

# ML\_shaz13\_CPU

June 20, 2024

2. (M/D) Comparar um conjunto de algoritmos de aprendizado de máquina (pelo menos cinco) em aplicações relacionadas ao seu projeto ou alguma outra aplicação real sugerida por você. Neste caso, espera-se que a aplicação seja útil e conte com uma base de dados interessante e de tamanho que permita a aplicação de algoritmos de aprendizado de máquina.

## 1 Coleta de Dados

### 1.0.1 <https://www.kaggle.com/datasets/shaz13/real-world-documents-collections>

```
[53]: import pandas as pd
import os
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
[54]: from typing import List
from ETL_download_dataset_kaggle import download_dataset, unzip_file
from ETL_process_image import process_dataset, draw_bounding_boxes

path_dataset = "./dataset/real_world_documents_collections/docs-sm"
all_folders = ['invoice', 'file_folder', 'memo', 'specification', 'letter',
↳ 'scientific_publication', 'scientific_report', 'budget', 'resume',
↳ 'handwritten', 'questionnaire', 'form', 'email', 'advertisement',
↳ 'presentation', 'news_article']
path_df_parquet = './DF_shaz13_real_world_documents_collections_V3.parquet'
required_folders = ['form', 'resume', 'letter', 'invoice', 'questionnaire']
# required_folders = ['form']
def etl_for_dataframe(path_dataset: str, path_df_parquet: str, required_folders:
↳ List[str]) -> pd.DataFrame:
    """
    Gerencia o fluxo de download e descompactação do dataset.
    Extrai textos das imagens
    Prepara textos para serem usado pela ML
    """
    zip_file = 'real-world-documents-collections.zip'
    extract_to = 'dataset/real_world_documents_collections'
    link_dataset = 'shaz13/real-world-documents-collections'
    if not os.path.exists(zip_file):
```

```

        # Baixar o dataset
        download_dataset(link_dataset)
        # descompacta arquivo ZIP
        unzip_file(zip_file, extract_to)
        df = process_dataset(path_dataset, path_df_parquet, required_folders)
        return df

df = etl_for_dataframe(path_dataset, path_df_parquet, required_folders)

```

## 2 Análise Exploratória de Dados (EDA)

```
[55]: df.columns
```

```
[55]: Index(['text', 'class_img', 'name', 'dict_ocr'], dtype='object')
```

```
[56]: print(df.head(5))
```

```

                                text      class_img \
0  bellomycarrigginc job university parkway aa ex...  questionnaire
1  inc broadway york screening questionnatre time...  questionnaire
2                                                ozle  questionnaire
3  alan november account smoking manufacture desc...  questionnaire
4  market monttor office use interviewer resp res...  questionnaire

                                name                                dict_ocr
0  505527865_505527879.jpg  {'block_num': [0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2...
1    89583564_89583573.jpg  {'block_num': [0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
2  503543120_503543128.jpg  {'block_num': [0, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3...
3      501525861.jpg  {'block_num': [0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2...
4      71224770.jpg  {'block_num': [0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2...

```

## 3 Limpeza dos Dados

```

[57]: # Substitui strings vazias por NaN
df['text'] = df['text'].replace('', np.nan)
# Conta o número de valores NaN na coluna 'text'
num_nan = df['text'].isna().sum()
print(f'Quantidade de linhas vazias: {num_nan}')

```

Quantidade de linhas vazias: 103

```

[58]: # Remove linhas onde a coluna 'text' está NaN
df = df.dropna(subset=['text'])
num_nan = df['text'].isna().sum()
print(f'Quantidade de linhas vazias após a remoção: {num_nan}')

```

Quantidade de linhas vazias após a remoção: 0

```
[59]: def text_smaller_than(df, amount: int):  
      # Filtrar linhas onde a string na coluna 'text' tem menos de 'amount'  
      ↪caracteres  
      short_text_rows = df[df['text'].str.len() < amount]  
      # Contar o número de tais linhas  
      num_short_text_rows = short_text_rows.shape[0]  
      return num_short_text_rows  
amount: int = 5  
num_short_text_rows = text_smaller_than(df, amount)  
print(f'Quantidade de textos que é menor que {amount} caracteres:↵  
      ↪{num_short_text_rows}')
```

Quantidade de textos que é menor que 5 caracteres: 24

```
[60]: # Remover as linhas onde a string na coluna 'text' tem menos de 'amount'  
      ↪caracteres  
df_filtered = df[df['text'].str.len() >= amount]  
df = df_filtered  
num_short_text_rows = text_smaller_than(df, amount)  
print(f'Quantidade de textos que é menor que {amount} caracteres:↵  
      ↪{num_short_text_rows}')
```

Quantidade de textos que é menor que 5 caracteres: 0

### 3.0.1 Analisar classes (target) do df

```
[61]: # Conta a quantidade de linhas que têm o mesmo valor na coluna 'class_img'  
class_img_counts = df['class_img'].value_counts()  
print(f'Quantidade imagens de cada classe: {class_img_counts}')
```

Quantidade imagens de cada classe: class\_img

resume	303
form	299
letter	297
invoice	276
questionnaire	267

Name: count, dtype: int64

```
[62]: # Contar as ocorrências de cada classe  
class_counts = df['class_img'].value_counts()  
  
# Configurar o estilo do gráfico  
sns.set(style="whitegrid")  
  
# Criar o gráfico de barras  
plt.figure(figsize=(10, 6))
```

```

ax = sns.barplot(x=class_counts.index, y=class_counts.values, palette="viridis")

# Adicionar anotações sobre as barras
for i, value in enumerate(class_counts.values):
    ax.text(i, value + 5, str(value), ha='center', va='bottom', fontsize=12)

# Adicionar título e rótulos dos eixos
plt.title("Document Distributions", fontsize=16)
plt.xlabel("Class", fontsize=14)
plt.ylabel("Count", fontsize=14)

# Ajustar o layout
plt.tight_layout()

# Exibir o gráfico
plt.show()

```

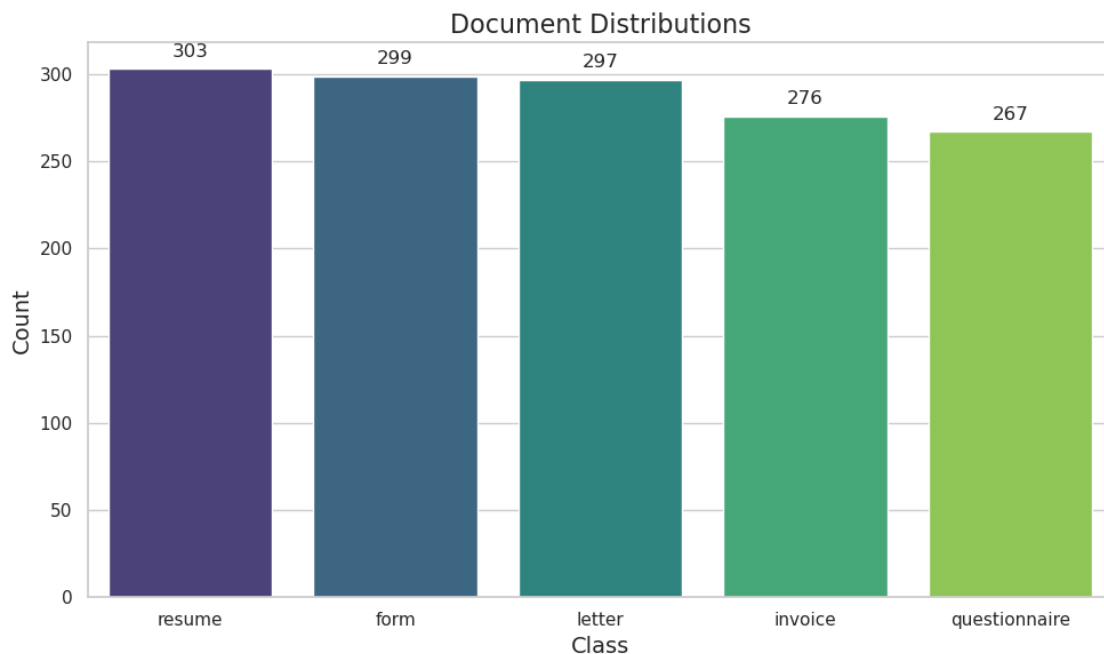
/tmp/ipykernel\_8786/1263128678.py:9: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```

ax = sns.barplot(x=class_counts.index, y=class_counts.values,
palette="viridis")

```



Nivelando aleatoriamente a quantidade de imagens de cada classe para ficar do mesmo tamanho que a menor

```
[63]: # Define o número de linhas que cada classe deve ter
min_class_count = class_img_counts.min()
# Amostra aleatória de cada classe para ter a mesma quantidade que a menor
↳ classe
balanced_df = df.groupby('class_img').apply(lambda x: x.
↳ sample(min_class_count)).reset_index(drop=True)
# Verifica a quantidade de linhas em cada classe no DataFrame balanceado
balanced_class_img_counts = balanced_df['class_img'].value_counts()
df = balanced_df
print(f'Quantidade de imagens em cada classe após balanceamento:
↳ \n{balanced_class_img_counts}')
```

Quantidade de imagens em cada classe após balanceamento:

```
class_img
form          267
invoice       267
letter        267
questionnaire 267
resume        267
Name: count, dtype: int64
```

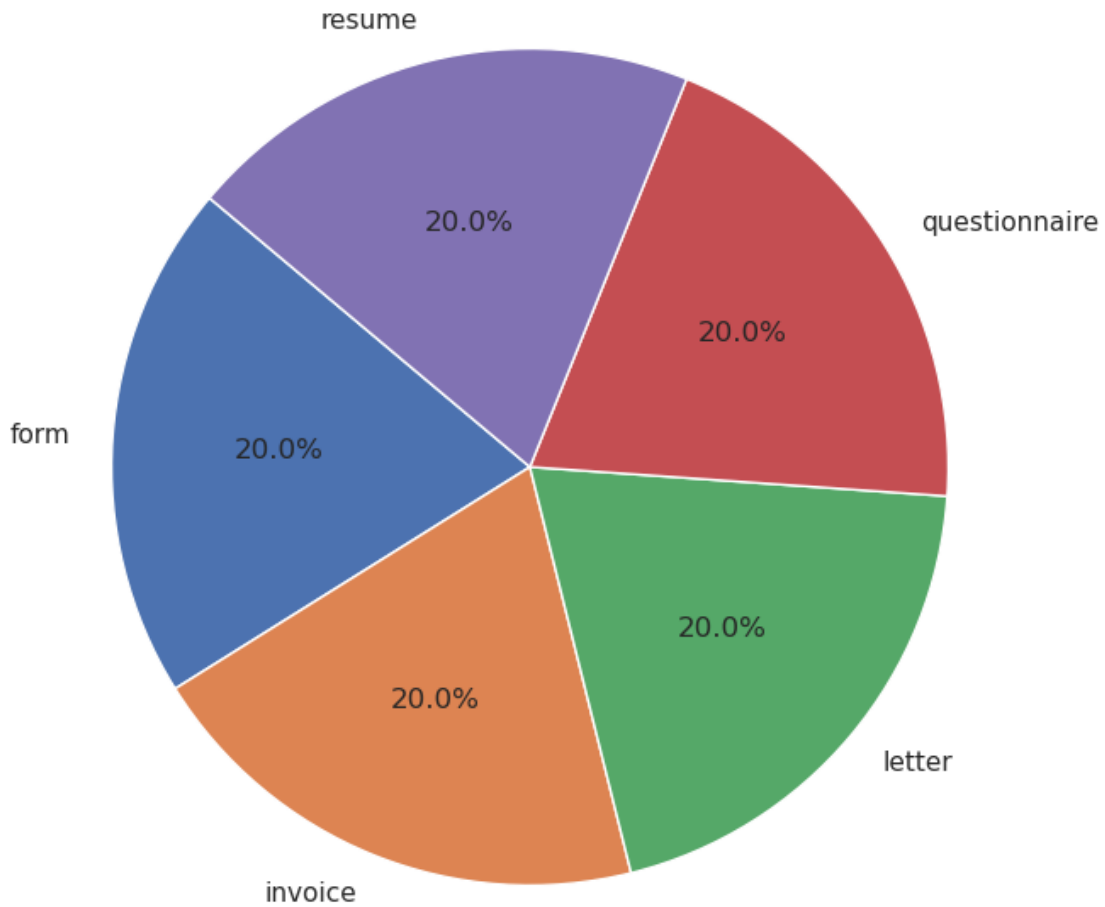
/tmp/ipykernel\_8786/2612964233.py:4: DeprecationWarning: DataFrameGroupBy.apply operated on the grouping columns. This behavior is deprecated, and in a future version of pandas the grouping columns will be excluded from the operation. Either pass `include\_groups=False` to exclude the groupings or explicitly select the grouping columns after groupby to silence this warning.

```
balanced_df = df.groupby('class_img').apply(lambda x:
x.sample(min_class_count)).reset_index(drop=True)
```

```
[64]: # Contar as ocorrências de cada classe
class_counts = df['class_img'].value_counts()

# Criar um gráfico de pizza
plt.figure(figsize=(8, 8))
plt.pie(class_counts, labels=class_counts.index, autopct='%1.1f%%',
↳ startangle=140)
plt.title("Document Distributions")
plt.show()
```

Document Distributions



```
[65]: print(df.head(5))
```

	text	class_img	\
0	e record destruction date january linda boyd a...	form	
1	infibo rso inbifo institut forschung gmbh cont...	form	
2	european community manufacturer limited telef...	form	
3	estimate date leo burnett medium e job number ...	form	
4	hammons schwalm j campbell po order amount req...	form	

	name	dict_ocr
0	516776562+-6570.jpg	{'block_num': [0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3...
1	2050974573.jpg	{'block_num': [0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
2	2046546273_6274.jpg	{'block_num': [0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2...

```
3      2045468000.jpg {'block_num': [0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2...
4  518446049+-6049.jpg {'block_num': [0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2...
```

### 3.0.2 Lista de palavras que são consideradas stopwords (palavras vazias) para o idioma inglês.

```
[66]: from nltk.corpus import stopwords
stopwords_list = set(stopwords.words("english"))

