' as a syn
onym of type is deprecated; in a future
version of numpy, it will be understood
as (type, (1,)) / '(1,)type'.
  _np_qint32 = np.dtype([("qint32", np.i
nt32, 1)])
C:\Users\houzeta\.conda\envs\nlpcourse\l
ib\site-packages\tensorboard\compat\tens
orflow_stub\dtypes.py:550: FutureWarnin
g: Passing (type, 1) or '1type' as a syn
onym of type is deprecated; in a future
version of numpy, it will be understood
as (type, (1,)) / '(1,)type'.
  np_resource = np.dtype([("resource", n
p.ubyte, 1)])
```

B.1 WHAT ARE THE MAJOR GOALS OF THE PROJECT?

(This section contains 659 words to express its' content)

In [3]:

```
# extractive summary by word count
text = """

The goals of the Critical Path Project (“CPP”; Project
Substance and Drug Product manufacture, manage regulat
integrate regulatory documents for submission of an IN
are designed to support a first-in-human NIH pre-Phase
novel active ingredient Q-GRFT (the PREVENT study). Th
Goal/Aim 1: Manufacture Q-Griffithsin (Q-GRFT) API and
This Aim comprises monitoring and assisting multiple t
comprehensive cGMP-compliant Chemistry, Manufacturing
submission to FDA. Aim 1 includes pre-GMP as well as c
Substance, or DS) and formulation of the DS in an appr
DP). These materials need to be produced in sufficient
studies (Aim 2 and Project 2), including regulation-co
female participants in the clinical study (Project 3).
microbicide from a gel to a solution of Q-GRFT. As suc
which has been produced under cGMP by its manufacturer
stabiity data for the new enema formulation will be av
Milestone 1 will have been met.
Goal/Aim 2: Conduct regulation-compliant preclinical s
This Aim entails providing design assistance, oversigh
contractors for the conduct of preclinical pharmacolog
trial. In compliance with regulatory statutes for prec
preclinical rectal irritation, sensitization and toler
regulation-compliant studies were completed during gra
Hence, Aim 2 of Project 1 is nearing completion.
Goal/Aim 3: Prepare and submit an IND application to s
The conduct of the proposed first-in-human clinical st
regulations and with approval from DAIDS. In Aim 3 of
```

```
generated in Aims 1 and 2 of Project 1 and all other c
reduce the risk of regulatory delays and to establish
submission of a pre-IND package to gather Agency input
Preparation of pre-IND and IND regulatory documents is
other projects and cores providing input to the Critic
guidelines.
The first goal of this Aim was submission of the pre-I
grant Year 2), with FDA's formal and complete response
development. The main regulatory Milestone of this Aim
grant Year 5). The Project is currently on track to me

"""
# tokenize
doc = nlp(text)
```

In [4]:

```
# create dictionary
word_dict = {}
# loop through every sentence and give it a weight
for word in doc:
    word = word.text.lower()
    if word in word_dict:
        word_dict[word] += 1
    else:
        word_dict[word] = 1
```

In [5]:

```
# create a list of tuple (sentence text, score, index)
sents = []
# score sentences
sent_score = 0
for index, sent in enumerate(doc.sents):
    for word in sent:
        word = word.text.lower()
        sent_score += word_dict[word]
    sents.append((sent.text.replace("\n", " "), sent_s
#Sorting the Sentences
```

In [6]:

```
# sort sentence by word occurrences
sents = sorted(sents, key=lambda x: -x[1])
# return top 3
sents = sorted(sents[:3], key=lambda x: x[2])
```

The final cell of this Jupyter Notebook below prints out the summary of the B1 Section of the RPPR in just a single sentence.

In [7]:

```
# compile them into text
summary_text = ""
for sent in sents:
    summary_text += sent[0] + " "

print(summary_text)
```

```
The IND modules 1-5 are currently in dev
elopment. The Project is currently on tr
ack to meet this Aim's Milestone.
```

The 19 word summary below accurately replaces the 659 words used to form the B1 Section of the attached RPPR.

"The IND modules 1-5 are currently in development. The Project is currently on track to meet this Aim's Milestone."