

pdf_reader

October 1, 2023

1 PDF Reader Chatbot

This project is based on Codebasics video: <https://www.youtube.com/watch?v=MoqgmWV1fm8>

1.1 Importing Libraries

```
[2]: from openai_key import secret_key
import langchain
from langchain import OpenAI
from langchain.chains import RetrievalQAWithSourcesChain
from langchain.chains.qa_with_sources.loading import load_qa_with_sources_chain
from langchain.text_splitter import RecursiveCharacterTextSplitter
from langchain.document_loaders import PyPDFLoader
from langchain.embeddings import OpenAIEmbeddings
from langchain.vectorstores import FAISS
import pickle
```

1.2 Loading Document

```
[4]: loader = PyPDFLoader('Flavio_Brienza_Abstract_Tesi.pdf')

data = loader.load()

data
```

```
[4]: [Document(page_content='DISSERTATION ABSTRACT \n(ENGLISH AND ITALIAN VERSIONS)
\nThe Digital Economy and Society Index Progress of European Companies: The
\nApplication of Machine Learning in Financial Services \n \nEnglish Version
\nThe following thesis work has two main goals: analyzing the Digital Economy
and Society \nIndex (DESI) progress of European companies and providing a
practical application of \nmachine learning in the financial services sector.
\nAbout the first one, to the traditional DESI 's indicators , more general
macroeconomic and \nsocial traits have been added in order to have a better
framework of the current situation \nand to understand on the most problematic
aspects of the digitalization process. \nThe second part is focused on the
building of a machine learning model to predict the \nfinancial risk bearable
from banks' clients to offer them the proper stocks investment. Before \ndoing
this , the shares of 40 (+1, the gold) different companies have been analyzed
```

and \nclustered using both statistic met hods and Natural Language Proc essing of the latest news \nabout them. \nIn both phases Python programming language has been used. \n \nItalian Version \nIl seguente lav oro di tesi h a due obiettivi principa li: analizzare il D igital Economy and \nSociety Ind ex (DESI) delle aziende europee e fornire un esempio pratico dell 'applicazione \ndel machine learning nel settore dei servizi finanziari. \nNella prima parte, oltr e ai tradizionali indicato ri del DESI, altr i elementi macroeconomici e \nsociali sono stati considerati per avere una rappresentazion e migliore del l'attuale situazione \neuropea per comprender e meglio le difficoltà legate alla digitalizzazione. \nLa seconda parte in vece è invece focalizzata sulla creazione di un modello di machine \nlearning in grado di predire il livello di rischio finan ziario sopportabile da i clienti di una \nbanca, in modo da poter offrire loro la miglior opzione di investimento in azioni . Prima di \nsviluppare questa sezione, le azioni di 40 compagnie (+1 , l'oro) sono state analizzate e \nclassificate usando sia metodi statistici che Natural Language Processing sulle ultime notizie \nriguardanti le aziende stesse. \nPer entramb e le parti è stato usato il linguaggio di programmazione P ython. ', metadata={'source': 'Flavio_Brienza_Abstract_Tesi.pdf', 'page': 0})]

1.3 Splitting Text

```
[8]: text_splitter = RecursiveCharacterTextSplitter(
    chunk_size=1000,
    chunk_overlap=200
)

docs = text_splitter.split_documents(data)
docs
```

```
[8]: [Document(page_content='DISSERTATION ABSTRACT \n(ENGLISH AND ITALIAN VERSIONS)
\nThe Digital Economy and Society Index Progress of European Companies: The
\nApplication of Machine Learning in Financial Services \n \nEnglish Version
\nThe following thesis work has t wo main goals: analyzing the Digital Economy
and Society \nIndex (DESI) progress of European companies and providing a
practical application of \nmachine learning in the financial servi ces sector.
\nAbout the first one, to the traditional DESI 's indicators , more general
macroeconomic and \nsocial traits have been added in order to have a better
framework of the current situation \nand to understand on the most problematic
aspects of the digitalization process. \nThe second part is focused on the
building of a machine learning model to predict the \nfinancial risk bearable
from banks' clients to offer them the proper stocks investment. Before \ndoing
this , the shares of 40 (+1, the gold) different companies have been analyzed
and', metadata={'source': 'Flavio_Brienza_Abstract_Tesi.pdf', 'page': 0}),
Document(page_content='financial risk bearable from banks' clients to offer
them the proper stocks investment. Before \ndoing this , the shares of 40 (+1,
the gold) different companies have been analyzed and \nclustered using both
```

statistic methods and Natural Language Processing of the latest news about them. In both phases Python programming language has been used.