def check_for_stopwords(text: str, stopwords_list: List[str]) -> List[str] | None:
    """
    Função para verificar se há stopwords em uma string
    """
    # Dividir a string em palavras
    words = text.split()
    # Verificar se há interseção entre as palavras e as stopwords
    stopwords_in_text = set(words) & stopwords_list
    if stopwords_in_text:
        return list(stopwords_in_text) # Retorna lista de stopwords encontradas
    else:
        return None # Retorna None se não há stopwords
# Aplicar a função check_for_stopwords à coluna 'text' e armazenar as stopwords encontradas
df['has_stopwords'] = df['text'].apply(lambda x: check_for_stopwords(x, stopwords_list))
# Filtrar o DataFrame para incluir apenas linhas onde 'has_stopwords' não é None
filtered_df = df[df['has_stopwords'].notna()]
# Exibir índice da linha, texto original e has_stopwords não vazio
for index, row in filtered_df.iterrows():
    print(f"Índice da linha: {index}")
    print(f"Texto original: {row['text']}")
    print(f"has_stopwords: {row['has_stopwords']}")
    print() # Linha em branco para separar os resultados
```

Índice da linha: 19

Texto original: certificate insurance confers right upon certificate holder  
certificate johnson higgins amend extend alter coverage afforded broad street  
company affording coverage in fire accident indemnity morris company inc  
subsidiary co incl kraft general city fire in co food incoscar mayer food corp  
miller brewing underwriter in co park avenue new york new york coverage policy  
insurance listed issued insured named policy requirement team condition respect  
may insurance afforded described subject exclusion condition nay aeduced paid  
personal nj la operation santa ana beverage inc consolidated beverage  
distributor downtown santa ana business association policy expiration date  
thereof company notice certificate named left failure mail notice shall impose  
obligation kino upon company agent representative representative corporation

has\_stopwords: ['in']

Índice da linha: 123

Texto original: nt va information hunton riverfront plaza east byrd street  
telecopier no richmond virginia floor telecopier william brandt esq firm philip  
mortis location new york telecopy number page including cover original follow  
yes cheryl ragsdale hunton floor th problem transmission contact telecopy  
operator rico also included draft believe must updated include least inbifo  
others maybe new information discussed discussed preparing alist case dts  
supplement necessary pr response operator date october time clientmattor tothe  
reader message th employee agent responsible copying please message u postal  
has\_stopwords: ['no']

Índice da linha: 147

Texto original: name information hunton williams plaza east byrd street richmond  
virginia nd floor telecopier david bemick kirkland elis thomas schroeder womble  
cariyle jim young bill jones day dan willoughby king spaulding alan kaufinan  
thomas albright jones day ny bill newbold thompson coburn david murphy wachtel  
lipton jim goold covington burling cooneyk davis polk william hendricks king  
spaulding kendrick well brown williamson daniel donahue reynolds tobacco charles  
wall philip john mulderig philip morris james wilson lorillard jamie clark shook  
hardy bacon page including cover original follow mail yes jack mcclard floor  
telecopier no  
has\_stopwords: ['no']

Índice da linha: 151

Texto original: foreign matter report date brand country quip blend  
manufacturing facility shift foreign matter found found location foreign matter  
found type found unit machine no found machine no common bulker foreign matter  
cause choke machine listed checked foreign matter yes bulker time bulker started  
bulker used ono time bulker finished foreign matter go cutter pound tobacco run  
bulker prior finding foreign matter pound tobacco remaining bulker type  
containerlocation tobacco run bulker plus screen checked yes found location can  
filled otv bulker used datetime filled cut date shipped shipped owp amount  
tobacco remaining time product quipped pound tobacco bulker quipped case quip  
case size export bulk rejects cleanout phone supervisor  
has\_stopwords: ['can', 'no']

Índice da linha: 178

Texto original: philip morris westchester avenue rye telephone fax sheet  
distribution mr dermot gatenby m jan goodheart mr owen smith mr roger whidden  
matthew n winokur page follow philip morris international corporate affair fax  
call valerie fax  
has\_stopwords: ['m']

Índice da linha: 230

Texto original: competitive product introduction progress report manufacturer r  
j j brand camel menthol da packing full flavor box light box reporting period



aug oct nov forward th following month market geography area involved reynolds  
chicago region buffalo region inthe chicagoland area product worked sale rep po  
metal sign counter mat danglers poster decal mobile hanging suppor expressway  
town wisconsin approximately madison fond du lac kenosha area attached map  
includes milwaukee metro area area po currently promotional coupon sticker  
permanent te camel support price point indicates distribution cost carton force  
product rapid po perm many instance appears unitfor camel panel fr lighted  
header flap order case sold fast also facsimile display cardboard unit column  
carton packing wisconsin product shipped however appear sale perm display row  
promotional occasionally told thatthe sale rep coupon next ime around  
distributor termsintro deal accepted product m per case distributor involvement  
\_dearborn grocery wholesaler asked solicit carton display sale rep sale rep  
display sale also receives display shipped offer supposed extended rep working  
indiana michigan southwestern area ohio jobber shipping gf product live nothing  
reported bgf deal wisconsin jobber reported shipping deal open pantry chain  
wisconsin storés one account wisconsin kwik trip store accepted product store  
shipped pack temp store coupon banner wholesale chicago seling pack bgf display  
floor cash carry customer  
has\_stopwords: ['m']

Índice da linha: 234

Texto original: fax humphrey farrington mcclain pc information contained  
privileged confidential individual entity named sheet reader tho intonded  
employee agent responsible deliver recipient heshe t hereby notified reading  
dissemination copying communication strictly prohibited received communication  
error please immediately notify u telephone return original message u address  
via u postal service upon request rolmburse return thank transmittal sheet see  
attached attached sackman liggett page including cover hard copy follow remark  
date  
has\_stopwords: ['t']

Índice da linha: 251

Texto original: herbert february vice president public policy winstonsalem m  
pauline crumpton greco junior high school e fowler avenue temple terrace fl dear  
m crumpton thank interest youth nonsmoking material table card booklet ordered  
enclosed poster sent separately poster placed classroom school bulletin board  
location student see table card placed desk cafeteria table etc two booklet  
designed used parent adult talking kid peer pressure peer influence use  
counselor interested teacher might also want give copy parent group leader  
provide additional copy parent group interest quantity booklet requested however  
reduced developed use adult rather distribution student interest hope material  
helpful effort persuade young people smoke herbert osmon enclosure smoker  
has\_stopwords: ['m']

Índice da linha: 259

Texto original: company august winstonsalem nc m marie sexton buchanan clean  
county committee route watkins building grundy va dear m sexton thank recent  
request pocket ashtray pocket ashtray program important part reynolds tobacco

environmental program appreciate interest support within next week receive following material vantage pocket ashtray beach golf bumperstickers beach tentcards golf reminder material give smoker year age older im sure aware cigarette target antilitter campaign smoker dont take time properly dispose cigarette litter compound problem cigarette litter filter packaging carton material disposed designated waste container haphazardly discarded pocket ashtray offer one solution litter problem first introduced japan pocket ashtray promoted courteous smoking country almost year reusable ashtray made plastic inside contains fireresistant material prevent burning package designed extinguish cigarette recommend putting cigarette first placing butt pouch work smoker

has\_stopwords: ['m']

Índice da linha: 310

Texto original: robert e trout sr adelina court agoura hill ca january reynolds tobacco company nc attn chairman board andor chief executive officer dear sir madam year old male born smoking winston cigarette year since age needle say currently addicted brand cigarette nicotine please allow opportunity profile smoking history past year purpose write letter going nicotine addition hypnotized numerously injected various blocker chemical used patch joined stopsmoking club name tried im trying time im going quite frankly im pissed offbig time might laughing saying took klutz year figure nicotine addictive life ive smoked knew addicted try might quit went hell every time put family hell well bottom line might take real hard notice took long get mad bet as im mission stop contaminating youngster especially new grandson becoming addicted tobacco seems though money talk world understand one individual cause much financial pain group hotshot lawyer defend kinda like lying public dancing around truth year well gentleman ive decided play little game know rule play win enclosed first invoice nicotine treatment solely responsible standby boy get better admission altered amount nicotine additive bullshit marking ploy brand cigarette tobacco additive stop go end hurt buying product even though know hurt people around

has\_stopwords: ['as']

Índice da linha: 357

Texto original: form comp philip morris date voucher explanation develop scout southeen pennsylvania manor road miquon pa payee federal tax id s number invoice amount project code accounting invoice approval return name michelle manning x cash control use bank code company check accounting use action vendor invoice invoice dati invoice invoice voucher due number bypass separate check separate check check description freight audit term code amount

has\_stopwords: ['s']

Índice da linha: 369

Texto original: m mary eec cost report guideline europe

has\_stopwords: ['m']

Índice da linha: 390

Texto original: political campaign contribution request state recipient  
informaton billy bowles cr houston m office district representative office party  
democrat leadership position committee contribution information amount code  
payable billy bowles campaign address address date needed date approved revised  
senior vice signature  
has\_stopwords: ['m']

Índice da linha: 397

Texto original: political campaign contribution request date september state  
local state recipient name tim ford address speaker office new capitol building  
jackson m current office district representativedistrict aspired office district  
party leadership position speaker house nja contribution information amount  
account code check pay tim ford campaign fund address different address note  
date needed fundraiser dateother information approved vice president signature  
vice president signature  
has\_stopwords: ['m']

Índice da linha: 455

Texto original: north ave new york benson invoice form advertising placement  
total placement le discount general agency fee invoice attention m elisa copy  
read benson hedge blue indiania black expo est director  
has\_stopwords: ['m']

Índice da linha: 494

Texto original: blk lt othe million triple super stop slice slice stk polar tnc  
ot pol polar ot free shine ch s shine soott ritz ork btr  
has\_stopwords: ['s']

Índice da linha: 535

Texto original: cr register contract research center telefax avenue du péage  
autoroute roi baudouin zaventem juil dédouanement marchandises monsieur nous  
vous prion dé bien vouloir dédouaner no frais dise en annexe pesant ca kg et  
passant la vers aujourd'hui le frais de sont notre charge si vous désirez un de  
notre part veuillez prendre contact avec moi grand merci d'avance veuillez agréer  
monsieur no salutation le plus distinguées van den buleke management accountant  
facture pro forma  
has\_stopwords: ['no']

Índice da linha: 537

Texto original: rjr interoffice memorandum december c w attachment summary  
consolidation effort tow fine paper see work led paul chapman ec jones led  
significant reduction category filter tow plug wrap cigarette paper reduced  
since several issue reflecting plan follows consolidation work done cigarette  
paper unless identified economic benefit fewer item tinsley attachment m jackie  
renegar  
has\_stopwords: ['m']

Índice da linha: 545

Texto original: inc new july j morgan president philip morris usa park avenue  
new york ny attentionmaureen dear maureen note confirm budget meeting office mr  
morgan m keane wednesday october awending ftom ctr dr james glenn chairman dr  
harmon c scientific director help sincerely lorraine treasurer charlene m  
has\_stopwords: ['m']

Índice da linha: 571

Texto original: leo burnett advertising prudential plaza illinois january mr  
robert wade publicity department finger lake racing association inc box  
canandaigua ny dear mr wade requested m esther fortner philip morris usa  
enclosed please find necessary fourcolor printing material marlboro ad al day ad  
scheduled run back cover finger lake race track daily racing program fourcolor  
bleed ad question please hesitate call also please forward copy program record  
thank sincerely leo diane smith print service coordinatormarlboro enclosure mr  
dudynskay lb m fortner pm mr marchi lb  
has\_stopwords: ['m']

Índice da linha: 575

Texto original: sample retailer letter individual retailer october gale held  
director state prevention system program center substance abuse prevention  
rockwall building th floor fisher lane rockville md m held retail trade  
professional wish object strenuously official harassment proposing industry  
guise enforcing section adamha reorganization act original language section  
appropriate congress simply requires state receiving adamha block grant enforce  
minimum age law purchase tobacco product manner reasonably expected reduce  
availability minor language acknowledges underage tobacco use legitimate issue  
league drug alcohol abuse two evil adamha fund meant combat also assumes rightly  
state retailer already committed enforcing obeying minimum age law purchase  
tobacco product onerous combative enforcement language proposed hhs contrast  
assumes adversarial relationship retailer imposes upon state retail trade  
excessive cost bureaucracy rule proposed hhs would require retailer bear  
estimated million annually compliance cost hhs suggests retailer absorb  
additional million year lawenforcement cost new licensing requirement  
distressing hhs would subject retailer vigilante zeal antismoking activist  
encouraging private employ child conducting frequent operation unwary retailer  
even offensive hhs suggests publishing name local store entrapped operation  
would wish steer patronage store violate please delete provision proposed rule  
sum fantastic excessive rule proposed hhs measure compliance section unnecessary  
existing state law conjunction responsible industry program promoting compliance  
law sufficient sincerely  
has\_stopwords: ['m']

Índice da linha: 592

Texto original: ale office real newspaper comunicus pretest date july mrd\_ mr p  
clark maria falconetti mr j wallace mr jack moore comunicus inc wednesday july  
present research finding real newspaper pretest presentation th floor large  
conference thank maria falconetti marketing research department cc mr h johnston  
mr w tredennick mr g harlow m mcreynolds sharpe m feakins

has\_stopwords: ['m']

Índice da linha: 601

Texto original: june m abner image entry industrial boulevard london ky virginia  
slims program setup dear enclosed find program instruction following virginia  
slims program virginia slims july v catalog maintainreward undeliverables  
virginia slims july v dm virginia slims july v catalog dm virginia slims july v  
catalog om club virginia slims july king catalog dm programming anc begun key  
response please provide test gump randomly selected document standard report  
ship data pm adult smoker database dump question please hesitate contact  
sincerely maria chiu j lerry e marini comb j mullins image entry browning chang  
claggett conroy l l grey

has\_stopwords: ['m']

Índice da linha: 610

Texto original: ce company yy th december mr cousin bat uk export ltd export  
house woking surrey gu b england dear tony many thanks letter dated th november  
certainly make sense await product test result deciding blend change jps taken  
account urgency need well comparative relevance two piece research hong kong  
market propose make reference hong kong product research word prepared consider  
r marlboro lucky strike however rely b w pretation research result since neither  
research agency bat hk knowledge coding reference kindly advise u joint  
recommendation soon possible kind regaras sincerely allen ck leung m scott jv  
uke

has\_stopwords: ['m']

Índice da linha: 616

Texto original: ine marketing park new york septenber mr willard upeburch buyer  
j reynolds tobacco co north main street winstonsalem korth carolina  
schechterluth doral tar nicotine chanj schechter luth project dear pleasant  
productive day epent associate yesterday know presented variety alternative  
handling tar nicotine nev doral carton reached method changed seven mechanicals  
ate decision offered send u photostatic copy altered mechanicals record please  
ask secretary put final tar number photostate order avoid future confusion aleo  
approved arcters ink color vith exception dark red somewhat lighter since  
wrapper already proofed proofing maining carton sample delicate color balance  
course available regard remaining modification wrapper carton filter menthol  
returned u ed consistent final change showa vill sent early next veek project  
draving close product reintroduced must say pleasure work brand group he

has\_stopwords: ['he']

Índice da linha: 625

Texto original: re street york dear sire writing book cigarette vould like end  
non snokers develop cigarette rether hebit better suoke one cen food vine  
porticuler ete book elso include different type blend utilized brand rettes  
bistory collecting usa could send printed type utilize type produce different  
brand trade pack history charge exper thank hearing

has\_stopwords: ['re']

Índice da linha: 627

Texto original: council researchusa new arthur april m kristen lawrence infoed western avenue guilderland dear m lawrence thank interest program research support response request pleased enclose copy explanatory material sent council tobacco research potential principal investigator inquired support research project material detail twostep application process executive committee scientific advisory board first evaluates preliminary application decides whether encourage submission full application decision based scientific merit also appropriateness inclusion one council multidisciplinary program area well recognition current support level similar project order office hand reference set full application form sample copy enclosed also find package statement policy information applicant recent annual report list currently supported grant contains abstract paper appeared last year acknowledging support tobacco research question please feel free contact sincerely eisenberg  
has\_stopwords: ['m']

Índice da linha: 633

Texto original: walsh inc dear mork te ole provide brand five major brand style benson no able nave parent tran teck support resulted  
has\_stopwords: ['no']

Índice da linha: 657

Texto original: rod kuegel ray branstetter allen harman althausen phone kentucky memorandum kean philip norvell universal leat kornegayjerry mabe reynolds ard leat oakes dimon danny mckinney date september spec enclosed general processing review change discusse may meeting incorporated applied crop meeting following noted change crop ofthe target pan inthe stem reporting target upper t moisture foreign matter check enhance crop year processing danny mokinney  
has\_stopwords: ['t']

Índice da linha: 661

Texto original: subject tcrc presentation celanese distribution interoffice memorandum date september mr jones october dining room building dr lance deutsch present result joint rjr study affecting softening filter smoking hot collapse study cover effect total denier denier per filament plasticizer level ventilation crimp index please invite area who might interested distribution bugene w barnhardt barry z black johnny l brook paul glenn james l harris leslie levi man norman michael poe colbert w seagraves connie ce mr j phillips m j wheeler jones mr daniel beeson mr tony carter mr dan r pugh dr mary e stowe mr ron willard  
has\_stopwords: ['m']

Índice da linha: 670

Texto original: e william free company inc newyork m j klein american tobacco company park avenue new york ny dear ellen decrease newspaper schedule covering seattle post intelligencer ceased publishing ad front page eff ner period auth auth de franco w mellett degenhardt donohue park avenue new york telex

has\_stopwords: ['m']

Índice da linha: 671

Texto original: janvary m vivian marra dear vivian returning call report v revision discussing belleve clarify operating policy respect material bates provide bw viceroy outdoor understanding mike mccue based group bates zml bates provide art layout cell overlay containing set type material indicate type placement apparently confusion resulting meeting bates provided aheet layout type cell overlay paint future provide sheet paint call report question therefore material provided throwing vpa read layout type cell overlay paint king would please arrange call report accordingly sincerely carter attachment mr heller mr j mecue mr l lewis j  
has\_stopwords: ['m']

Índice da linha: 684

Texto original: tobacco company august m janine girard mclane northwest po box tacoma wa dear m girard po attached please find signed authorization m joyellen johnston buyerjackpot food mart co ship one display per month company jackpot food mart currently scheduled r j reynolds valueadded promotion automatic plusout commence september scheduled camel bgf package promotion question regarding authorization please contact sincerely oguin special account manager r j reynolds tobacco company johnston jackpot food mart allen work smoker  
has\_stopwords: ['m']

Índice da linha: 691

Texto original: e william free company inc london january chester anERICAN tobacco company park avenue new york new york dear mr chester confirm conversation two new permanent paint unit york previously approved major deegan wo bridge unit queen expressway brooklyn recommend taking advantage thé availability unit starting soon possible gold design receive artwork mechanical stats tuesday estimate paint completed ry barring unforeseen total feb march space cost february please note cost space cost reflect fabrication charge able estimate fabrication cost receive artwork please advise wish u proceed painting thank messrs morse jacknan dann laymon m wright park avenue new york  
has\_stopwords: ['m']

Índice da linha: 700

Texto original: brown tobacco april m arlyn elizee student working tobacco office tobacco control se c tallahassee dear swat thank recent letter youth smoking prevention pointed shared cause preventing tobacco teenager pleased invited continue dialogue issue related youth tobacco use happy report progress part sponsoring program prevent youth smoking tightening control marketing program sure directed toward youth description program company sponsoring organization dedicated helping kid aim isto provide encouragement organization measure variety youth smoking prevention program program successful work help expand wider group child teenager ofthese program based idea building social resistance skill among successful technique preventing youth smoking address root cause researcher demonstrated peer pressure resistance skill communication

decisionmaking skill selfesteem assertiveness responsibility work help kid avoid  
temptation produced bw web site  
has\_stopwords: ['m']

Índice da linha: 710

Texto original: ror biomedical investigation new yori ny arthur phd april  
katherine johnson jd associate director administration institute cancer research  
care datapoint drive suite san antonio texas dear m johnson thank interest  
program research support response request pleased enclose copy explanatory  
material sent council tobacco research potential principal investigator inquired  
support research project material detail two step application process executive  
committee scientific advisory board first evaluates preliminary application  
decides whether encourage submission full application decision based scientific  
merit also appropriateness inclusion one multidisciplinary program area well  
recognition current support level similar project procedure take two month order  
office hand reference set full application form sample copy enclosed also find  
package statement policy information applicant recent annual report list  
currently supported grant contains abstract paper appeared last year  
acknowledging support council tobacco research question please feel free contact  
sincerely arthur eisenberg encl  
has\_stopwords: ['m']

Índice da linha: 715

Texto original: interoffice subject competitive activity date march mr wm hobbs  
brown williamson brown williamson brown williamson offered jobber extended term  
order excess normal weekly volume attached u mr frank armstrong st louis  
regional manager lorillara mr hank bass area sale manager informed u lorillard  
offer direct account second allocation kent iii order build trade inventory due  
pending truck strike april attached report brand mr warren ishan tulsa division  
manager philip morris march issue call news revealed following point interest  
philip morris started publishing newsy breezy pictureladen newsletter designed  
sale force called force company employ scientist support staff research center  
richmond review promotion used promote duty business also state duty free sale  
representative give review promotion kentucky derby churchill down military sale  
force responsibility use computer aid provided philip marketing strategist  
philip military sale specialist u copy article available mrd marketing research  
department attachment form  
has\_stopwords: ['down']

Índice da linha: 723

Texto original: interoffice meeting metzenbaum october amendment restrict  
importation foreign tobacco distribution p johnson meeting held raleigh october  
among technical represen tatives major tobacco manufacturing company purpose  
pose liste pesticide used tobacco country use status pesticide u tobacco  
production tion requested mr horace r kornegay use dealing rose metzenbaum  
amendment requiring certification pesticide cancelled suspended u used tobacco  
imported u raleigh meeting chaired dr preston leake company attended m leslie  
lewis mr p johnson j reynolds wording amendment changed slightly since firet



drafted originally would forbidden importation tobacco exposed pesticide  
cancelled suspended use tobacco u current wording would require certification  
none pesticide registration cancelled pended use tobacco u used production  
tobacco offered importation u exposure pesticide would difficult avoid pesticide  
suspended accumulated soil due long halflife remain year come furthermore use  
word exposure would imply zero tolerance residue pended pesticide could  
practically impossible achieve rew wording appeats much realistic  
has\_stopwords: ['m']

Índice da linha: 733

Texto original: medical th street oklahoma city ok administration c sommers  
director tobacco research th street new york ny deor dr sommers writing letter  
inquire cbout research proposal concerning study blood pressure lowering effect  
heparin tobacco research council preliminary observation indicate heparin ho  
potent antihypertensive effect mandal et al microvas re susic et al hypertension  
submitted proposal include study designed investigate mechanism hypotensive  
effect heparin expected proposed study contribute therapeutic approach  
hypertensive investigation conducted institute medical research beograd  
university oklahoma city drs kentera beograd yugoslavia ore established  
investigator field skill facility depth investigation designed area initiate  
various hemo dynamic study measurement humoral factor might involved hypotensive  
effect renal histopathologic study light electron microscopy done oklahoma  
however fluorecence microscopy study done beograd dr interested extend protocol  
studying effect nicotine compound tobacco cadmium whether heparin may modify  
effect tobacco according advice enclosing approximate budget year period three  
year  
has\_stopwords: ['re']

Índice da linha: 739

Texto original: september messrs l ogburn jr g e harlow johnston cummins hall dr  
w perry jr e blackiner m mackinnon burke nordine hifi category protocol research  
presentation result protocol research first phase major high filtration study  
changed wednesday september main conference room th floor alan cox arcaw cc mr  
novak  
has\_stopwords: ['m']

Índice da linha: 744

Texto original: february reynolds tobacco company north main street winstonsalem  
nc united state america m donna tuttleolm rd department dear madam acknowledge  
receipt letter dated january regarding delay however reference confidentiality  
agreement brought aforementioned letter would like clarify herewith solely  
agreement reynolds tekasago corp usa one perceived legal binding effect  
whatsoever reynolds takasago singapore would like propose drawing new separate  
ity agreement reynolds takasago singapore case disclosure information require  
ingredient formula component made possible hope kind understanding predicament  
looking forward early reply regarding disposition matter thank kind attention  
faithfully takasago international singapore pre ltd marketing sal  
has\_stopwords: ['m']

Índice da linha: 755

Texto original: reddie grose valentine patent european patent attorney rade  
agent london wcix pl telephone fax patrick ad bsc aidan j d jackson bsc jj day  
philip management corp law department patent section pbox richmond virginia  
united state america attn james e schardt asst gen counsel ref date apr dear ret  
finnish patent philip morris product inc carbon heat source note instruction  
abandon case act accordingly case removed record faithfully reddie grose  
has\_stopwords: ['d']

Índice da linha: 762

Texto original: tobacco institute street northwest suite dc senior fax august m  
elizabeth e helwig vice president first union national bank north street  
richmond va institute employee pension plan trust dear m helwig thank quick  
response request listing plan payment william inc covington burling cooper  
lybrand attached copy form schedule c part please check record provide  
information regarding payment reported line schedule c part william adam plan  
administrator attachment  
has\_stopwords: ['m']

Índice da linha: 770

Texto original: tobacco company august m suzanne verdonne marketing research  
counselor university plaza shopping center silas creek parkway ext winstonsalem  
north carolina dear suzanne subject product test screening attaching copy  
demographic design test currently scheduled please notice unless speci fied  
percent smoker brand category selection normal rate distribution applied ie  
specified reflect normal tribution also attached copy various way looking  
tribution smoker helpful question please call sincerely ray rodgers marketing  
research department  
has\_stopwords: ['m']

Índice da linha: 773

Texto original: best copy recent ovidence smoking involved development  
pneumoconiosis said kotins theme simple one remove risk worker one point  
employee feel enter workplace optimum physical condition free disease character  
istic associated increased susceptibility disease groat majority caseslung  
cancer uranimum miner asbestos worker coke oven worker obvious example workplace  
clearly element increased risk certainly exclusive one one engincering  
technological work practice control requires increased risk dimension cigarette  
smoking addressed well panel discussion role knowledge high risk group  
occupational health policy practice moderator michael united steel worker  
america said opposed kind smoking ban kotin reported effect johnsmanville said  
educational campaign preferable referring various comment others regard genetic  
screening wondered extent industry used device order avoid cleaning work environ  
ment kilian dow chemical co said company sup ported ahh research university  
texas however view failure antialcohol campaign even acceptable test ever  
developed detect susceptible person probably little effect mortality morbidity  
nicholas ashford mit disagreed kotins mark risk susceptibility past history

occupational environmental exposure important relation ceptilibity said stated  
management reduce compensation removing hazard high risk individual potential  
claimant dont hire ident disease precursor disease kotin also panelist said hed  
hard put find abnormality related disease diaphragm urogenic fold smoking major  
determinant smoking hypersusceptibility factor said company hire smoker job even  
slightest exposure pulmonary irritant company fund stopsmoking program workors  
accepts possibility may somebody constitutionally able give smoking physician  
would think he  
has\_stopwords: ['he']

Índice da linha: 808

Texto original: nicholas re discussion guide camel collector pack qualitative  
introduction postit brand fax transmittal respondent profile k profle brand much  
smoke per dey brand usage camel collector pack signage retro collector pack  
first repeat theme collector pack immediate reaction brand people might pack  
would attitude based pack probe key fun clever distinctive set  
has\_stopwords: ['re']

Índice da linha: 868

Texto original: marketing information system inc sylvan avenue book ad cigarette  
would t look ad long go next ad respondent finished take back ask skip im book  
hedge notice benson hedge brand advertised brand word whet main hedge ad benson  
vo buy benson hedge ad detail fully clarify exch response benson hedge ad go  
store buy cigarette would ask ad away place benson hedge packing front four  
brand one ad showed  
has\_stopwords: ['t']

Índice da linha: 920

Texto original: favor de lo que disclose para calificar needed qualify los  
grupos group spanish buenas con que hoy estanos haciendo un survey en su hacerle  
preguntas research conducting study like ask question hedle hablar con alguns en  
este efios yeers age older may speak residing hovsehold yeers age older one  
livin home tee age terminate speaking someone ask parar cuando categoria que  
contiene list could stop get contains age read list menos de terminate continue  
terminate terminate ma de terminate  
has\_stopwords: ['ma']

Índice da linha: 930

Texto original: cigarette test report form parti answered read player  
advertisement finished smoking pack cigarette start smoking player k cigarette  
anything dislike t cigarette anything like t cigarette pair different phrase  
might used describe t one phrase side pagewith box numbered feelaparticular  
phrase applies cigarette phrase would pick abox please check one box pair phrase  
like cigarette much lot taste lot taste good harsh mouth throat good taste harsh  
mouth throat e taste tasted low tar low tar good good aftertaste aftertaste g  
cool smoking smoking strong menthol mild menthol taste taste l good good menthol  
j refreshing important please keep report form handy telephone return u  
has\_stopwords: ['t']

Índice da linha: 950

Texto original: brc section self int want take look fantasize tell story int  
look like guy city kind pressured job possibly fastmoving made time relax get  
thought notice he smoking generally speaking teking thing he used taking weekly  
type getting away lot af thought together get end tied together ard trying get  
back perspective mean city look like scene desolate peaceful erea way dressed  
hair look like weuld office position white collar abcut way doas go together  
would dress way yes sweater hush puppy right bo think alone think alone  
capturing lot thing missed quite look like content right hink think x  
has\_stopwords: ['he']

Índice da linha: 959

Texto original: smoking le stari smoking think ore put one ofthe box mace ore o  
yeu put scoreboard dislike cigors like tell yes question yu way many cigors  
smoke oll given note that even smoked little smoke ony little nax page ore ave  
il little  
has\_stopwords: ['o']

Índice da linha: 963

Texto original: two first relative work following read smoked seven day  
approximately many cigarette usually smoke per day study list family currently  
member one brand consider usual t chat style regular sized automobile food  
processing company advertising agency magazine company tobacco manufacturer  
insurance company marketing research fira yes le soft drink past che past year  
past year carettes tally yes tally continue nonfilter  
has\_stopwords: ['t']

Índice da linha: 965

Texto original: smoker test area mail panel smoker control area total smoker  
household called listed large computer printout sheet telephone number also  
listed sheet sorted two group test mail control sure mark correct area every  
card filled survey want interview every adult smoker household fill m card  
smoker interview example call number find husband grown son smoke would  
interview husband also son would two different m card filled household use  
prenumbered m card interview first smoker use flip side enter number name adult  
smoker household including first smoker enter family id number upper right  
corner m card used interview smoker family marking card find household adult  
smoke least one cigarette day fill prenumbered m card household showing area  
card col number smoker col phone number flip side specific market area enter  
flip side initial date first smoker use prenumbered m card show area card col  
number smoker household first smoker flip side smoker name adult household name  
including first smoker name flip side card first smoker name phone number  
specific market area eg san francisco santa barbara etc initial date response  
interview family income smoker use blank m card show area smoker card col family  
id number taken first smoker m card flip side card smoker name phone number  
initial date response interview lole znoos  
has\_stopwords: ['m']

Índice da linha: 971

Texto original: ip smoker list survey smoking indicted portion upcoming philip more usa service yes put preferred smoker list phone buy u cou way brand brand  
has\_stopwords: ['more']

Índice da linha: 982

Texto original: mr john huraphrey answer following question help u deliver information better informed issue u appreciate time willingness take moment answer question convenience postage paid reply envelope isthe smoking smoking g letter facetoface meeting please indicate issue would like information cigarette excise increase ban severe smoking prevention effo t federal lawsuit industry type information would like receive key issue please check apply one page issue summary talking point affect state study survey result elected official story presently pm usa com unicate consumer via however may implement communication would interested hearing u medium please indicate email address response  
has\_stopwords: ['t']

Índice da linha: 992

Texto original: tobacco company m kathleen worth chestnut avenue ca dear valued customer recently contacted reynolds consumer relation department gave u opportunity resporid request assistance either promotion product related issue better serve customer asking please take minute complete questionnaire well treated company representative completing reverse side letter please enclosed postagepaid envelope thank taking time let u know staff met expectation value opinion appreciate business mark conrad consumer satisfaction research manager work smoker  
has\_stopwords: ['m']

Índice da linha: 1005

Texto original: internship feedback survey philip morris experience compared expectation exceeded expectation met expectation expectation hed hoped recall recall summer exposed side business n tp view may safe soy me made meaningful tu time dene able tap involved recall within least interesting activity involved wity ow  
has\_stopwords: ['me']

Índice da linha: 1043

Texto original: survey inc park avenue york new york february respondent cluster frame interviewer cigarette study continued behavior long smoking one month regular basis one month month year year year year year average many cigarette day cigarette twenty cigerettes one pack forty pack forty le nore cigarette smoking daily time last year moreless cigarette current brand brand cigarette snoking present brand hand t regular king size tip filter nonmentholated cigarette  
has\_stopwords: ['t']

Índice da linha: 1060

Texto original: section interviewer im going show sone thing use imaginaticn  
tell little story like way he dressed cant tell story government auyble lawyer  
int think thinking know never tell man thinking aloof idea look content happy ox  
sad look like scmething serious mind eye smile facc soker succesful he arrived  
ne lawyer maybe want good lawyer job good maybe he thinking one case he cot one  
want il wiy deep thought think personal relationship would many int think lot  
friend general wo he select friend reminas mo someone used know ago jerk int  
krew well knew well well would let anyone know  
has\_stopwords: ['he']

Índice da linha: 1067

Texto original: comparison smoke evaluation experimental lot bae ncsu t milder  
better flavor preferred comment  
has\_stopwords: ['t']

Índice da linha: 1070

Texto original: university california san francisco curriculum vitae joseph h  
rapp professor surgery department surgery school medicine research  
cardiovascular research institute education university oregon eugene oregon b  
biology university oregon medical school md medicine portland oregon  
postaraduate training baltimore city hospital intern medicine baltimore maryland  
england medical center resident surgery boston massachusetts oregon health  
university research fellow portland oregon northwest permanente portland oregon  
university california clinical fellow san california vascular cardiovascular  
research research fellow university california san francisco california dr  
phillipson b rapp jh connor colestipol plus nicotinic acid treatment  
heterozygous familial lancet rapp jh connor lin d inahara porter jm lipide human  
atherosclerotic xanthoma clue plaque progression j lipid re wk rapp direct trunk  
aortic arch j vasc surg presented rapp jh qvarfordt p krupski wc wk stoney rj  
hypercholesterolemia early restenosis carotid endarterectomy surgery presented  
western surgical society krupoki wc olive gc weber ca rapp jh comparative effect  
hypertension nicotine injuryinduced myointimal thickening surgery rapp jh kerlan  
ring bj stoney rj preoperative balloon tipped catheter placement proximal  
control visceral artery aneurysm surgery surgery pan xm rapp jh wc hale jd  
sheldon p kaufman l identification aortic magnetic resonance imaging j vase surg  
rapp jh hw hamilton rl reilly wk stoney rj goldstone j kane jp particle size  
distribution human atherosclerotic plaque j surg  
has\_stopwords: ['re', 'd']

Índice da linha: 1073

Texto original: biographical sketch key paul institution location fielo ohio  
university athens ohio b zoology westem ano present federal puble tile al year  
two pathology case western reserve experimental case western reserve american  
hear association northeast ohi fellow pathology western reserve pathology case  