Italian Version

Il seguente lavoro di tesi ha due obiettivi principali: analizzare il Digital Economy and Society Index (DESI) delle aziende europee e fornire un esempio pratico dell'applicazione del machine learning nel settore dei servizi finanziari.

Nella prima parte, oltre ai tradizionali indicatori del DESI, altri elementi macroeconomici e sociali sono stati considerati per avere una rappresentazione migliore dell'attuale situazione europea per comprendere e meglio le difficoltà legate alla digitalizzazione.

La seconda parte invece è focalizzata sulla creazione di un modello di machine learning in grado di predire il livello di rischio finanziario sopportabile da i clienti di una banca, in modo da poter offrire loro la miglior opzione di investimento in azioni. Prima di sviluppare questa sezione, le azioni di 40 compagnie (+1, l'oro) sono state analizzate e classificate usando sia metodi statistici che Natural Language Processing sulle ultime notizie riguardanti le aziende stesse.

Per entrambi le parti è stato usato il linguaggio di programmazione Python.

```
Document(page_content='europea per comprendere e meglio le difficoltà legate alla digitalizzazione. La seconda parte in vece è invece focalizzata sulla creazione di un modello di machine learning in grado di predire il livello di rischio finanziario sopportabile da i clienti di una banca, in modo da poter offrire loro la miglior opzione di investimento in azioni. Prima di sviluppare questa sezione, le azioni di 40 compagnie (+1, l'oro) sono state analizzate e classificate usando sia metodi statistici che Natural Language Processing sulle ultime notizie riguardanti le aziende stesse. Per entrambi le parti è stato usato il linguaggio di programmazione Python.', metadata={'source': 'Flavio_Brienza_Abstract_Tesi.pdf', 'page': 0})]
```

1.4 Embedding

OpenAI library will be used.

```
[9]: embeddings = OpenAIEmbeddings(openai_api_key=secret_key)

vector_index = FAISS.from_documents(docs, embeddings)
```

Storing the results

```
[10]: file_path="vectors.pkl"
with open(file_path, "wb") as f:
    pickle.dump(vector_index, f)
```

Calling them back

```
[11]: import os

[12]: if os.path.exists(file_path):
    with open(file_path, "rb") as f:
        vectorIndex = pickle.load(f)
```

1.5 Creating the Chain

```
[20]: llm = OpenAI(temperature=1, max_tokens=500, openai_api_key=secret_key)

chain = RetrievalQAWithSourcesChain.from_llm(llm=llm, retriever=vectorIndex.
↳as_retriever())

chain
```

```
[20]: RetrievalQAWithSourcesChain(combine_documents_chain=MapReduceDocumentsChain(llm_
chain=LLMChain(prompt=PromptTemplate(input_variables=['context', 'question'],
template='Use the following portion of a long document to see if any of the text
is relevant to answer the question. \nReturn any relevant text
verbatim.\n{context}\nQuestion: {question}\nRelevant text, if any:'),
llm=OpenAI(client=<class 'openai.api_resources.completion.Completion'>,
temperature=1.0, max_tokens=500,
openai_api_key='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',
openai_api_base='', openai_organization='', openai_proxy='')), reduce_documents_
chain=ReduceDocumentsChain(combine_documents_chain=StuffDocumentsChain(llm_chain
=LLMChain(prompt=PromptTemplate(input_variables=['summaries', 'question'],
template='Given the following extracted parts of a long document and a question,
create a final answer with references ("SOURCES"). \nIf you don\'t know the
answer, just say that you don\'t know. Don\'t try to make up an answer.\nALWAYS
return a "SOURCES" part in your answer.\n\nQUESTION: Which state/country\'s law
governs the interpretation of the contract?\n=====\nContent: This Agreement
is governed by English law and the parties submit to the exclusive jurisdiction
of the English courts in relation to any dispute (contractual or non-
contractual) concerning this Agreement save that either party may apply to any
court for an injunction or other relief to protect its Intellectual Property
Rights.\nSource: 28-pl\nContent: No Waiver. Failure or delay in exercising any
right or remedy under this Agreement shall not constitute a waiver of such (or
any other) right or remedy.\n\n11.7 Severability. The invalidity, illegality or
unenforceability of any term (or part of a term) of this Agreement shall not
affect the continuation in force of the remainder of the term (if any) and this
Agreement.\n\n11.8 No Agency. Except as expressly stated otherwise, nothing in
this Agreement shall create an agency, partnership or joint venture of any kind
between the parties.\n\n11.9 No Third-Party Beneficiaries.\nSource:
30-pl\nContent: (b) if Google believes, in good faith, that the Distributor has
violated or caused Google to violate any Anti-Bribery Laws (as defined in
Clause 8.5) or that such a violation is reasonably likely to occur,\nSource:
4-pl\n=====\nFINAL ANSWER: This Agreement is governed by English
law.\nSOURCES: 28-pl\n\nQUESTION: What did the president say about Michael
Jackson?\n=====\nContent: Madam Speaker, Madam Vice President, our First
Lady and Second Gentleman. Members of Congress and the Cabinet. Justices of the
Supreme Court. My fellow Americans. \n\nLast year COVID-19 kept us apart. This
year we are finally together again. \n\nTonight, we meet as Democrats
Republicans and Independents. But most importantly as Americans. \n\nWith a duty
```