reserve university instructor theeapy cuyahoga community college professor  
pathology case reserve case research sbir sttr speciat member usarmdc breast  
cancer program cell nhlbi special american association olio medical health  
assistance fellowship award anderson xid plasma biomedical polymer pankowsky zit

topham od aad morphologie adsorbed plasma vascular j vase bonfield activation  
polymer topham pankowsky da andetson blood protein labeling k bp nakabayashi  
anderson maes re p anderson ofthe hageman supernates ziats anderson stp kennedy  
yun jk vi surface protein  
has\_stopwords: ['re']

Índice da linha: 1077

Texto original: fe biographical sketch provide information resham kulkarni  
postdoctoral fellow institution texas university texas microbiology university  
poona india phil biotechnology university poona india m microbiology university  
poona india b microbiology research professional experience win present  
employment present membership federal two exceed two page present postdoctoral  
research associate dept biology texas university research assistant dept biology  
texas university teaching assistant dept biology texas university research  
fellow national chemical laboratory pune india american society microbiology  
award scientific industrial research india junior research fellowship  
biotechnology government india scholarship biotechnology government open merit  
scholarship government maharashtra india ph rav form page ff use  
has\_stopwords: ['m']

Índice da linha: 1080

Texto original: ff carole form page page format carole liedtke associate  
professor institution location fielo miami university oxford oh ba microbiology  
case western reserve univ cleve oh phd anatomy case western reserve univ cleve oh  
physiology professional experience concluding previous experience membership  
federal fo pertinent exceed two page research professional experience research  
assistant case western reserve university fellow dept anatomy fellow dept  
pediatrics pharmacology department case western reserve university professor  
pediatrics developmental anatomy professor pediatrics physiology professor  
pediatrics physiology biophysics professor department biology st francis xavier  
university antigonish nova scotia honor honor magna cum laude general honor  
honor microbiology phi beta kappa alpha lambda delta lota sigma pi phi sigma who  
among student university college herbert steuhr memorial award research award  
cell general physiology section american physiological society government  
committee site visit team national center research resource nih general clinical  
research center publication total cm liedtke anion brush border membrane  
isolated rat small intestine biochem biophys re comm cm liedtke u hopfer  
electrolyte small intestine proceeding international congress hydrogen lon  
epithelium schulz g sachs forte kj ullrich editor elsevier amsterdam pp cm  
liedtke u hopfer kinetic feature cotransport mechanism isotope exchange  
condition membrane biochem liedtke u hopfer mechanism across small intestinal  
brush border membrane absence cotransport j physiol gastrointest liver gg  
liedtke hopfer mechanism across small intestinal brush border membrane  
demonstration exchange cl cotransport j physiol gastrointest liver physiol  
liedtke boat rudolph neurohormonal receptor camp binding protein rabbit  
mucosasubmucosa biochem biophys acta cm rudolph boat modulation mucin secretion  
physiol cell  
has\_stopwords: ['who', 're']

Índice da linha: 1085

Texto original: biographical sketch georg brunner assistant professor  
institution location study university west germany plant physiology university  
kaiserlatuem west germany phd plant physiology protease research professional  
experience fellow institute immunology genetics german cancer research center  
heidelberg west germany research fellow department cell biology nyu medical  
center new york tion proteolytic modification monoclonal antibody physiolo  
platarum v brunner g waller ca mechanism transendothelial cell passage matrix  
degradation metastatic tumor cell prog appl brunner g v fibrin autography  
plasminogen activator electrophoretic transfer fibrin agar gel anal biochem  
brunner g pohl erkell v schirmacher induction urokinase activity phenotype  
bladder carcinoma transfection haras oncogene j cancer re clin brunner simon mm  
md kramer activation prourokinase human cellassociated serine proteinase hutsp  
feb lett meissauer kramer md hofmann erkell lj jacob e schirmacher v g brunner  
urokinase type tissuetype plasminogen activator essential vitro invasion human  
melanoma cell exp cell re brunner g gabrilove j rifkin db wilson phospholipase c  
release basic fibroblast growth factor human bone marrow culture biologically  
active complex phosphatidylinositol anchored heparan sulfate proteoglycan j cell  
biol vettel u barshavit r simon brunner g viodavsky md kramer coordinate  
secretion functional synergism cellassociated serine proteinase mtsp  
endoglycosidases activated cell meissauer kramer schirmacher v brunner  
generation cell surfacebound plasmin cellassociated urokinasetype secreted  
tissuetype plasminogen activator key event melanoma cell invasiveness vitro exp  
cell re g brunner vettel u kramer mb v cellrelated expressed lymphoma cell  
activates endogenous prourokinase blood md kramer u vettel schmitt j reinartz g  
brunner monoclonal antibody plasminogen activator plasminogen fibrinolysis suppl  
kramer mullerbardorff pj frosch g brunner sodium dodecyl sulfate sdsinduced  
activati cutaneous plasminogen activator system molecular pathway possibly  
involved sdsinduced skin tn cell tissue culture model dermatological ed et al  
springer berlin heidelberg germany pp brunner nguyen gabrilove j rifkin db basic  
fibroblast growth factor human bone marrow peripheral blood cell blood u vettel  
g brunner r barshavit md kramer chargedependent binding granzyme mtsp basement  
membrane eur j immunol press  
has\_stopwords: ['re']

Índice da linha: 1086

Texto original: principal biographical sketch key personnel consultant begin  
withthe director photocopy phd poston muchien sun graduate student education  
begin professional education suich nursing include location year study taiwan  
university b zoology taiwan republic china taiwan university m anatomy taiwan  
republic china research concluding honor key include principal investigator  
individual execution key personnel wil wih doctoral professional project  
substantive way fo development execution project include present membership  
federal public advisory complete referances publication past three end fo  
application page education present graduate student vanderbilt university  
nashville tennessee graduate student national university talwan china  
undergraduate student national taiwan university taipei taiwan republic china



thesis functional domain titin molecule muchien sun national taiwan university  
taipei taiwan republic china publication wang sun mc jeng cj location cterminus  
region sarcomere biophysical research communication sm jeng cj sun mc study  
interaction titin myosin histology histopathology page bottom  
has\_stopwords: ['m']

Índice da linha: 1090

Texto original: ff biographical key page name michael gosland pharmd assistant  
year pacific university pacific stockton ca pharmd palo alto va hospital  
university residency stanford university school medicine pharmacology research  
professional experience expe employment experience professional experience honor  
clinical pharmacist adjunct professor university pacific school pharmacy palo  
alto veteran administration medical center palo alto ca b department veteran  
affair performance award exceeding performance requirement clinical pharmacy  
specialist adjuvant faculty c cancer society certificate merit oncology group  
swog leukemia subcommittee may present american association cancer research aacr  
preclinical pharmacology experimental therapeutic subcommittee may present  
american society hospital pharmacist ashp oncology special interest group  
present professor college pharmacy university kentucky lexington xy november  
joint appointment university kentuckylucille markey cancer center june present  
joint appointment che university kentucky department medicine division  
hematologyoncology j astra clinical pharmacy pharmacology young award  
bibliography gosland mp lum bl sikic bi reversal cefoperazone resistance  
doxorubicin vinblastine resistant human sarcoma colls cancer re gosland mp bl  
sikic bi modulation multidrug resistance relationship human glycoprotein proc  
cancer re kaubisch gosland mp lum bl sikic bi halsey j etiz b c correlation  
pharmacokinetics parameter toxicity cp pro soc clin pharmacol ther tum bl bi  
pharmacokinetic behavior intraperitoneal ip carboplatin pro soc clin pharmcol  
ther paper mp lun bl sikic bi modulation resistance structureactivity  
relationship human glycoprotein pro assoc cancer mp lum bl sikic bi modulators  
multidrug binding multidrug resistant sensitive cell proc assoc cancer rea  
gosland mp lum bl sikic bi superiority cyclosporina csa versus verapamil vpm  
modulating multidrug resistance mdr human celle pro clin pharmacol ther paper  
presented gosland mp dittman l lum bl mondan ce sikic th perfused rat liver ao  
model studying bffects modulators multidrug resistance excretion anticancer  
druga initial etoposide proc assoc cancer rea duplicate copy use  
has\_stopwords: ['re']

Índice da linha: 1093

Texto original: james r walker professor la chemistrybacteriology nj biochemical  
science professional experience research associate konrad bloch chem cambridge  
asm visiting professor professor microbiology professor professor microbiology  
dept bibliography expression walker ramsey w g haldenwang interaction z protein  
vivo ramsey coli j bacteriol region identification product mol gen genet walker  
coli dna fragment base pair length blinkowa j escherichia coli nucl acid re  
has\_stopwords: ['re']

Índice da linha: 1095

Texto original: fisher charles scientist education texas university college station tx biochemistry u georgia athens ga m university georgia athens ga professional experience research scientist university texas southwestern medical center dallas texas instructor university texas southwestern medical center dallas texas research associate case western reserve university department medicine va hospital cleveland ohio postdoctoral research associate case western reserve university department pediatrics metro general hospital cleveland ohio research associate northeastern ohio university college medicine rootstown ohio distinguished research award case western school medicine usda ad hoc grant reviewer american society biological chemist associate member sigma xi research society  
has\_stopwords: ['m']

Índice da linha: 1098

Texto original: senior research assistant m institution location degree study university texas b microbiology paso tx university colorado health m microbiology center denver co immunology research andor professional experience list alpha lambda delta freshman honor university texas el paso el paso texas list university texas paso performance numerous assay oil field microbial isolates petroleum el paso texas dr james zajic supervisor beta beta beta biology honor alpha chi scaior honor university texas el student gpa teaching research assistant cloning herpes hepatitis b viral gene instructor laboratory medical dental student suspended doctoral dissertation work cell receptor immunology department university colorado health sciences center denver colorado research associate applied tissue cell culture hybridoma technology elisa screening embryo microinjection transfer southem dna transfer mrna national jewish center immunological respiratory medicine denver colorado senior research assistant cell culture radioligand binding assay immunoassay ion flux assay recombinant dna toward characterization nicotinic acetylcholine receptor subtypes neurochemistry division neurobiology barrow institute arizona  
has\_stopwords: ['m']

Índice da linha: 1102

Texto original: biographical sketch key begm subramanian kiranur n professor wan univ madras madras india chemistry univ madras madras india msc chemistry indian inst science bangalore phd india ano professional experience n honor key way \_any government tree two page professor university illinois chicago research ing present associate professor university chicago research teaching assistant professor university illinois chicago research teaching research associate yale university new connecticut research fellow yale university new connecticut research research university michigan ann arbor michigan research teaching fellow mcmaster university hamilton ontario canada research teaching honor faculty research award fra american cancer society swabius cancer research award yale university p reddy medal sirish gold medal obtaining first rank bsc degree exam madras university government india merit scholarship publication pertinent one shown bergsma dj olive dm hartzell subramanian territorial limit functional anatomy sv replication origin natl acad sci usa byrne bj davis m yamaguchi j bergsma dj subramanian kn definition early promoter region

demonstration host range bias enhancement effect sv bp repeat proc natl acad  
usa j subramanian kn deletion mutant lacking bp repeated sequence viable  
noncomplementable deficiency nucleic re  
has\_stopwords: ['m', 're']

Índice da linha: 1108

Texto original: richard isaac current status education summer summer near vitae  
asst prof med cellular laboratory jackson mass gen hosp boston usa citizen  
birthdate october bsbiology university california irvine bachemistry university  
california university ington dept university washington dept microbiology  
student research univ calif irvine dr eb fleischers laboratory bioinorganic  
chemistry porphines student teaching assistant univ calif irvine  
instructionsupervision organic chemistry laboratory atomic energy commision  
summer trainee iowa state univ ames iowa research dr d laboratory structure  
platinum palladium complex inorganic crystal student research univ calif irvine  
dr eb laboratory bioinorganic chemistry sulphonated porphines graduate study  
hchain rna sequence mouse thymocytes supervisor dr ursula storb cancer  
immunology training grant research antibody dependent cellular cytotoxicity  
research assistant dr u storb transcription tlymphocytes national research  
service award molecular cellular biology training grant dr u storb research  
assistant dr u storb spring winter teaching assistant duty technique dna  
electrophoresis postdoctoral research control idiotypic expression dr ml gefer  
dept biology cancer society postdoctoral fellowship national research service  
award inst allergy infectious disease joint appointment assistant mass general  
hosp instructor harvard med school  
has\_stopwords: ['d']

Índice da linha: 1109

Texto original: biographical sketch title rita professor institution field  
brandeis university waltham ba biology university pennsylvania philadelphia pa  
immunology professional experience kay tho key master ina membership federal  
exceed employment usphs predoctoral trainee univ c doherty supervisor m society  
postdoctoral scholar koprowski director nih ucla dept microbiology benjamin  
bonavida assistant research pathologist departmont pathology ucla medical  
assistant professor department ucla medical school department pathology ucla  
medical department pathology ucla medical cent honor university national scholar  
announced one frequentlycited immunology new investigator research award  
resoarch career development award nih study epidemiology disease control oct rfa  
review cause consequence thymic nov board journal immunology publication rom  
dohery pc gerhard w bennink j generation crossreactive cell immunization virus  
exp med listed current one cited paper survey pc efftos bennink cytotoxic cell  
response immunization virus proc sci usa biddison snodgrass eros doherty  
induction recognized prior target cells tnp exp med doherty pc biddison model  
immune system icnucla symposium molecular vol p etfros doherty pc secondary cell  
response mouse infected virus cell immunol doherty pc bennink j immunity virus  
virus perspective raven new york pp rb doherty pc influenza early appearance  
crossreactive lung mouse primed heterologous type immunol frankel w doherty pc  
function virus hybridoma antibody j frankel gerhard w monoclonal virat block

effector cell operating j pc bennink frankel dual cell functional indication  
virus h molecule may associate cell membrane trend ferrone plenum press york  
has\_stopwords: ['m']

Índice da linha: 1111

Texto original: fryer biographical sketch key gereid md school end mayo  
institution michigan ann arbor mt with high university school ann arbor mi cum  
leude medicine experience honor key personne wil level federal oo two rotating  
general internship philadelphia general hospital pa rated flight surgeon united  
state fir force whitman force mo residency ta internal medicine university  
jackson memorial hospital fl allergy clinical immunology university rochester  
medical center rochester ny assistant physician strong memorial hospital  
rochester ny instructor microbiology university rochester school dentistry  
rochester ny instructor medicine microbiology mayo graduate schoo medicine  
rochester medicine allergy mayo clinic rochester mn assistant professor medicine  
microbiology mayo graduste school rochester mn associate professor medicine  
immunology microbiology mayo medical school rochester mn professor medicine  
immunology mayo medical school rochester department immunology mayo rochester mn  
professor immunology medicine mayo medical school rochester mn present ed j bd  
board j allergy clin mbr allergy immunol allergy infect fellow acad allergy  
fellow coll assoc rheum mbr central soc clin re mbr fed clin re advancement mbr  
assoc phys diplomate board int diplomate board allergy immunology john sheldon  
steve lang memorial lectureship stoll lecturer soc david talmage memorial  
lectureship sth aspen allergy cont mbr board scientific counselor allergy infect  
immunological study sectionnih recipient nih merit award landmark allergy award  
reciplent e total filley wv holley ke kephart gm gleich gj identification  
eosinophil granule major basic protein lung tissucs patient bronchial asthma  
lancet july ff  
has\_stopwords: ['re']

Índice da linha: 1114

Texto original: biographical sketch name manuela martinsgreen education  
university california davis phd zoology university california berkeley lbl mol  
university california berkeley lbl nrса fellow mol biol laboratory cell  
molecular biology lawrence berkeley laboratory berkeley ca mentor mina j bissell  
service aw laboratory cell molecular biology lawrence berkeley laboratory  
berkeley ca mentor mina j bissell adjunct assistant professor laboratory  
molecular oncology rockefeller university ny head laboratory hidesaburo staff  
research scientist faculty series cell molecular biology division lawrence  
berkeley laboratory berkeley ca head department mina j bissell assistant  
professor biology university california riverside ca lat honor fulbright travel  
grant graduate study usa departmental fellowship department plant pathology uc  
riverside elected phi kappa phi nih traineeship cellular molecular training  
grant uc davis regent fellowship uc davis graduate research award uc davis  
jastroshields research award uc davis nrса postdoctoral fellowship nih lawrence  
berkeley laboratory biography included american men woman science biography  
included who west associate editor journal molecular oncology regent faculty  
fellowship uc riverside biography included who world biography included who

medicine healthcare american cancer society society celi present american  
society developmental biology present electron microscopy society america  
northern society electron microscopy woman cell biology present healing society  
present  
has\_stopwords: ['who']

Índice da linha: 1116

Texto original: biographical sketch give following information key personnel  
collaborator begin with the principal director photocopy page person carol c  
korte research technician education begin education nursing institution location  
northern university dekalb il m biolog professional noy two page microscopy  
present experience negative staining preparation tissue sectioning viewing  
photographic development negative print data analysis quantitation morphological  
character included digitizing point counting measurement experience design  
implementation experiment various analysis tool tissue culture present routine  
maintainance cell without contamination molecular biology present including  
routine subcloning dna preparation restriction analysis sequencing library  
screening cell biology present immunocytochemical staining drosophita tissue  
followed light microscopy digital imaging analysis research assistant dr l yasui  
illinois univeristy dekalb il cell survival experiment electron microscopy  
morphological technologist chief virology technician pennsylvania state  
univeristy animal diagnostic laboratory assistant present half time dr c  
department biology pennsylvania state university assistant present half time dr  
thomas department pennsylvania state university pa honour member sigma xi place  
student competition held midwest society microscopist logy publication yasui  
morphological characterization radiation sensitive cell scanning microscopy  
has\_stopwords: ['m']

Índice da linha: 1118

Texto original: group edward principal andor professional ny research university  
chicago discovery radioactivity decay energy radioactive halflife technical  
operation officer armed force special weapon project depart ment defense  
planning direction laboratory field experiment nuclear effect radioactive  
fallout nevada test site eniwetok associate enrico fermi institute study  
university heading research group study radioactive strontium air plant dairy  
product food animal human bone specimen chief research group head g air force  
cambridge research bedford mass experimental study atmospheric aspect global  
radio active fallout nuclear test present research group leader center  
atmospheric research study end lowlevel alpha spectroscopy absolute counting  
interaction radioactive aerosol atmosphere biosphere property dynamic indoor  
radon progeny aerosol honor postdoctoral appointment enrico fermi institute  
nuclear study chicago delegate united nation scientific committee effect atomic  
radiation president international commission atmospheric chemistry radioactivity  
iamapiugg membership health physic international radiation protection  
association union fellow aaa sigma ac diviston nuclear chemistry technology  
representative recent publication poet martell americium contamination denver  
health physic poet moore ea martell lead bismuth polonium atmosphere accurate  
ratio measurement application aerosol residence time deter mination geophys re

has\_stopwords: ['re']

Índice da linha: 1119

Texto original: curriculum pi z academic background institutic hebrew university  
jerusalem biochemistry hebrew university jerusalem biochemistry university  
rochester biophysics b major research interest structure molecular mechanism  
transport animal cell membrane new biophysical tool studying biological  
transport biochemical delineation ot altered transport mechanism pathological  
condition iron iron overload malaria iron chelators antimalarial c previous  
employment r national inst membrane transport hebrew university membrane  
biophysics university rochester visiting as prof membrane biochem university  
jerusalem professor membrane biophysics institute health cfni fibrosis  
university jerusalem selected publication cabantehik z eidelman cabantchik  
continuous monitoring transport fluorescent vesicle biophys acta review biomembr  
eidelman cabantchik fluorescence method continuous monitoring transport meth  
enzymol erythrocyte band structure function subcellular biochemistry vol blood  
edpp cabantchik transport domain erythrocyte anion exchanger j membrane biol p  
englert hc lang hg greger r conjugate new tool probing topography anion  
transport protein cell cc shanzer libman j glickstein h lytton reversed  
siderophores agent proc natl acad lytton cabantchik j shanzer nbddfo chelator  
antimalarial agent mol pharmacol ed new methodology study transport across ell  
membrane plenum press london cabantchik greger r chemical probe anion  
transporter mammalian cell j physiol c breuer wv riordan jr modulation cftr j  
biol chem

has\_stopwords: ['as']

Índice da linha: 1124

Texto original: biographical sketch dr andrew stolz date november college boston  
universi boston ba cum laude independent concentration chemistry medical school  
albert einstein college medicine bronx ny internship montefiore hospital bronx  
ny july june residency july june junior ine fellowship uclawadsworth fellowship  
program los angeles ca july gastroenterology award june american liver  
foundation postdoctoral fellowship july june associate investigator award va  
career development program july research associate award va career development  
program september westem society clinical investigation academic appointment  
assistant professor medicine step il unive los angeles assistant professor  
medicine step university california los angeles assistant professor medicine  
step iv university california angeles associate professor medi ty california  
university southem california publication publication yamada sugiyama belknap w  
kaplo quantitation characterization newly described orgat radioimmunoassay  
biochim biophys acta j n kda bile acid binder human hepatic sugiyama stolz  
sugimoto evidence common high affinity binding si ithocholic acid lipid re stolz  
sugiyama kublenkamp j osadchey b yamada belknap w w glutathione b molecular form  
hepatology stolz takikawa h sugiyama j kaplowitz dehydrogenase activity bile  
acid binder ra identification kinetics physiologi bile acid catalyzed rat  
hepatic study purified enzyme isolated hepatocytes invest takikawa h ookhtens  
stolz kaplowitz

has\_stopwords: ['re']

Índice da linha: 1126

Texto original: interagency committee smoking health national advisory committee meeting october speaker terence c golden mr golden general service prior served assistant secretary department treasury previously partner crow texas mr golden received b degree engineering notre university m nuclear massachusetts institute technology master business administration harvard graduate school business barbara hulka currently professor department epidemiology school public health north carolina chapel hulka served university since previously held position associate professor also worked federal government consulting committee chaired passive smoking national prior also includes position board regent american college preventive medicine well seat fertility maternal health drug advisory fda hulka earned master degree columbia school medicine also holds degree juliard school music pinney mr pinney founding director office smoking health executive director institute study smoking behavior policy john f kennedy school government harvard university mr pinney also general partner corporate health policy group washington received ba degree history yale university  
has\_stopwords: ['m']

Índice da linha: 1132

Texto original: biographical sketch give key personnel collaborator begin principal director photocopy page person position yong q chen assistant professor nursing year institution location fudan university china b biology free university brussels brussels belgium molecular biology state detroit ml professional with key tho page professional\_experience assistant professor dept pathology wayne state research associate dept radiation oncology wayne state university research assistant dept molecular biology free univ brussels brussels honor solvary award january brussels ional metastasis research association cancer research b chen j j cancer deforesta f hertoghs avalosse b comelis j rommelaere j selective killing simian virus transformed human fibroblast parvovirus h cancer re hilgers g chen j rommelaere j deficient expression enhanced parvovirus ataxia telangiectasia xrays uv light carcinogenesis b chen q j duponchel n p namba j parvoviral dna function hostcell transformation accomplishment oncology chen q tuynder comelis j boukamp p fusenig rommelaere j sensitization human keratinocytes killing parvovirus h take place malignant transformation require carcinogenesis hilgers g abraham p j chen q schouten cornelis j j van der eb j impaired recovery mutagenic soslike response ataxia telangiectasia cell mutagenesis avalosse b l chen cassiman j j rommelaere j nuclear protein normal human fibroblast specifically interacts minute virus mouse correlate cell resistance parvovirus molecular carcinogenesis comelis j chen spruyt n duponchel n cotmore tattersall j differential human cell killing autonomous parvovirus mvm h regulated al transcriptional level j virology chang y yq chen la fitzgerald kv honn analysis integrin mrna human rodent tumor cell biochem biophys re commun honn k v grossi chang chen lipoxigenases integrin receptor metastasis eicosanoids bioactive cancer radiation injury kvhonn ed pp kluwer academic publisher boston  
has\_stopwords: ['y', 're']

Índice da linha: 1135

Texto original: grant karin biographical sketch anning lin begin institution  
location college nanking university nanking chine biochemistry university  
alabama pharmacology university california san diego field teaching assistant  
department biochemistry nanking university china graduate student department  
pharmacology university alabama birmingham pres researcher department  
pharmacology university san diego membership chinese society biochemistry  
publication abstract lin chao sh cheng lt change calmodulin level phospho  
activity development chick brain chinese biochem j lin wallace rw partial  
purification characterization inositol trisphosphate kinase human platelet paseb  
j lin barnes wallace rw phosphorylation protein kinase inactivates inositol  
trisphosphate kinase purified human platelet biochem biophys re commun lin  
wallace rw barnes identification kda polypeptide inositol kinase human platelet  
faseb j lin barnes wallace rw property inositol trisphosphate kinase human  
platelet submitted lin frost j deng smeal kikkawa u hunter brenner karin kinase  
ii negative regulator cjun dna binding ap activity coll submitted albert deng  
lin meinkoth jl schonthal mc karin feramisco jr protein phosphatase  
differentially regulates promoter containing sre tre element j gell biol  
submitted  
has\_stopwords: ['re']

Índice da linha: 1136

Texto original: curriculum vitae name philip hultgren date birth july hartford  
connecticut social security number mailing address dept physiology biophysics  
university vermont college burlington vermont assistant lthaca college ny  
research assistant university philadelphia pa physiol postdoctoral fellow  
university vermont burlington vt physiol biophys research assistant professor  
university vermont vt physiol biophys honor university fellowship summer temple  
university biomedical research grant fellow temple university agan research  
vermont affiliate american heart association new investigator research award nih  
membership american heart association council basic science international  
society heart research publication article thesis pb development individually  
prescribe exercise initial training session master thesis ithaca college ithaca  
ny burke ej hultgren pb physician future able prescribe exercise j med education  
hultgren pb burke ej methodology prescription exercise review austral j sport  
med hultgren pb effect volume left ventricular hypertrophy myocardial blood flow  
rest doctoral dissertation university school medicine philadelphia pa rove aa  
hultgren pb ritzer tf carey ra myocardial blood flow response exercise training  
dog j appl physiol pb bove aa myocardial blood flow mechanic volume  
overloadinduced left ventricular hypertrophy dog cardiovasc re  
has\_stopwords: ['re']

Índice da linha: 1139

Texto original: ff biographical sketch provide following infermation key form  
page photocopy position title zheng cui assistant professor educationtraining  
begin postdoctoral degree institution location field study medical college  
tsuenyi md medical kweichow province china shanghai institute entomology m  
entomology academy science shanghai china university massachusetts amherst phd



biochemistry research professional experience concluding previous employment  
 exceeds two page pertinent publication two page postdoctoral fellow molecular  
 biology signal transduction mammalian cells department medicine harvard medical  
 school boston massachusetts research associate molecular biology lipid  
 metabolism tumor development growth control liver lipid lipoprotein research  
 group department biochemistry university alberta edmonton canada assistant  
 professor department biochemistry associate cancer biology bowman gray school  
 medicine winstonsalem north carolina present member comprehensive cancer center  
 wake forest university bowman gray school medicine winstonsalem north carolina  
 inors award university massachusetts graduate study scholarship american heart  
 association postdoctoral fellowship publication cui z hsu study insect hormonal  
 regulation glutamatealanine transaminase glutamateaspartate transaminase bombyx  
 mori linnaeus contr shanghai inst entomol influence cui z hsu purification  
 characterization lalanineglyoxylate transaminase fat body philosamia cynthia  
 ricini donavan hormonal impact activity vivo contr shanghai inst entomol nordin  
 j h cui z yin coldinduced glycerol accumulat developmentally regulated j insect  
 physiol ostrinia nublilis larva cui z mason single nucleotide substitution aib  
 locus yeast mitochondrial gene rrna confers resistance erythromycin  
 coldsensitive ribosome assembly curr genet cui z zubiaur bloch b michel seidman  
 j g neer e expression g protein subunit balbc cell lead agonistspecific change  
 growth regulation j biol chem vance cui z debate phospholipid golgi complex  
 interpreting blocking pc biosynthesis trend cell biol vance j e chen voelker r  
 vance cloning expression novel phosphatidylethanolamine specific biochemical  
 cytological marker unique membrane fraction rat liver j biol chem page number  
 throughout b  
 has\_stopwords: ['m']

Índice da linha: 1140

Texto original: ff pringipal biographical sketch give key md phd bauer e sumpio  
 md phd position professor su ana institution ano location hopkins baltimore  
 maryland chemistry cornell university medical college new york new york  
 physiology ano professional present federal pubic year two page professional  
 general surgery residency yale univ sch medicine vascular surgery fellowship  
 univ hill assistant professor surgery yale univ school medicine associate  
 professor surgery yale univ school medicine honor j wylie travelling fellow  
 vascular surgery schering scholar american college surgeon society surgeon  
 fellow american society shock society society nephrology association surgeon  
 federation clinical research society vascular medicine biology international  
 society sugery new england society vascular surgery society university surgeon  
 surgical biology club society american heart association society cell biology  
 engineering society maack kinetics competition selectivity tubular absorption j  
 physiol f sumpio kirkwood j urinary interferon microglobulin interferon therapy  
 cancer re sumpio camargo maack kinetics renal catabolism absorbed protein  
 influence lysosomal ph contribution nephrology chaudry baue ae nuclear magnetic  
 resonance study showing injury following forum sunpio chaudry clemens mg baue  
 accelerated recovery isolated rat kidney warm ischemia j physiol camargo mjf  
 maack renal hydrolysis absorbed protein influence load lysosomal ph j physiol  
 sumpio chaudry ae aneliorates reperfusion injury following determined arch surg

sumpio chaudry baue reduction druginduced nephrotoxicity effect isolated  
perfused j surg re sumpio chaudry ih baue ae reduction druginduced  
nephrotoxicity effect perfused kidney j surg re ph rev page ff sage atthe  
has\_stopwords: ['re']

Índice da linha: 1142

Texto original: jerome bailey asst re april university rochester rochester new  
york b biochemistry university delaware newark delaware phd chemistry position  
research fellow division biology research city hope duarte ca research fellow  
division immunology beckman research institute city hope duarte ca present  
assistant research scientist division immunology beckman research institute city  
hope duarte publication jerome bailey bailey jm shively je carboxyterminal  
sequencing solid phase technique protein chemistry il automation application  
academic press bailey lin ln brandt ma proline alanine hinge region  
phosphoglycerate kinase j protein chem colman bailey jm decamp huang vollmer  
affinity adenine nucleotide site enzyme procnew york acad sci bailey shively  
carboxyterminal sequencing formation hydrolysis cterminal biochemistry grant pg  
decamp bailey colman colman rw low km phosphodiesterase human platelet  
stimulation activity phosphorylation affinity labeling active site advance  
second messenger phosphoprotein research press grant pg decamp dl bailey jm  
colman rw colman three new potential affinity label inactivation human platelet  
low km camp phosphodiesterase cyclic biochemistry ma bailey jm resplendor  
sitedirected mutagenesis histidine hinge region yeast phosphoglycerate kinase  
effect catalytic activity activation sulfate biochemistry  
has\_stopwords: ['ma', 're']

Índice da linha: 1145

Texto original: douglas biographical sketch information forthe key personnel  
consuitants collaborator begin director photocopy page douglas head section  
cerebrovascular surgery begin baccalaureate education nursing inclide degree  
institution ano location year fielo study northwestern university em hpme honor  
program medical education medicine mayo minnesota postgraduate professional  
experience year wo page exceed two page honor mention academy research award  
competition american academy neurological surgery national foundation march dime  
summer research grant medical student rag fund response nervous system neoplasia  
yale medical school surgical grant ccf grant experience pediatrics university  
colorado health science center residency mayo clinic resident neurosurgery mayo  
fellowship dr sundt clinic aj strong new castle uk professor surgery  
neurosurgery yale school medicine director neurosurgical cerebrovascular disease  
yale school medicine chief neurosurgery west vamc certified american board  
neurosurgery professor surgery yale school medicine professor surgery university  
head cerebrovascular surgery section cerebrovascular surgery cleveland  
foundation publication anesey j scott pg veis chyatte soluble type collagen  
precursor rat skin biochem biophys re commun diagnostic importance normal  
finding n eng j med chyatte chyatte c note abortion visual closure j psychol  
sundt tm cerebral vasospasm subarachnoid clinic proceeding chyatte c pain  
sensitivity reduced mild cutaneous electrostimulation psychologica rusch n sundt  
tm prevention chronic experimental cerebral vasospasm ibuprofen highdose

methylprednisone j neurosurg chyatte sundt tm response chronic experimental  
cerebral vasospasm dexamethasone j neurosurg rao gjs chyatte nadler hl  
enhancement glycoprotein human skin fibroblast cationic polypeptide et  
biophysica acta romero sd chyatte byer de romero jc yaksh tl measurement  
prostaglandin cerebrospinal fluid cat dog man j neurochem rusch nj chyatte sundt  
tm vanhoutte pm source activator calcium human basilar artery stroke ph  
has\_stopwords: ['re']

Índice da linha: 1147

Texto original: biographical sketch name position philip trackman phd research  
assistant professor biochemistry education year field institution conferred  
study college wooster wooster ohio chemistry boston university boston  
biochemistry brandeis university waltham research andor professional experience  
present associate member american chemical society american cancer society  
postdoctoral fellow staff scientist novo laboratory wilton ct group leader novo  
laboratory wilton ct present research assistant biochemistry boston university  
schocl medicine boston publication jordan r e milbury p sullivan k trackman p c  
kagan h study oxidase bovine ligamentum nucha bovine aorta adv biol med p c  
kagan h amine inhibitor substrate oxdase j biol chem kagan h tseng l trackman p  
c k rapaka r urry w repeat polypeptide model elastin substrate oxidase j bicl  
chem trackman p c zoski c g kagan h development peroxidasecoupled fluorometric  
lysyl oxidase anal biochem p c abele r h biochem biophys re commun abele r h  
synthesis deoxymethylthioadenosine activity identification ribulose j biol chem  
tang trackman p c kagan reaction bovine aortic lysyl oxidase biol chem  
has\_stopwords: ['re']

Índice da linha: 1155

Texto original: sketch timothy mather technical assistant research program  
oklahoma medical research foundation birth education michigan state university  
physiology university michigan biochemistry professional present technical  
assistant research oklahoma medical research foundation research worker  
department medicine columbia university direction dr karen kaplan instructor  
college il research assistant biochemistry university michigan instructor  
chemistry hermann lietz schule schloss homenwehroa west germany laboratory  
technician department biochemistry university michigan research assistant  
evanston hospital evanston mather aminoff rheology blood patient malignant clin  
re abstr mather aminoff purification exoglycosidoses cl penfringens fed proc  
has\_stopwords: ['re']

Índice da linha: 1159

Texto original: tobacco research advisory board meeting new york new york people  
presenc aboverefererced meeting advisory richard bing roswell sab gordon sato  
joined sab sheldon somers hobbs president vice president joined ctr staff  
director robert vice research director ford research director vincent director  
stone research director leake technical zahn drumond candidate sab shown  
paragraph betow ma obtained fiftyfour application reservation rating deferred  
pending site new ctr sab business heard positive  
has\_stopwords: ['ma']

Índice da linha: 1162

Texto original: director sketch give information key snd collaborator begin principal director photocopy page name position title thomas jue professor education institution study university ba california state university hayward ba chemistry university california davis chemistry professional experience concluding order employment order title author pest three year publication applicetion exceed two page postgraduate research associate department chemistry university california davis research scientist molecular biophysics biochemistry university honor graduate opportunity fellowship university california davis u patent apparatus method determining presence substance sample nmr producing thereof publication recent total bogan j jue computer controlled indirect detection experiment j mag re jue winnowing resonance vivo nmr spectrum j mag re strategy sift metabolite relaxation rate vivo nmr spectrum j mag re two one simultaneous winnowing signal using h pulse j re kay l jue b bangerter p sensitivity enhancement measurement via polarization transfer j mag re jue jab lohman r ordige rg shulman natural abundance nrm spectrum glycogen human mag re med decoupling interaction edited spectrum j mag re jue chung rg shulman measuring redox perfused liver nrm mag re hanstock cc rothman rg shulman proton spectroscopy human brain j mag ph rev page page ff page consecutively atthe bottom throughout use auch b  