to one another to the American people to the Constitution. \n\nAnd with an unwavering resolve that freedom will always triumph over tyranny. \n\nSix days ago, Russia's Vladimir Putin sought to shake the foundations of the free world thinking he could make it bend to his menacing ways. But he badly miscalculated. \n\nHe thought he could roll into Ukraine and the world would roll over. Instead he met a wall of strength he never imagined. \n\nHe met the Ukrainian people. \n\nFrom President Zelenskyy to every Ukrainian, their fearlessness, their courage, their determination, inspires the world. \n\nGroups of citizens blocking tanks with their bodies. Everyone from students to retirees teachers turned soldiers defending their homeland. \nSource: 0-pl\nContent: And we won't stop. \n\nWe have lost so much to COVID-19. Time with one another. And worst of all, so much loss of life. \n\nLet's use this moment to reset. Let's stop looking at COVID-19 as a partisan dividing line and see it for what it is: A God-awful disease. \n\nLet's stop seeing each other as enemies, and start seeing each other for who we really are: Fellow Americans. \n\nWe can't change how divided we've been. But we can change how we move forward-on COVID-19 and other issues we must face together. \n\nI recently visited the New York City Police Department days after the funerals of Officer Wilbert Mora and his partner, Officer Jason Rivera. \n\nThey were responding to a 9-1-1 call when a man shot and killed them with a stolen gun. \n\nOfficer Mora was 27 years old. \n\nOfficer Rivera was 22. \n\nBoth Dominican Americans who'd grown up on the same streets they later chose to patrol as police officers. \n\nI spoke with their families and told them that we are forever in debt for their sacrifice, and we will carry on their mission to restore the trust and safety every community deserves. \nSource: 24-pl\nContent: And a proud Ukrainian people, who have known 30 years of independence, have repeatedly shown that they will not tolerate anyone who tries to take their country backwards. \n\nTo all Americans, I will be honest with you, as I've always promised. A Russian dictator, invading a foreign country, has costs around the world. \n\nAnd I'm taking robust action to make sure the pain of our sanctions is targeted at Russia's economy. And I will use every tool at our disposal to protect American businesses and consumers. \n\nTonight, I can announce that the United States has worked with 30 other countries to release 60 Million barrels of oil from reserves around the world. \n\nAmerica will lead that effort, releasing 30 Million barrels from our own Strategic Petroleum Reserve. And we stand ready to do more if necessary, unified with our allies. \n\nThese steps will help blunt gas prices here at home. And I know the news about what's happening can seem alarming. \n\nBut I want you to know that we are going to be okay. \nSource: 5-pl\nContent: More support for patients and families. \n\nTo get there, I call on Congress to fund ARPA-H, the Advanced Research Projects Agency for Health. \n\nIt's based on DARPA-the Defense Department project that led to the Internet, GPS, and so much more. \n\nARPA-H will have a singular purpose-to drive breakthroughs in cancer, Alzheimer's, diabetes, and more. \n\nA unity agenda for the nation. \n\nWe can do this. \n\nMy fellow Americans-tonight, we have gathered in a sacred space-the citadel of our democracy. \n\nIn this Capitol, generation after generation, Americans have debated great questions amid great strife, and have done great things. \n\nWe have fought for freedom, expanded

```

liberty, defeated totalitarianism and terror. \n\nAnd built the strongest,
freest, and most prosperous nation the world has ever known. \n\nNow is the
hour. \n\nOur moment of responsibility. \n\nOur test of resolve and conscience,
of history itself. \n\nIt is in this moment that our character is formed. Our
purpose is found. Our future is forged. \n\nWell I know this nation.\nSource:
34-pl\n\n=====\nFINAL ANSWER: The president did not mention Michael
Jackson.\nSOURCES:\n\nQUESTION:
{question}\n\n=====\n{summaries}\n\n=====\nFINAL ANSWER:'),
llm=OpenAI(client=<class 'openai.api_resources.completion.Completion'>,
temperature=1.0, max_tokens=500,
openai_api_key='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',
openai_api_base='', openai_organization='', openai_proxy=''),
document_prompt=PromptTemplate(input_variables=['page_content', 'source'],
template='Content: {page_content}\nSource: {source}'),
document_variable_name='summaries')), document_variable_name='context'),
retriever=VectorStoreRetriever(tags=['FAISS'],
vectorstore=<langchain.vectorstores.faiss.FAISS object at 0x0000025D9EBC13F0>))

```

Asking a question

```

[21]: question = 'How many companies have been considered to create portfolios?'

langchain.debug = True

chain({'question':question}, return_only_outputs=True)

```

```
[chain/start] [1:chain:RetrievalQAWithSourcesChain]
```

Entering Chain run with input:

```

{
  "question": "How many companies have been considered to create portfolios?"
}

```

```
[chain/start] [1:chain:RetrievalQAWithSourcesChain >
```

3:chain:MapReduceDocumentsChain] Entering Chain run with input:

```
[inputs]
```

```
[chain/start] [1:chain:RetrievalQAWithSourcesChain >
```

3:chain:MapReduceDocumentsChain > 4:chain:LLMChain] Entering Chain run with

input:

```

{
  "input_list": [
    {
      "context": "financial risk bearable from banks' clients to offer them the
proper stocks investment. Before \ndoing this , the shares of 40 (+1, the gold)
different companies have been analyzed and \nclustered using both statistic met
hods and Natural Language Proc essing of the latest news \nabout them. \nIn

```

both phases Python programming language has been used. \n \nItalian Version
 \nIl seguente lavoro di tesi ha due obiettivi principali: analizzare il Digital Economy and Society Index (DESI) delle aziende europee e fornire un esempio pratico dell'applicazione \ndel machine learning nel settore dei servizi finanziari. \nNella prima parte, oltre ai tradizionali indicatori del DESI, altri elementi macroeconomici e \nsociali sono stati considerati per avere una rappresentazione migliore dell'attuale situazione \neuropea per comprendere meglio le difficoltà legate alla digitalizzazione. \nLa seconda parte in vece è invece focalizzata sulla creazione di un modello di machine",

"question": "How many companies have been considered to create portfolios?"

},
 {

"context": "europea per comprendere e meglio le difficoltà legate alla digitalizzazione. \nLa seconda parte in vece è invece focalizzata sulla creazione di un modello di machine \nlearning in grado di predire il livello di rischio finanziario sopportabile da i clienti di una \nbanca, in modo da poter offrire loro la miglior opzione di investimento in azioni . Prima di \nsviluppare questa sezione, le azioni di 40 compagnie (+1 , l'oro) sono state analizzate e \nclassificate usando sia metodi statistici che Natural Language Processing sulle ultime notizie \nriguardanti le aziende stesse. \nPer entrambi e le parti è stato usato il linguaggio di programmazione Python.",

"question": "How many companies have been considered to create portfolios?"