has\_stopwords: ['re']

Índice da linha: 1163

Texto original: preliminary application peter j good biographical sketch assistant professor year field study state university ames ia b towa state university ames zoology univ wisconsin madison wi phd molecular cellular biology lmg nichd nih bethesda md developmental biology ri staff fellow laboratory molecular genetics national institute assistant professor department biochemistry biologylsu medical center shreveport la award service nih trainee program molecular cellular biology uni intramural research training award bethesda good pj welch w ryu jb late spliced rna simian virus synthesized common pool transcripis j virol good pj rc welch barkan somasekhar je mertz vp vp translated independently spliced late rna specie simian virus j virol richter k pj good dawid developmentally regulated nervous xenopus encodes putative rna binding protein new biologist robert ab p kondaiah f rosa watanabe p good daniclpour n roche rebbert lb dawid sporn mesoderm induction xenopus distinguishes various tge isoforms growth factor good pj k richter lb dawid nervous isotype beta subunit nat expressed early development xenopus laevis proc natl acad sci horisberger jd p jaunin pj good bc rossier geering coexpression subunit result functional nak pump xenopus oocyte acad taira jamrich pj good dawid lim homeobox gene expressed specifically organizer region xenopus gastrula embryo gene development dawid lb taira pj good rebagliati role growth factor embryonic induction xenopus laevis mol rep dev good pj mr rebbert dawid three new member rnp protein family xenopus nucleic acid re good pj conserved family elavlike gene vertebrate proc natl acad sci knecht ak good lb dawid rm harland patterning differentiation noggininduced tissue absence development wisconsin madison wi

has\_stopwords: ['re']

Índice da linha: 1171

Texto original: miller biographical sketch following information forthe key j  
brenner assistaat institution location study oxford university uk philosophy  
university london medical college radiation physic st bartholomew hospital  
university surrey uk physic prof experience concluding employment experience  
present federal public advisory com plete reference ali year pertinent  
application oo exceed two page technician department electrical university  
london department anesthetic oxford fellowship los alamo laboratory nm member  
los alanos national laboratory research radiological reveucch laboratory college  
physician surgeon columbia university york ny assistant professor department  
radiation radiological research laboratory college physician surgeon columbia  
univ icrptcru quality factor member nci site review university rochester member  
iaea advisory group nuclear data radiobiology member program committee th annual  
meeting research society la vega university senate columbia university honor  
oxford university carter physic prize young award fourth international  
conference aspect nuclear structure nuclear reaction winnipeg young scientist  
fellowship seventh international nci new investigator research avarad radiation  
research selected recent publication brenner j quality factor radiat re brenner  
dj stochastic treatment time decay specie created heavy chargedparticle radiat  
prot dosimetry goodhead dt brenner dj bance chmielewski jj bradbury jn dosimetry  
comparison characterisation al k ultrasoft xray mrc coldcathode source phys blol  
zaider brenner evaluation specitic quality function human fibroblast rad prot  
dosim ph page de use

has\_stopwords: ['re']

Índice da linha: 1174

Texto original: biographical name position title birthdate beng ho phd professor  
neuropharmacology national taiwan university taiwan b chemistry university  
oregon eugene oregon m chemistry university washington wash chemistry state  
university new york buffalo post doe chemistry academic appointment chief  
neurochemistry neuropharmacology section texas research institute mental science  
houston texas assistant professor mental science university texas graduate schoo  
biomedical science gsbs houston texas associate professor mental science  
university texas gsbs houston texas present mental science university texas gsbs  
houston texas professor neuropharmacology department neurooncology medica  
oncology university texas md anderson cancer publication past year excluding  
abstract taylor mathew rj ho et al serotonin level platelet uptake premenstrual  
tension neuropsychobiology taylor dll silverman ho effect carboline serotonin  
binding rat brain pharm pharmacol samorajski hartford kralik p ho smith change  
serum prolalin growth hormone monoamine oxidase acetylcoenzyme hydrolase age  
hydergine brit clin prac research progress dementia suppl khan ho davis al trace  
element level human alcoholic brain alcohol khan major et al trace element  
disulfiram metal ion gabay j harris bt ho ed pp new york alan li englert ho  
passive avoidance deficit mouse following ethylcholine aziridinium chloride  
treatment pharmacol biochem behav ho bt evaluation two distinctive  
betacarbolines serotonin binding human platelet re comm chem path pharmacol ho

ho et al correlation plasma cortisol csf  
has\_stopwords: ['m', 're']

Índice da linha: 1175

Texto original: ff principal investigatorprogram director biographic sketch  
james zhijun lu postdoctoral fellow education location study central university  
b biology agriculture china wuhan university china m biochemistry beijing normal  
university mol biology professional experience present postdoc fellow dr dennis  
lab university rochester currently working replication regulation human  
papillomaviruses b focusing role two major open reading frame product e e using  
molecular biology technique including cloning gene structure analysis vivo gene  
expression dnaprotein interaction postdoc fellow dr lan lab university rochester  
following work carried high expression purification gtp binding protein  
potential cellular signal transducer coli development antibody raba mutagenesis  
cdna change codon order decrease gtpbinding ability expression wildtype  
mutagenized form raba pc cell respectively functional analysis signal  
transduction effect cell differentiation phd student national lab cell biology  
beijing normal university phd thesis included antisense human decreased  
endogenous expression level human stomach cancer cell line overexpression  
induced transformation normal niht cell establishment cell line inducibly  
expressing vharas  
has\_stopwords: ['m']

Índice da linha: 1177

Texto original: ff director biographical sketch give ine information key  
consultant listed page begin principal page zbigniew darzynkiewicz md phd may  
cation year fie university school medicine warsaw poland highest honor  
university school medicine phd cell biology warsaw poland research ano experienc  
employment present federa government advisory chronological ang com dunng year  
representative earier ti exceed two page resident surgery warsaw poland  
predoctoral fellow dept histology schoo medicine warsaw poland research  
associate pharmacology suny buffalo ny sr asst prof dept histology school  
medicine warsaw poland visiting scientist exp cell research laboratory medical  
nobel institute karolinska institutet stockholm sweden staff scientist boston  
biomedical research institute boston mass asst member sloankettering institute  
new york ny asst prof cornell univ graduate school med sci new york ny assoc  
member sloankettering institute new york ny assoc prof univ grad school med sci  
new york head exp cell re laboratory sloankettering institute member institute  
new york ny professor cornell univ grad school med sci new york ny present nasa  
consultant flow cytometry space station director flow cytometry core facility  
network sloankettering institute new york ny professor medicine new york medical  
college valhalla ny professor microbiology immunology new york medical college  
valhalla ny professor experimental pathology new york medical college valhalla  
ny director cancer research institute new york medical college ny president cell  
kinetics society presidentelect international society analytical cytology north  
american editor cell proliferation associate editor editorial board member  
cytometry jhistochem cytochem leukemia leukemia re dr darzynkiewicz publication  
field cell biology cytochemistry cancer research past three year published paper

following representative publication bibliography del bino g skierski  
darzynkiewicz z diverse effect camptothecin inhibitor cycle lymphocytic molt  
myelogenous

has\_stopwords: ['re']

Índice da linha: 1179

Texto original: john p sundberg university vermont burlington vt b purdue  
university west lafayette university connecticut storrs ct credential diplomate  
american college veterinary pathologist current position staff scientist jackson  
laboratory bar harbor relavent publication sundberg jp junge lancaster wd  
immunoperoxidase localization papillomaviruses hyperplastic neoplastic  
epithelial lesion animal j vet re sundberg jp williams e hill lancaster wd  
nielsen papillomaviruses cutaneous fibroma whitetailed mule deer j vet re mk  
reichmann sundberg jp cloning characterization equine cutaneous papillomavirus  
virology sundberg jp mk schmidt didier reichmann mb cloning characterization  
canine oral papillomavirus j vet re sundberg jp mk reichmann mouse  
papillomavirus pathology characterization virus cancer cell press sundberg jp  
papillomavirus infection animal gissmann l syrjanen k kos l ed springerverlag  
heidelberg west germany pre sundberg jp animal model papillomavirus research  
hofschneder ph munk k ed virus human tumor basel switzerland pre

has\_stopwords: ['re']

Índice da linha: 1180

Texto original: principal investigator program biographical sketch kay personnel  
with the photocopy paye tre re associate institution location university india  
osmania university india university india professional experience present  
federal ca ine yea two page profession lecturer zoology university hyderabad  
india research associate dept college amherst mass research fellow british  
industrial biological research association carshalton surrey england research  
associate fellow honorary lecturer biochemistry king hospital medical school  
london re assistant professor pathology albany albany new york re associate  
pathology albany college albany new york assoc prof dept pharmacology rush  
medical college chicago publication representative koh synthesis polyadenosine  
diphosphate ribose swine tissue k dna aorta sorta janakidevi k heparin removal  
lysinerich histone fraction poly polymerase dna polymerase template activity  
isolated swing aortic nucleus exp cell re thomas wa reiner k janakidevi  
florentin lee dynamic arterial cell atherosclerotic lesion black heterozygous  
dehydrogenase molec path janakidevi k state chromatin poly adpribosylation  
regulatory protein dev bioll pp murray k janakidevi rna polymerase protein  
kinase activity forum aortic nucleus swine fed aash cholesterol diet biochem med  
lee kt k janakidevi reiner property dehydrogenase hybrid hare hybrid hafe model  
special reference atherogenesis exp molec path

has\_stopwords: ['re']

Índice da linha: 1182

Texto original: angie rizzino professor july stony brook b biology suny stony  
brook m biochemistry stony brook phd molecular biology postdoctoral fellow dr  
brian mccarthy training molecular biology eukaryote dr gordon sato training cell

biology research fellow roche institute molecular biology collaborated dr  
 michael sherman development mouse embryo serumfree medium research biologist  
 university california san diego research focused growth differentiation  
 embryonic cell expert national cancer research focused production growth factor  
 growth factor receptor embryonic professor eppley research cancer allied disease  
 research focused growth factor receptor professor eppley institute research  
 cancer allied disease biology growth factor receptor professor eppley institute  
 research cancer allied disease molecular biology growth factor publication  
 partial list publication levy bv distribution polyadenylated rna sequence  
 drosophila cell exp cell re rizzino sato g growth embryonal carcinoma cell  
 serumfree proc nail acad sci usa rizzino sherman mi cultivation early mouse  
 embryo serumfree culture medium exp cell re c differentiation embryonal  
 carcinoma cell line f defined medium proc natl acad sci usa rapp ur dome  
 formation retrovirus induced lung adenocarcinoma cell line cancer re rizzino two  
 multipotent embryonal carcinoma cell line irreversibly differentiate defined  
 medium dev biol rizzino orme l delarco embryonal carcinoma cell growth  
 differentiation production response molecule transforming growth factor activity  
 exp cell re rizzino fibronectin laminin promote two multipotent embryonal  
 carcinoma cell line spring cell ed lm silver gr martin strickland pp cold harbor  
 laboratory cold spring harbor  
 has\_stopwords: ['m', 're']

Índice da linha: 1187

Texto original: biographical sketch th folowing key begin withthe robert dale  
 nelson associate professor begin wih aa institution university wisconsin river  
 fall biology university minnesota minneapolis human genetics human genetics  
 minnesota minneapolis university minnesota minneapolis postdos research  
 professional experience wit al win ina present federal government year oo two  
 page professional experience instructor surgery university minnesota assistant  
 professor department surgery university associate graduate faculty department  
 uom consultant surgical research laboratory st med center associate professor  
 departuent surgery university minnesota associate professor department  
 dermatology university minnesota honor freshman chenistry university wisconsin  
 national health fellowship national institute health research career year  
 university wisconsin river fall fall publication fast dj pm schlievert rd nelson  
 nonpurulent response toxic shock syndrome producing staphylococcus  
 toxinstimulated production tumor necrosis factor j immunol vercellotti gm hq yin  
 gustafson rd nelson h jacob platelet activating factor prime neutrophil response  
 agonist role promoting neutrophilmediated endothelial damage blood mccormack rt  
 rd nelson l solem tw lebien decreased expression common acute lymphoblastic  
 leukemia antigen callacd neutrophil patient thermal injury brit j haematol  
 nelson rd herron agarose method method enzymology immunochemical technique di  
 sabato ed academic press ny wright cd rd nelson candidacidal activity  
 myeloperoxidase characterization myeloperoxidaseyeast complex formation biochem  
 biophys re fast dj pm rd toxic shock syndrome associated staphylococcal  
 pyrogenic toxin potent inducer tumor necrosis factor production infect immun  
 nelson rd hasslen dh ahrenholz ld solem polymorphonuclear leukocyte function  
 following burn mechanical injury regulation kinetics con trauma shock sepsis e



faist j green ed springerverlag berlin heidelberg pp page age  
has\_stopwords: ['re']

Índice da linha: 1191

Texto original: sketch name norval w king jr birthdate april u citizen yes  
education maryland preveterinary medicine university georgia major research  
comparative pathology ultrastructural pathology interest position director  
collaborative research program new england regional primate research center  
medical school professor comparative pathology england regional primate research  
center harvard medical school general pathology medical curriculun harvard  
medical school division pathobiology new england primate research hm associate  
comparative pathology electron microscopic histology menber faculty medicine ims  
veterinary pathologist div pathobiology new england regional primate re ctr  
pathology electron microscopic ims pathologist div pathobiology rog primate re  
ctr diplomate american college veterinary pathologist fellow pathology ims  
assitant veterinary pathologist animal re ctr captain u veterinary corp  
assistant vet pathology div u army med re nutrition lab fitzsimons gon denver  
colo animal practice salisbury salisbu maryland selective publication karpas  
melendez lv king nw isolation spontaneous disease squirrel monkey saimiri  
sciureus archiv furdie gesamte virusforschung rd melendez lv king nw garcia  
herpesvirus saimiri malignant spider monkey new susceptibe host j med primatol  
has\_stopwords: ['re']

Índice da linha: 1192

Texto original: ff biographical sketch kay personnel name position professor wih  
institution location fielo college new york city university york university pa  
philadelphia lawrence university california livermore ca professional experience  
concluding oder key ofthe project membership federal relerences three year  
exceed two research assistant hunter college new york ny biology college new  
york student laboratory university livernore ca biologist division lawrence  
california ca professor medicine new york university medical center medicine new  
york university nedical center director new york university medical center ny  
cell new york science journal present environmental section penn activity sperm  
penn gelatin sub proteolytic penn sperm ucrl phd barton thesis ponn tinorian h  
division delay sea urchin induced protease re shore r pena shore b gel  
electrophoresis focusing plasna apolipoproteins crc press lewis j albert change  
size proliferation lesion cockerel spontaneous lesion ate procedure page  
has\_stopwords: ['re']

Índice da linha: 1193

Texto original: biographical sketch provide page postion robert niederman  
professor year institution location fieldof university connecticut bacteriology  
university connecticut m bacteriology university urbana dvm veterinary medicine  
university urbana bacteriology professional experience concluding present  
present membership federal advisory ordoor year publication year exceeds page  
exceed two page professional experience assistant professor associate professor  
dept professor dept molecular biology biochemistry vice chair rutgers university  
new brunswick nj research teaching biochemistry research fellow biochemistry

dept biochemistry university bristol u k visiting professor dept molecular  
biology biotechnology univ sheffield u k postdoctoral fellow dept physiological  
chemistry roche institute molecular biology nutley research molecular biology  
aec postdoctoral fellow dept biochemistry michigan state university e lanning mi  
research enzymology nih div dept university research assistant bacteriology div  
dept dairy science university illinois urbar research microbial physiology  
research microbial physiology biochemistry research assistant dept animal  
disease university connecticut storrs ct research veterinar y virology honor  
award editorial board journal bacteriology public health service research career  
development award gm national institute general medical science merck grant  
faculty development merck company foundation aec postdoctoral fellowship nih  
postdoctoral fellowship editor veterinarian academic honor university  
connecticut publication olivera l niederman r effect phospholipase az carotenoid  
bacteriochloro complex rhodobacter westerhuis w h j farchaus j niederman r  
altered spectral property b light harvesting pigmentprotein complex rhodobacter  
sphaeroides lacking pufx photobiol olivera l w h j niederman r alteration  
rhodobacter capsulatus sitedirected change bacteriochlorophyllbinding site b  
lightharvesting complex biochim biophys acta wu j niederman r topological  
organization rieske ironsulphur protein subunit iv cytochrome complex  
rhodobacter sphaeroides biochem j hu youvan niederman carotenoid binding  
combinatorial lh mutant rhodobacter capsulatus p mathis ed light biosphere vol  
pp kluwer boston  
has\_stopwords: ['m']

Índice da linha: 1194

Texto original: j view president life division little ine canbridge born  
béucation cornell university biochemistry nat acad re council cancer research  
pharmacole harvey britain professor exp boston unive school med kiplinger g  
failure phencxybenzamine prevent chronic carbon chloride nat cancer inst york  
rogers production throrbi embolus hamster cheek pouch phthalanilides related exp  
jule rogers york physiological disposition study cancer interest tobacco smoke  
ne yo scie source american men science th index  
has\_stopwords: ['re']

Índice da linha: 1195

Texto original: patrizia longone ost classico baccal biological professional  
experience inthe biological chemistry university genoa genoa italy pediatric  
oncology doctoral fellow georgetown washington dc nathan kline insitute cemer  
neuropharmacology new york university orangeburg ny present doct fellow  
department psychiatry psychiatric chicago uic school society society  
neuroscience publication g morelli pit longone de flora characterization guanine  
erythrocyte membrane biochem byophys re longone p comagliaferraris p ponzoni  
retinoic acid rasensitive raresistant human neuroblastoma cell biochem re comm  
wojcik w holopainen w longone p riva mocchetti regulation receplors primary  
culture cerebellar granule neuron rat biol fukuda ed pp g racagni  
has\_stopwords: ['re']

Índice da linha: 1196

Texto original: brief curriculum vitae scientific bibliography birthdate katusic  
 md assistant professor pharmacology year field conferred study faculty  
 university belgrade yugoslavia faculty university belgrade yugoslavia  
 pharmacology medical faculty university belgrade yugoslavia pharmacology  
 research professional experience academic position professor department  
 pharmacology medical faculty university belgrade research fellow department  
 physiology biophysics mayo clinic mayo foundation rochester mn professor  
 department pharmacology medical faculty university belgrade scientist department  
 physiology biophysics mayo clinic mayo foundation rochester nn professor  
 department pharmacology medical faculty university belgrade research fellow  