},
 {

"context": "DISSERTATION ABSTRACT \n(ENGLISH AND ITALIAN VERSIONS)
 \nThe Digital Economy and Society Index Progress of European Companies: The
 \nApplication of Machine Learning in Financial Services \n \nEnglish Version
 \nThe following thesis work has two main goals: analyzing the Digital Economy and Society \nIndex (DESI) progress of European companies and providing a practical application of \nmachine learning in the financial services sector. \nAbout the first one, to the traditional DESI 's indicators , more general macroeconomic and \nsocial traits have been added in order to have a better framework of the current situation \nand to understand on the most problematic aspects of the digitalization process. \nThe second part is focused on the building of a machine learning model to predict the \nfinancial risk bearable from banks' clients to offer them the proper stocks investment. Before \ndoing this , the shares of 40 (+1, the gold) different companies have been analyzed and",

"question": "How many companies have been considered to create portfolios?"

}
]
 }

```
[llm/start] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain > 4:chain:LLMChain > 5:llm:OpenAI] Entering LLM
run with input:
{
  "prompts": [
    "Use the following portion of a long document to see if any of the text is
    relevant to answer the question. \nReturn any relevant text verbatim.\nfinancial
    risk bearable from banks' clients to offer them the proper stocks investment.
    Before \ndoing this , the shares of 40 (+1, the gold) different companies have
    been analyzed and \nclustered using both statistic met hods and Natural Language
    Proc essing of the latest news \nabout them. \nIn both phases Python
    programming language has been used. \n\nItalian Version \nIl seguente lav oro
    di tesi h a due obiettivi principa li: analizzare il D igital Economy and
    \nSociety Ind ex (DESI) delle aziende europee e fornire un esempio pratico dell
    'applicazione \ndel machine learning nel settore dei servizi finanziari.
    \nNella prima parte, oltr e ai tradizionali indicato ri del DESI, altr i
    elementi macroeconomici e \nsociali sono stati considerati per avere una
    rappresentazion e migliore del l'attuale situazione \neuropea per comprender e
    meglio le difficoltà legate alla digitalizzazione. \nLa seconda parte in vece è
    invece focalizzata sulla creazione di un modello di machine\nQuestion: How many
    companies have been considered to create portfolios?\nRelevant text, if any:"
  ]
}
[llm/start] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain > 4:chain:LLMChain > 6:llm:OpenAI] Entering LLM
run with input:
{
  "prompts": [
    "Use the following portion of a long document to see if any of the text is
    relevant to answer the question. \nReturn any relevant text verbatim.\neuropea
    per comprendere e meglio le difficoltà legate alla digitalizzazione. \nLa
    seconda parte in vece è invece focalizzata sulla creazione di un modello di
    machine \nlearning in grado di predire il livello di rischio finan ziario
    supportabile da i clienti di una \nbanca, in modo da poter offrire loro la
    miglior opzione di investimento in azioni . Prima di \nsviluppare questa
    sezione, le azioni di 40 compagnie (+1 , l'oro) sono state analizzate e
    \nclassificate usando sia metodi statistici che Natural Language Processing
    sulle ultime notizie \nriguardanti le aziende stesse. \nPer entramb e le parti
    è stato usato il linguaggio di programmazione P ython.\nQuestion: How many
    companies have been considered to create portfolios?\nRelevant text, if any:"
  ]
}
}
```



```
[llm/start] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain > 4:chain:LLMChain > 7:llm:OpenAI] Entering LLM
run with input:
```

```
{
  "prompts": [
    "Use the following portion of a long document to see if any of the text is
    relevant to answer the question. \nReturn any relevant text
    verbatim.\nDISSERTATION ABSTRACT \n(ENGLISH AND ITALIAN VERSIONS) \n
    The Digital Economy and Society Index Progress of European Companies: The
    \nApplication of Machine Learning in Financial Services \n \nEnglish Version
    \n
    The following thesis work has two main goals: analyzing the Digital Economy
    and Society \nIndex (DESI) progress of European companies and providing a
    practical application of \nmachine learning in the financial services sector.
    \n
    About the first one, to the traditional DESI's indicators, more general
    macroeconomic and \nsocial traits have been added in order to have a better
    framework of the current situation \nand to understand on the most problematic
    aspects of the digitalization process. \n
    The second part is focused on the building of a machine learning model to
    predict the \nfinancial risk bearable from banks' clients to offer them the
    proper stocks investment. Before \ndoing this, the shares of 40 (+1, the gold)
    different companies have been analyzed and\n
    Question: How many companies have been considered to create portfolios?\n
    Relevant text, if any:"
  ]
}
```

```
[llm/end] [1:chain:RetrievalQAWithSourcesChain >
```

```
3:chain:MapReduceDocumentsChain > 4:chain:LLMChain > 5:llm:OpenAI] [1.70s]
Exiting LLM run with output:
```

```
{
  "generations": [
    [
      {
        "text": " \nShares of 40 (+1, the gold) different companies have been
analyzed and clustered.",
        "generation_info": {
          "finish_reason": "stop",
          "logprobs": null
        }
      }
    ]
  ],
  "llm_output": {
    "token_usage": {
      "completion_tokens": 106,
      "total_tokens": 1012,
```

```

        "prompt_tokens": 906
    },
    "model_name": "text-davinci-003"
},
"run": null
}
[llm/end] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain > 4:chain:LLMChain > 6:llm:OpenAI] [1.70s]
Exiting LLM run with output:
{
  "generations": [
    [
      {
        "text": " \"Prima di sviluppare questa sezione, le azioni di 40
compagnie (+1 , l'oro) sono state analizzate e classificate usando sia metodi
statistici che Natural Language Processing sulle ultime notizie riguardanti le
aziende stesse.\",
        "generation_info": {
          "finish_reason": "stop",
          "logprobs": null
        }
      }
    ]
  ],
  "llm_output": {
    "token_usage": {},
    "model_name": "text-davinci-003"
  },
  "run": null
}
[llm/end] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain > 4:chain:LLMChain > 7:llm:OpenAI] [1.70s]
Exiting LLM run with output:
{
  "generations": [
    [
      {
        "text": "\n\nThe shares of 40 (+1, the gold) different companies have been
analyzed and",
        "generation_info": {
          "finish_reason": "stop",
          "logprobs": null
        }
      }
    ]
  ]
}

```

```

    ],
    "llm_output": {
        "token_usage": {},
        "model_name": "text-davinci-003"
    },
    "run": null
}
[chain/end] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain > 4:chain:LLMChain] [1.71s] Exiting Chain run
with output:
{
  "outputs": [
    {
      "text": " \nShares of 40 (+1, the gold) different companies have been
analyzed and clustered."
    },
    {
      "text": " \n\"Prima di sviluppare questa sezione, le azioni di 40 compagnie
(+1 , l'oro) sono state analizzate e classificate usando sia metodi statistici
che Natural Language Processing sulle ultime notizie riguardanti le aziende
stesse.\"\"
    },
    {
      "text": " \n\"The shares of 40 (+1, the gold) different companies have been
analyzed and\"
    }
  ]
}
[chain/start] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain > 8:chain:LLMChain] Entering Chain run with
input:
{
  "question": "How many companies have been considered to create portfolios?",
  "summaries": "Content: \nShares of 40 (+1, the gold) different companies have
been analyzed and clustered.\nSource:
Flavio_Brienza_Abstract_Tesi.pdf\n\nContent: \n\"Prima di sviluppare questa
sezione, le azioni di 40 compagnie (+1 , l'oro) sono state analizzate e
classificate usando sia metodi statistici che Natural Language Processing sulle
ultime notizie riguardanti le aziende stesse.\"\" \nSource:
Flavio_Brienza_Abstract_Tesi.pdf\n\nContent: \n\"The shares of 40 (+1, the gold)
different companies have been analyzed and\nSource:
Flavio_Brienza_Abstract_Tesi.pdf\"
}