 department physiology mayo clinic mayo foundation rochester mn present assistant  
 professor pharmacology mayo medical school rochester mn associate department  
 physiology mayo clinic mayo foundation rochester mn present research associate  
 department anesthesiology mayo clinic mayo foundation rochester hn membership  
 federation clinical research heart association council basic science society  
 pharmacology experimental therapeutic american physiological society  
 international union angiology society vascular medicine research yugoslav  
 pharmacological society yugoslav physiological society experience regular  
 reviewer american journal physiology heart circulatory physiology circulation  
 research european journal pharmacology hypertension life science quarterly  
 journal experimental physiology selected publication total number paper abstract  
 katusic z divergent effect cyproheptadine r ht antagonist response rat eur j  
 krstic katusic divergent effect pirenperone antagonist pressor response  
 guineapigs arch int physiol biochim z shepherd jt pm vasopressin cause endothel  
 dependent canine basilar artery circ re vanhoutte katusic z shepherd jt  
 vasopressin induces dependent relaxation cerebral coronary systemic artery j  
 hypertension  
 has\_stopwords: ['re']

Índice da linha: 1201

Texto original: mercola daniel mdphd associate university california los angeles  
 ba university california los angeles m university california los angeles phd  
 biophysics university southampton bm medicine oxford postdoctoral prof hodgkin  
 nobel laureate oxford member biological science southampton medical study  
 california state license md california san diego board certification american  
 board pathology pathology facp california san diego pathology school medicine n  
 morris jws mercola daand preparation property nitrotyrosine insulin blundell  
 tlcutfield jf cutfield sm dodson ej gghodgkin daand vijayan atomic position  
 crystal nature blundell tlcutfield dodson dodson gg dc mercola dathe crystal  
 insulin cold quant biol mereola damorris jws use interaction study chain folding  
 insulin solutionbiochemisiry blundell dodson g g mercola crystal reflection  
 chemistry protein chemistry blundell cutfield j cutfield dodson e dodson  
 hodgkind mercola dimensional atomic structure insulin relationship mercola  
 bullard b priest j crystallization nature bullard b mercola mommaerts whf  
 evidence circular dichroism structure tropomyosin biochem biophys acta bentley g  
 dodson e hodgkin mercola insulin insulin nature strickland e mercola dichroism  
 pig insulin monomer dimer hexamers coupling approximation bigchemistry levine  
 amercola j mand coffman magnetic resonance study j mol biol wollmer j

strassburger w thiele h dodson mercola mobility dichroism protein biophysical  
toponinc proton  
has\_stopwords: ['m']

Índice da linha: 1202

Texto original: fe biographical sketch nd ad haldar subrata research assistant  
professor education win year institution location study university calcutta  
india biochemistry university calcutta india ph september august associate  
environment government india september january research fellow university  
pennsylvania philadelphia laboratory dr john g monroe february september  
postdoctoral research fellow wistar institute philadelphia laboratory professor  
carlo croce october december research fellow fels institute temple university  
laboratory professor carlo croce january june fels institute temple university  
philadelphia july present research assistant professor jefferson cancer  
institute thomas jefferson universit march pres reviewer national cancer  
institute bethesda maryland publication ghosh chatterjee k haldar tiwari r  
chatterjee gc study membrane atpase activity lipid lev el tor normal ine  
phosphatase catalase resistant condition acta ha haldar chatterjee gc study alk  
brio el tor normal rifampi microbiologica polonica microbiol basu haldar  
chakraborty k santra chatterjee gc effect cysteine m antioxidant defense basu  
santra haldar le several scaven jon acta gc effect enzyme rat kidney influence  
vitamin e inol enzymol  
has\_stopwords: ['m']

Índice da linha: 1203

Texto original: la fahri biographical sketch information begin wih photocopy  
fahri assistant research molecular biologist begin institution location fieldof  
univ med school turkey none univ fairfield tova as maharishi intl univ towa  
molec biol maharishi intl fairfield professional experience wih key personnel  
typically include individual doctoral professional degree project al master  
exceed page graduate student department molecular cellular biology maharishi  
international university fairfield john fagan thesis advisor visiting scientist  
mbcrr dana farber cancer institute harvard medical school boston ha phd  
supervisor fellow ucsd departnent la jolla ca michael karin phd supervisor  
assistant research ucsd depart pharmacology la jolla ca honor b summa cum laude  
m distinction outstanding phd student outstanding master student avar  
outstanding biology major award graduating class national dean list visiting  
scientist fellowship harvard medical school sept june marchjuly visiting  
scientist nih julynov postdoctoral fellowship disease research program  
postdoctoral fellowship anerican heart association publication saatcioglu f  
perry dj pasco fagan jb multiple dnabinding factor interact overlapping  
specificity aryl hydrocarbon response element cytochrome gene mol cell biol  
saatcioglu perry dj pasco d fagan jb aryl hydrocarbon ah dnabinding activity j  
biol chem saatcioglu f bartunek p deng zenke karin conserved cterminal sequence  
deleted essential transactivation interference ap induction erythroid  
differentiation ligand binding hormone mol cell biol saatcioglu deng karin novel  
ciselement mediating ligand independent activation cerba implication hormonal  
regulation cell yangyen chambard jc deng saatcioglu positive negative gene

regulation nuclear receptor bur j clin pharmacol saatcioglu f helmberg karin  
modulation immune function interference nuclear receptor ap fund clin form page  
has\_stopwords: ['m', 'd', 'as']

Índice da linha: 1210

Texto original: biographical sketch henry tien lau positiontitle assistant  
professor pediatric surgery date birth november education city college new york  
new york b howard university washington dc postgraduate postdoctoral training  
intern surgery harlem hospital new york city resident surgery memorial sloan  
kettering new york city fellow transplantation immunology columbia university  
presbyterian hospital new york resident surgery harlem hospital new york city  
assistant chief resident pediatric surgery resident pediatric surgery hospital  
philadelphia visiting fellow liver transplantation university chicago academic  
appointment instructor pediatric surgery dept surgery university pennsylvania  
school medicine assistant professor pediatric surgery dept surgery university  
pennsylvania school medicine publication lau ht raw schmidt bj et al pain  
insensitivity metabolic disease lancet lau ht hardy altman rp deceased pulmonary  
macrophage bactericidal activity splenectomized rat j surg re lau ht reemtsma k  
hardy pancreatic islet allograft donorspecific blood transfusion treated  
ultraviolet irradiation science lau ht reemtsma k hardy prolongation rat islet  
allograft survival direct ultraviolet irradiation graft science hardy lau ht  
weber c reemtsma k pancreatic islet transplantation  
has\_stopwords: ['re']

Índice da linha: 1213

Texto original: curriculum vitae nicholas j gross ph birthdate august malaya  
citizenship british married two son degree ba cambridge england cambridge  
england cambridge london england london chicago md flex illinois education  
cambridge university england guy hospital london england guy hospital london  
house physician whittington hospital london house physician central middlesex  
hospital london house physician west end hospital london sr house officer  
hammersmith hospital london house officer national heart london re med hohns  
hopkins hospital baltimore fellow med university chicago usphs fellow cardiology  
university chicago instructor med university chicago assistant professor  
department medicine university chicago head pulmonary section present visiting  
colleague institute disease chest brompton hospital london postgraduate medical  
school london  
has\_stopwords: ['re']

Índice da linha: 1214

Texto original: tucker patterson office address nw th ave box gainesville  
gainesville april education university arkansas fayetteville b university south  
carolina columbia phd pharmacology university florida gainesville  
pharmacodynamics fellow resent research andor professional experience university  
south carolina columbia teaching assistant university south carolina columbia  
poison advisor palmetto poison south carolina columbia instructor pharmacology  
university florida gainesville postdoctoral fellow neurobiology aging present  
university florida gainesville neurochemical foundation pharmacodynamics dept

pharmacodynamics spring publication patterson ta av terry jr jw kosh prevention  
physostigmine dfp diazinoninduced acute toxicity monoethylcholine naminodeanol j  
pharmacol patterson ta jw kosh simultaneous quantitation acetylcholine choline  
tissue using gas chromatography electron mass spectrometry biol mass spectrom  
patterson ta jw kosh elucidation rapid invivo metabolism arecoline gen pharmacol  
patterson jw kosh modification duration action pharmacodynamics arecoline  
isoompa pharmacol re meldrum tu patterson r dawson effect blood pressure urinary  
sodium potassium calcium excretion disease huxtable plenum press new york ny  
press patterson ta ek kim mj meldrum r dawson jr glutamate efflux rat brain  
slice comparison depolarizing agent potassium aminopyridine veratrine submitted  
publication  
has\_stopwords: ['re']

Índice da linha: 1217

Texto original: vitae principal pooks road bethesda married three child born  
april professional student medicine university medical school jerusalem  
internship hospital jerusalem field research coronary disease graduated doctor  
medicine resident n hadassa university resident chest discases hadassa  
university hospital jerusalem resident medicine university hospital staff  
medicine hospital nih trainee medicine university washington dc house second  
assistant physician first assistant physician chief physician lecturer  
physiology university medical school assistant professor medicine george  
washington university medical school research associate veteran administration  
hospital washington director pulnorary physiology laboratory veteran hospital  
washington dc assistant professor physiology georgetown university clinical  
investigator veteran adainistration at aik afcr practice document subject  
agreement sf  
has\_stopwords: ['at']

Índice da linha: 1226

Texto original: sm antelman david j edward professor duke university durham nc  
university north carolina chapel hil university pittsburgh pittsburgh pa post  
doc biochemistry professor department psychiatry western clinic university  
pittsburgh school medicine university pittsburgh pa assistant professor  
department pharmacologyphysiology school dental medicine university pittsburgh  
pittsburgh professor department pharmacologyphysiology schoo dental medicine  
university pittsburgh pittsburgh pa department dental medicine university  
pittsburgh pittsburgh pa selected publication edward blau k phenvlethylamines  
gasliquid chromatography electroncapture anal biochem edward blau vivo formation  
young rat injected j neurochem edward blau k aromatic acid derived phenylalanine  
rat experimentally induced biochem j dj blau brain rat experimentally induced j  
dj burn mo tricyclic antidepressant upon human platelet oxidase lite edward dj  
chang interacting catalytic site human platelet biophys re commun edward dj  
chang form monoamine oxidase rabbit platelet life sci student ak edward dj type  
b monoamine oxidase brain biochem pharmacol edward dj malsbury cw type b  
monoamine oxidase discrete brain region pineal gland golden hamster life sci  
edward dj phenylethanolamine specific substrate type b oxidase life sci edward  
pak ky selective radiochemical labeling type b active site rat liver monoamine

oxidase biophys re commun edward dj doshi hanin analysis phenylethylamines gas chromatographychemical ionization mass spectrometry anal biochem edward quantitation glycol human rat normal condition clin chim acta edward dj use acid novel derivatizing analysis catecholamine histamine related amine gas chromatography chromatogr edward molecular property monoamine oxidase schizophrenia bulletin  
has\_stopwords: ['re']

Índice da linha: 1242

Texto original: biographical sketch name position title zaven assistant professor pathology neuroscience institute location degree year field study boston university boston boston university boston physic university baltimore california developmental pasadena ca research experience student dr douglas fambrough john hopkins university fellow dr paul patterson california institute technology present assistant professor depts pathologyneuroscience albert einstein college medicine award aeronautics space administration nasa predoctoral training grant health predoctoral training muscular mda research research service award individual postdoctoral fellowship manuscript kaprielian fambrough dm expression fast slow isoforms developing chick skeletal muscle dev biol kaprielian z campbell fambrough identification cerebellar purkinje cell mol brain re karin kaprielian z fambrough expression avian cultured mouse myogenic cell mol cell biol suzue ian patterson monoclonal antibody defines gradient mammalian nervous system neuron kaprielian z bandman e fambrough dm expression atpase isoforms regenerating denervated dystrophic chicken skeletal muscle dev biol kapriclian z patterson surface cytoskeletal marker inthe mammalian nervous system j  
has\_stopwords: ['re']

Índice da linha: 1243

Texto original: william e munger ph curriculum vitae department pathology university connecticut health center farmington ct telephone number general information march whittier ca married social security education university california santa barbara university connecticut appointment fellow department pathology university connecticut center farmington ct present publication munger w e lindquist effect ester alloimmune cytolysis cancer re w lindquist r acetate itpa augments killing ability lyt lytexpressing subpopulation killer preparation h munger w cohen c cohen effect lymphokine intraperitoneal neoplasm preparation mecley munger w picciano p cohen c interaction neoplastic cell vitro preparation w e lindquist augmentation si phorbol ester tpa fed proc w e lindquist acetate tpa augments ailoinmune cellmediated cytotoxicity cmc amer assoc cancer re munger w e lindquist r r acetate activates low lyt antigen expressing poorly cytolytic killer subpopulation fed proc abst  
has\_stopwords: ['re']

Índice da linha: 1252

Texto original: dale allen park assistant professor western michigan university b biologychemistry university south alabama college medicine basie med university south alabama college physiology position postdoctoral fellow

department physiology university south alabama college medicine mobile alabama  
postdoctoral assistant department university cincinnati college medicine  
cincinnati ohio postdoctoral fellow department physiology university college  
medicine cincinnati ohio present professor department physiology university  
alabama birmingham center birmingham alabama selected publication park bulkley g  
granger dn hamilton jm injury small intestine role superoxide radical  
gastroenterology grogaard b park da granger dn mecord jm fosberg jo radical  
mucosal albumin j physiol gg park da grogaard b granger dn comparison complete  
arterial occlusion model studying intestinal ischemia surgery granger dn park da  
role oxygen radical physiologist pathogenesis intestinal ischemia park da  
granger dn oxygen derived intestinal in oxygen radical scavenger system volume  
medical aspect edited greenwald ra cohen g elsevier biomedical new pp da granger  
dn va hydroxyl radical j physiol g role xanthine oxidase park da bulkley gb  
granger dn role free radical digestive tract pathology park da bulkley gb  
granger dn role free ischemia preservation surgery  
has\_stopwords: ['in']

Índice da linha: 1257

Texto original: biographical sketch michaël r hoffmann phd title professor  
environmental chemistry california institute technology institution california  
institute technology ba university phd brown university phd california institute  
technology professional experience assistant professor associate professor  
department civil mineral engineering university minnesota associate professor  
engineering science california institute technology visiting professor dibendorf  
swiss federal technical institute professor environmental engineering science  
california institute technology present professor environmental chemistry  
division engineering applied science california institute technology present  
pertinent publication j collett b c daube j w munger r hoffmann comparison two  
cloudwater rotating arm collector active strand collector atmos environ j  
collett b c daube j w munger hoffmann chemical intercepted cloudwater sierra  
nevada j collett b c j w munger r hoffmann intensive study sierra cloudwater  
chemistry relationship precursor aerosol gas concentration atmos igawa fukushi r  
hoffmann selective aldehydes across anionexchange membrane via formation  
bisulfite ind eng chem re w gunz hoffmann atmospheric chemistry peroxide atmos  
collett b c daube j w munger r hoffmann spatial temporal variation precipitation  
cloud interception sierra nevada tellus press c kormann w bahnmann hoffmann  
photolysis aqueous solution organic molecule presence suspension sci technol  
revision j w munger j collett b c daube r hoffmann cloudwater chemistry santa  
barbara channel environ sci revision  
has\_stopwords: ['re']

Índice da linha: 1259

Texto original: mansoor ahmed senior research associate new college madras india  
bsc zoology university madras india msc environmental toxicology university  
madras india environmental toxicology junior research fellow dept pharmacology  
environmental toxicology pgibms university madras india senior research fellow  
carcinogenicity mutagenicity laboratory dept toxicology frederick institute  
plant protection toxicology tamilnadu india foreign research fellow dept



hematology research institute nuclear medicine biology hirosshima university  
japan postdoctoral fellow jefferson cancer institute philadelphia pa july fellow  
department radiation oncology thomas jefferson university hospital philadelphia  
pa research associate dept radiation medicine university kentucky lexington ky m  
muthu p ahmed mm cardiotoxicity study interaction lead digoxin calcium indian j  
med re ahmed regu k lack genotoxicity menthol chromosome aberration sister  
chromatid exchange assay using human iymphocytes vitro toxicology vitro tanaka k  
takechi asou h takauchi k tashiro kamada n oncogene activation hematological  
malignancy proc re inst nuclear med biol hirosshima univ tanaka k ahmed shigeta c  
oguma n kamada loss heterozygosity locus short arm chromosome chronic myelocytic  
leukemia cancer genet cytogenet mm tanaka k kamada n molecular aspect indian  
patient syndrome chronic myelocytic leukemia non hodgkin re inst nuclear med  
biol hirosshima univ  
has\_stopwords: ['m', 're']

Índice da linha: 1260

Texto original: name title institution university university notre dame  
university biographical sketch robert byrne research associate professor october  
location conferred study dayton dayton notre phd biochemistry chicago il  
biochemistry research experience membership american cher american as  
publication acid university notre dame notre dame predoctoral research associate  
directorship dr castellino involved basic research regarding relationship  
vitamin kdependent blood coagulation protein plasminogen university chicago  
chicago il research associate engaged series study aimed understanding structure  
function metabolic property major apoproteins plasma high density namely apo ai  
apo university chicago chicago il research associate instructor continued basic  
research focusing use proteolytic enzyme structural probe hdl university chicago  
chicago il research associate assistant professor continuing study interaction  
focusing property elastase human pmn cleaves apo aii hdl apo b usphs  
postdoctoral fellowship hl university chicago chicago il professional related  
organization council american heart mical society biological chemist byrne fj  
effect plasminogen structure activation j biol  
has\_stopwords: ['as']

Índice da linha: 1273

Texto original: principal investigatorprogram director pfeffer lawrence  
biographical sketch following information key consuitants photocopy page person  
lawrence pfeffer professor year study university new york albany albany ny  
university graduate school medical science new york ny phd biochemistry  
rockefeller university new york fellow virology mol biol page professional  
experience undergraduate reseach fellow suny albany albany ny research fellow  
university new york ny fellow rockefeller university new york ny associate  
rockefeller university new york ny professor rockefeller university new york ny  
professor rockefeller university new york ny professor department pathology  
university tennessee college memphis tn present professor director graduate  
study department pathology university tennessee college medicine memphis tn  
present editorial board journal interferon research honor award magna cum laude  
signum laudis honor society suny albany postdoctoral trainee national cancer

institute pathobiology cancer workshop postdoctoral trainee junior faculty  
research award american cancer research leukemia society america scholar award  