```

```
[llm/start] [1:chain:RetrievalQAWithSourcesChain >
```

```
3:chain:MapReduceDocumentsChain > 8:chain:LLMChain > 9:llm:OpenAI] Entering LLM  
run with input:
```

```
{  
  "prompts": [  
    "Given the following extracted parts of a long document and a question,  
    create a final answer with references (\"SOURCES\"). \nIf you don't know the  
    answer, just say that you don't know. Don't try to make up an answer.\nALWAYS  
    return a \"SOURCES\" part in your answer.\n\nQUESTION: Which state/country's law  
    governs the interpretation of the contract?\n=====\nContent: This Agreement  
    is governed by English law and the parties submit to the exclusive jurisdiction  
    of the English courts in relation to any dispute (contractual or non-  
    contractual) concerning this Agreement save that either party may apply to any  
    court for an injunction or other relief to protect its Intellectual Property  
    Rights.\nSource: 28-pl\nContent: No Waiver. Failure or delay in exercising any  
    right or remedy under this Agreement shall not constitute a waiver of such (or  
    any other) right or remedy.\n\n11.7 Severability. The invalidity, illegality or  
    unenforceability of any term (or part of a term) of this Agreement shall not  
    affect the continuation in force of the remainder of the term (if any) and this  
    Agreement.\n\n11.8 No Agency. Except as expressly stated otherwise, nothing in  
    this Agreement shall create an agency, partnership or joint venture of any kind  
    between the parties.\n\n11.9 No Third-Party Beneficiaries.\nSource:  
    30-pl\nContent: (b) if Google believes, in good faith, that the Distributor has  
    violated or caused Google to violate any Anti-Bribery Laws (as defined in  
    Clause 8.5) or that such a violation is reasonably likely to occur,\nSource:  
    4-pl\n=====\nFINAL ANSWER: This Agreement is governed by English  
    law.\nSOURCES: 28-pl\nQUESTION: What did the president say about Michael  
    Jackson?\n=====\nContent: Madam Speaker, Madam Vice President, our First  
    Lady and Second Gentleman. Members of Congress and the Cabinet. Justices of the  
    Supreme Court. My fellow Americans. \n\nLast year COVID-19 kept us apart. This  
    year we are finally together again. \n\nTonight, we meet as Democrats  
    Republicans and Independents. But most importantly as Americans. \n\nWith a duty  
    to one another to the American people to the Constitution. \n\nAnd with an  
    unwavering resolve that freedom will always triumph over tyranny. \n\nSix days  
    ago, Russia's Vladimir Putin sought to shake the foundations of the free world  
    thinking he could make it bend to his menacing ways. But he badly miscalculated.  
    \n\nHe thought he could roll into Ukraine and the world would roll over. Instead  
    he met a wall of strength he never imagined. \n\nHe met the Ukrainian people.  
    \n\nFrom President Zelenskyy to every Ukrainian, their fearlessness, their  
    courage, their determination, inspires the world. \n\nGroups of citizens  
    blocking tanks with their bodies. Everyone from students to retirees teachers  
    turned soldiers defending their homeland.\nSource: 0-pl\nContent: And we won't  
    stop. \n\nWe have lost so much to COVID-19. Time with one another. And worst of  
    all, so much loss of life. \n\nLet's use this moment to reset. Let's stop  
    looking at COVID-19 as a partisan dividing line and see it for what it is: A  
    God-awful disease. \n\nLet's stop seeing each other as enemies, and start
```

seeing each other for who we really are: Fellow Americans. \n\nWe can't change how divided we've been. But we can change how we move forward-on COVID-19 and other issues we must face together. \n\nI recently visited the New York City Police Department days after the funerals of Officer Wilbert Mora and his partner, Officer Jason Rivera. \n\nThey were responding to a 9-1-1 call when a man shot and killed them with a stolen gun. \n\nOfficer Mora was 27 years old. \n\nOfficer Rivera was 22. \n\nBoth Dominican Americans who'd grown up on the same streets they later chose to patrol as police officers. \n\nI spoke with their families and told them that we are forever in debt for their sacrifice, and we will carry on their mission to restore the trust and safety every community deserves.\nSource: 24-pl\nContent: And a proud Ukrainian people, who have known 30 years of independence, have repeatedly shown that they will not tolerate anyone who tries to take their country backwards. \n\nTo all Americans, I will be honest with you, as I've always promised. A Russian dictator, invading a foreign country, has costs around the world. \n\nAnd I'm taking robust action to make sure the pain of our sanctions is targeted at Russia's economy. And I will use every tool at our disposal to protect American businesses and consumers. \n\nTonight, I can announce that the United States has worked with 30 other countries to release 60 Million barrels of oil from reserves around the world. \n\nAmerica will lead that effort, releasing 30 Million barrels from our own Strategic Petroleum Reserve. And we stand ready to do more if necessary, unified with our allies. \n\nThese steps will help blunt gas prices here at home. And I know the news about what's happening can seem alarming. \n\nBut I want you to know that we are going to be okay.\nSource: 5-pl\nContent: More support for patients and families. \n\nTo get there, I call on Congress to fund ARPA-H, the Advanced Research Projects Agency for Health. \n\nIt's based on DARPA-the Defense Department project that led to the Internet, GPS, and so much more. \n\nARPA-H will have a singular purpose-to drive breakthroughs in cancer, Alzheimer's, diabetes, and more. \n\nA unity agenda for the nation. \n\nWe can do this. \n\nMy fellow Americans-tonight, we have gathered in a sacred space-the citadel of our democracy. \n\nIn this Capitol, generation after generation, Americans have debated great questions amid great strife, and have done great things. \n\nWe have fought for freedom, expanded liberty, defeated totalitarianism and terror. \n\nAnd built the strongest, freest, and most prosperous nation the world has ever known. \n\nNow is the hour. \n\nOur moment of responsibility. \n\nOur test of resolve and conscience, of history itself. \n\nIt is in this moment that our character is formed. Our purpose is found. Our future is forged. \n\nWell I know this nation.\nSource: 34-pl\n\n====\nFINAL ANSWER: The president did not mention Michael Jackson.\nSOURCES:\nQUESTION: How many companies have been considered to create portfolios?\n====\nContent: \nShares of 40 (+1, the gold) different companies have been analyzed and clustered.\nSource: Flavio_Brienza_Abstract_Tesi.pdf\nContent: \"Prima di sviluppare questa sezione, le azioni di 40 compagnie (+1, l'oro) sono state analizzate e classificate usando sia metodi statistici che Natural Language Processing sulle ultime notizie riguardanti le aziende stesse.\" \nSource: Flavio_Brienza_Abstract_Tesi.pdf\nContent: \nThe shares of 40 (+1, the gold) different companies have been analyzed and\nSource:

```

Flavio_Brienza_Abstract_Tesi.pdf\n===== \nFINAL ANSWER:"
]
}
[llm/end] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain > 8:chain:LLMChain > 9:llm:OpenAI] [1.05s]
Exiting LLM run with output:
{
  "generations": [
    [
      {
        "text": " Forty (+1, the gold) companies have been considered to create
portfolios.\nSOURCES: Flavio_Brienza_Abstract_Tesi.pdf",
        "generation_info": {
          "finish_reason": "stop",
          "logprobs": null
        }
      }
    ]
  ],
  "llm_output": {
    "token_usage": {
      "completion_tokens": 33,
      "total_tokens": 1708,
      "prompt_tokens": 1675
    },
    "model_name": "text-davinci-003"
  },
  "run": null
}
[chain/end] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain > 8:chain:LLMChain] [1.06s] Exiting Chain run
with output:
{
  "text": " Forty (+1, the gold) companies have been considered to create
portfolios.\nSOURCES: Flavio_Brienza_Abstract_Tesi.pdf"
}
[chain/end] [1:chain:RetrievalQAWithSourcesChain >
3:chain:MapReduceDocumentsChain] [2.77s] Exiting Chain run with output:
{
  "output_text": " Forty (+1, the gold) companies have been considered to create
portfolios.\nSOURCES: Flavio_Brienza_Abstract_Tesi.pdf"
}

```

```
[chain/end] [1:chain:RetrievalQAWithSourcesChain] [3.00s]
```

Exiting Chain run with output:

```
{  
  "answer": " Forty (+1, the gold) companies have been considered to create  
portfolios.\n",  
  "sources": "Flavio_Brienza_Abstract_Tesi.pdf"  
}
```

```
[21]: {'answer': ' Forty (+1, the gold) companies have been considered to create  
portfolios.\n',  
       'sources': 'Flavio_Brienza_Abstract_Tesi.pdf'}
```

The answer is correct.