council international society interferon research present northeast present  
america present who science engineering present bibliography selected list  
publication pfeffer lm pincus fleissner polymorphism endogenous murine leukemia  
virus revealed focusing polyacrylamide gel virology lm lipkin kopelovich l  
differential genetic cultured human skin fibroblast transformation kirsten  
murine sarcoma virus cell murphy j tamm interferon effect growth division human  
fibroblast exp cell re pfeffer lm wang e tamm interferon effect microfilament  
organization cellular fibronectin cell motility human fibroblast j biol wang e  
tamm interferon inhibits redistribution cell surface component j exp med e  
pfeffer lm tamm interferon increase abundance submembranous microfilaments  
suspension culture proc natl acad sci usa pfeffer lm kwok bcp landsberger fr  
tamm stimulates phosphatidylcholine synthesis inhibits cholesterol ester  
synthesis cell proc natl acad sci usa  
has\_stopwords: ['who', 're']

Índice da linha: 1278

Texto original: ff samuel biographical sketch key personnel begin adunyah  
assistant professor institution location sci kumasi ghana biochemistry state  
univ stillwater m univ louisville louisville ky phd biochemistry univ mexico  
albuquerque cancer biology research ano professional experience win key persone  
present advisory author io oo two page professional experience sr re assoc  
hematologyoncology sect dept medicine univ alabama school medicine birmingham al  
re instructorassoc scientist dept medicine uab birmingham asst professor dept  
biochemistry meharry medical college present honor award summer studentship  
sandoz ag basel switzerland swiss agronomy summer traineeship zurich swiss  
research student scholarship eth zurich graduate assistantship oklahoma state  
university national collegiate natural science usa achievement academy award nci  
biomedical travel award american society hematology travel award society  
membership american society hematology american society advancement science  
american association cancer research american society biochemistry molecular  
biology representative publication niggli v adunyah se penniston jt carafoli  
purified human erythrocyte effect calmodulin phospholipid j biol chem niggli v  
adunyah carafoli acidic phospholipid unsaturated fatty acid limited mimic effect  
calmodulin purified human erythrocyte j bio vincenzi ff adunyah niggli v  
carafoli e purified red blood cell evidence direct inhibition presumed anti  
calmodulin drug absence calmodulin cell calcium  
has\_stopwords: ['m', 're']

Índice da linha: 1281

Texto original: abbreviated bibliographical sketch koch yitzhak department  
neurobiology weizmann institute science rehovot study institute hebrew  
university jerusalem msc hebrew university phd appointment institute weizmann  
institute dept neurobiology niehs research triangle park israel endocrine  
society president institute dept hormone research head bethesda md usa visiting  
scientist weizmann institute dept hormone research associate prof weizmann  
institute dept hormone research senior scientist weizmann institute dept hormone

research scientist michigan state university research assoc partial list  
 publication koch wilchek fridkin p zor hr production characterization synthetic  
 gonadotropinreleasing hormone biochem re commun koch chobsieng p zor u fridkin  
 lindner hr suppression gonadotropin secretion prevention ovulation rat antiserum  
 synthetic hormone re och chobsieng p fridkin degradation luteinizing hormone  
 hypothalamic tissue biochem biophys re commun koch gonadotropinreleasing hormone  
 human brain radioimmunoassay nature koch effect antibody luteinizing hormone  
 reproduction james excerpta medica amsterdam fridkin kalir rotman koch synthesis  
 luteinizing hormonereleasing hormone j solidphase biochem koch hazum fridkin  
 resistance enzymic degradation analogue possessing increased biological activity  
 biophys re okon koch gonadotropin releasing hormone organ human brain nature  
 koch hazum fridkin gonadotropin releasing hormone science z zor u r koch  
 difference pituitary amp response gonadotropinreleasing hormone j physiol e okon  
 localization releasing hormone human intern rev exper r koch luteinizing  
 hormonereleasing hormone rat pituitary cell life fridkin koch study enzymic  
 luteinizing hormone hormone rat pituitary plasma membrane biochem biophys re  
 commun kalcheim c koch bisexual behaviour male rat treated neonatally antibody  
 luteinizing hormonereleasing hormone lhrh j compphysiol koch characterization  
 subcellular localization analog binding brain  
 has\_stopwords: ['re']

Índice da linha: 1282

Texto original: investigatoriprogram director biographical sketch give following  
 information key personnel consultant listed page begin principal director  
 photocopy page ouglas c hixson assoc research oncologist professional education  
 postdoctoral begin location university purdue university w lafayette m  
 biobiochemis univ texas graduate school biomedical science houston phd biomed  
 science professional wih present present federal government com plete year  
 exceed two page b hite predoctoral fellow virology univ texas md anderson  
 hospital houston texas investigator dept virology univ texas md anderson  
 hospital tumor int houston texas research associate depts biochemistry virology  
 univ texas md anderson ho tumor institute houston texas biochemist assistant  
 professor biochemistry univ texas system cancer science parkresearch division  
 smithville texas associate research oncologist rhode island hospital providence  
 rhode island associate professor medicine pathology brown university providence  
 publication gm giuliano gilliam hixson walborg lectininduced hepatoma cell  
 lectin receptor activity cellsurface glycopeptides effect neuraminidase cancer  
 re giuliano hixson walborg jr role cell surface sialic acid lectinindw  
 agglutination hepatoma disease l bolis j hoffman le raven press new york ny pp  
 neri g mc giuliano capetillo eg gilliams dc hixson walborg jr regulatory surface  
 lectininduced cytoagglutination rat hepatoma cell acta medica romana starling dc  
 hixson davis walborg jr surface property rat hepatocyte increased concanavalin  
 ainduced agglutinability following papain digestion exp cell re chan jc dc  
 hixson bowen detection vesicular stomatitis virus murine leukemia pseudotypes  
 immunoelectron microscopy virology hixson dc jm bowen scanlon nash l dmochowski  
 strain related variation ferritin conjugated plant lectin murine mammary tumor  
 re glenney dc hixson ef walborg jr inhibition concanavalin ainduced hepatoma  
 cell cytochalasins metabolic inhibitor role cellsurface morphology yep

periodate lysine paraformaldehyde fixation method crosslinking plasma membrane  
histochem cytochem at the page  
has\_stopwords: ['m', 're']

Índice da linha: 1287

Texto original: biographical sketch provide following key form page page person  
xiaofan assistant professor at university wuhan china b biochemistry university  
california los angeles phd biol chem whitehead institute postdoctoral re  
cambridge fellow professional experience previous experience present public  
author reference year pertinent year exceeds two exceed two page professional  
experience assistant professor duke univ medical center durham nc mechanism  
negative growth control fellow whitehead institute biomedical research cambridge  
advisor dr weinberg molecular cloning graduate student department biological  
chemistry school medicine university california los angeles thesis advisor dr  
mechanism b lymphocyte development regulation academic award regent university  
fellowship damon runyon walter winchell cancer research fund fellow leukemia  
society america inc scholar publication selected mercola wang j calame  
transcriptional enhancer activity immunoglobulin heavy chain locus science g  
daneils skehel j wiley wang higa h paulson j hostmediated selection influenza  
virus receptor variant sialic acid galactose specific clone biol chem wang calame k  
endogenous immunoglobulin heavy chain enhancer activate tandem promoter  
separated large distance cell wang calame k sv enhancer binding factor required  
activation cell horowitz gerber wang e li ra deletion dna sequence mesenchymal  
proc natl acad sci usa wang xf weinberg ra cheifetz massague j absence growth  
factor receptor growth inhibitory response retinoblastoma cell science ph rev  
form page page  
has\_stopwords: ['re']

Índice da linha: 1290

Texto original: biographical sketch give following information for the key  
personnel consultant collaborator begin principal director photocopy page person  
richard lemon md phd institution degree field study connecticut ct ba chemistry  
math penn state university university park pa phd phys chem george washington  
university school of professional experience the exceed two nih nigrms postdoc  
re fellow lab biochem gen nhlbi nih general med re service nciva medical  
oncology branch washington dc staff fellow lab viral carcinogenesis nat cancer  
institute nih Bethesda md george todaro lab chief resident pediatrics dept  
pediatrics john hopkins hosp baltimore md nrsa fellow assistant member dept  
tumor cell biology st jude childrens research hospital memphis tn prof dept div  
hematol utah medical center salt lake city ut member program human molecular  
biology genetics univ utah present assoc prof dept division oncol dept human  
genetics utah medical center salt lake city ut honor phi lambda upsilon honorary  
society phi kappa phi honorary society jacobi medical society award pediatrics  
john hopkins honorary memorial award pediatrics american cancer society clinical  
oncology career development award pew scholar award biomedical science pew  
charitable trust bibliography total lemon r obrien sherr cj new genetic locus  
bevi human chromosome baboon type c virus human cells cell lemon nash wg obrien  
benveniste sherr cj gene bevi human integration site baboon type c cna provirus

human cell cell le beau pettenati mj lemon r diaz mo westbrook ca larson ra  
sherr rowley assignment gmcsf csf fm gene human chromosome provides evidence  
linkage family gene regulating hematopoiesis involvement deletion q myeloid  
disorder cold spring harbor symposium pettenati mj le beau mm lemon r shima ea  
kawasaki e larson sherr diaz rowley jd assignment q evidence clustering gene  
regulating hematopoiesis involvement delq myeloid disorder proc natl acad sci  
usa lebeau mm lemon r carrino jj pettenati souza lm diaz mo rowley jd  
chromosomal localization human gcsf gene proximal breakpoint ofthe acute  
promyelocytic leukemia leukemia ra lemon lebeau mm sr noack tack bf molecular  
human complement component c localization ofthe structural gene chromosome  
biochemistry  
has\_stopwords: ['re']

Índice da linha: 1293

Texto original: biographical sketch marcelo jacobsloreana professor education  
institution location sao paulo sao paulo b chemistry university osaka japan m  
biochemistry institute technology cambridge biology university geneva geneva  
switzerland professional experience graduate research master university osaka  
osaka japan advisor dr tsugita elucidation amino acid sequence wild type  
doubleframeshift mutant lysozyme phage work confirmed forthe genetic code via  
vivo experiment graduate research phd mlt cambridge massachusetts advisor dr  
baglioni major area research mechanism protein synthesis using cell free system  
b characterization globin histone mrna c demonstration characterization matenal  
messenger rna unfertilized sea urchin egg postdoctoral research university  
geneva geneva switzerland advisor dr crippa experiment gene expression  
drosophila oogenesis assistant professor department developmental genetics  
anatomy case western reserve university cleveland ohio associate professor dept  
genetics case western reserve university cleveland ohio professor dept genetics  
case western reserve university cleveland ohio service ad hoc reviewer following  
journal biochimica biophysica acta chromosoma development developmental biology  
insect molecular biology journal biological chemistry molecular cellular biology  
molecular reproduction development nucleic acid research proceeding national  
academy science science wilhelm archive developmental biology member advisory  
board crc year book developmental biology ad hoc reviewer national science  
foundation research council canada grant application ad hoc reviewer nih study  
section review fogarty international fellowship application mar ad hoc reviewer  
feb member nih genetics study section ad hoc reviewer nih general medical  
science minority review committee feb oct feb ad hoc nih tropical medicine  
parasitology study section oct ad hoc reviewer nth special review committee rfa  
nov publication following selected list housman jacobsloreana rajbhandary ul  
lodish hf initiation haemoglobin synthesis nature jacobsloreana baglioni c  
messenger rna globin postribosomal supernatant rabbit reticulocyte proc natl  
acad sci usa gross kw jacobsloreana baglioni c gross pr cellfree translation  
matenal messenger rna sea urchin egg proc natl acad sci usa jacobsloreana crippa  
mass fractionation drosophila egg chamber dev biol mermod jj jacobsloreana crippa  
change rate rna synthesis ribosomal gene number oogenesis drosophila dev biol z  
jacobsloreana drosophila egg chamber develop mature egg cultured vivo science  
biographical sketch

has\_stopwords: ['m']

Índice da linha: 1294

Texto original: lance terada research fellow pulmonary medicine education ba  
amherst college amherst university hawaii honolulu training experience  
internship residency internal medicine hospital cleveland oh clinical fellow  
pulmonary medicine university colorado present research fellow pulmonary  
medicine university colorado honor award letter commendation department  
pharmacology university hawaii alpha omega alpha second place research  
university hospital cleveland publication phelps waggoner l terada p williamson  
binding merocyanine normal leukemic erythroid cell cell l neonatal pneumonia  
community hospital retrospectivte study hawaii med j jw l terada jg douglas  
respiratory epithelium prostaglandin production granulocyte reduced oxygen  
molecule clin re patt harken ah burton lk rodell piermattei schorr wj parker nb  
berger em horesh ir linas cheronis jc terada l repine xanthine oxidase derived  
hydrogen peroxide contributes cerebral ischemiareperfusion induced edmea gerbil  
j clin invest accepted publication jm terada l grosso whitman gj velasco se patt  
harken ah repine je hydrogen perxoide xanthine oxidase contributes reperfusion  
injury ischemic isolated perfused rat heart publication  
has\_stopwords: ['re']

Índice da linha: 1296

Texto original: principal biographical sketch give information key personnel  
listed page beginning principal investigatorprogram director photocopy person  
professor pediatrics degree institution location fielo study ashland college  
ashland math fla gainesville fla phd md univ gainesville fla fellow biochem univ  
house research anoor professional experience honor en public committe past three  
yea publication exceeo xi graduate student research award florida chapter  
medicine society pediatric research midwest society pediatric research thoracic  
society medical faculty award researchuniversity florida participant various  
state perinatal symposium perinatal association perinatal care several national  
heart lung institute site review teans clinical fellowship review  
assistantassociate professor pediatrics professor pediatrics university  
wisconsin department pediatrics award univ wisspent lab dr f chytil vanderbilt  
invited speakerross conference dysplasia march representative publication zach  
effect dexamethasone maternal retinol binding protein j obstet gynecol retinyl  
ester synthesis isolated perfused ventilated neonatal rabbit lung internat j vit  
kakkad b chytil f perinatal rat lung retinol vitamin retinyl palmitate pediat re  
hustead va gutcher rd relationship vitamin retinol status lung disease perterm  
infant j pediat zachnan rd cotter pw end tsao conversion choline  
phosphatidylcholine ventilated perfused neonatal rabbit biol neonate tsao rd  
determination fluid single onedimensional thin layer chromatography method clin  
chim acta rd tsao uptake lung pediat re gutcher gr rd tsao analysis  
phospholipids organelle developing rabbit lung lipid tseo gutcher gr zachman rd  
effect hydrocortisone phosphatidyicholine maternal fetal rabbit lung liver peds  
re olson jr orbeck h graven sn rd isolated perfused rabbit lung preparation  
model oxygen toxicity j appl physoll hayden olson e rd biosynthesis fetal lung  
rabbit j effect administration maternal

has\_stopwords: ['re']

Índice da linha: 1297

Texto original: georg f springer md curriculum vitae may born february berlin germany eldest son elisabeth née kalvin dr hc ferdinand springer publisher springerverlag publishing firm founded expropriated dictatorial government family occupation force german army russia tank crew wounded seriously saved person drowning u citizen gymnasium humanistic univ heidelberg germany elected president student committee representing univ heidelberg student u military univ founding opening springerverlag heidelberg m medicine summa laude univ basel switzerland cwm laude extracurricular training organic chemistry reichstein nobel laureate internship univ penn philadeiphia pediatric research fellow prof p biochemistry r kuhn nobel joint germfree research project walter reed army med center washington dc univ penn prof immunopathol univ penn assoc prof pathol research immunologist univ penn hosp phila gen hosp chief blood bank examiner city philadelphia prof northwestern u med school chicago director dept evanston hosp prof surgery northw univ cofounder member northwest univ cancer center chicago tl heather bligh cancer biol research lab prof microbiol chicago med school va med center north chicago il discovery chronological two novel human bloodgroup active sugar higher plant proof antiblood group tn antibody man predominantly elicited intestinal flora addition antigen ag inhalation ingestion biological physical full chemical characterization lectin isolation characterization septic shock receptor microbial toxin attach eliciting lethal shock demonstration importance monosaccharide specific precipit lectin contrast general lattice formation immunologicalserological specific precipitin reaction patent cibageigy demonstration heavy contamination commercial influenza virus vaccine blood forssman ag danger pregnancy discovery one gene control blood group ag virus receptorspecificity including tumor causing virus eg human mn epitope eps specific myxovirus receptor ii infectious mononucleosis paul bunnell ag specific receptor putative rpt epsteinbarr tumor virus ii chicken blood group rpt rous sarcoma characterization human group ag eps mild stepwise degradation biosynthesis discovery human tn ag specific pancarcinoma camarkers ii autoags iii adhesion iv motility molecule tn immunoaccessible diseased healthy tissue except leukemia relative absolute density tn ca cell permit stratification patient adjuvant therapy post primary surgery human ca elicit throughout strong readily detectable cellular humoral high diagnostic even yr prior positive biopsy prognostic value discovery may cary meps successful active wit ag asti breast ca patient stage iii iv yr probability combined asti result survival rate due chance yr survival rate obtained use due chance p le yr le person yr le statistically probability p x

has\_stopwords: ['m']

Índice da linha: 1298

Texto original: name grigoryan grigory date birth january place birth samarkand former ussr marital son home address work address linwood close camberwell institute psychiatry department london s psychology de crespigny park denmark uk hill london s af uk phone ext fax education degree sci medexperimental psychology institute higher nervous activity neurophysiology ussr academy

science phd experimental psychology institute higher nervous activity  
neurophysiology ussr academy science diploma medical degree first moscow medical  
institute ussr employment research worker university london institute psychiatry  
department psychology london uk leading scientist institute higher nervous  
activity neurophysiology science department conditioned reflex moscow ussr  
senior research scientist institute higher nervous activity neurophysiology ussr  
academy science department conditioned reflex moscow ussr research scientist  
institute higher nervous activity neurophysiology ussr academy science  
department conditioned reflex moscow ussr international research experience  
visiting scientist laboratory prof grastyan institute physiology university  
medical school petz hungary funded ussr academy visiting scientist laboratory  
jagray department psychology institute psychiatry university london uk month  
month month funded royal society uk  
has\_stopwords: ['s']

Índice da linha: 1301

Texto original: director last fst richard biographical sketch key form page  
melinda borrello postdoctoral fellow degree fieldof university rochester  
rochester ny b microimmunology university rochester rochester ny m  
microimmunology university rochester rochester ny phd ity rochester rochester ny  
immunology federal advisory publication past three last exceed two summer  
research fellow department oral biology suny buffalo new york advisor dr  
rosemary dziak summer undergraduate research fellow department oral biology suny  
buffalo buffalo new york advisor dr rosemary dziak independent researcher  
department microbiology immunology university rochester new york advisor dr  
david w scott fellow department immunology university rochester rochester new  
york advisor dr richard p phipps present fellow university rochester cancer  
center eastman dental center rochester new york advisor dr richard p phipps  
coadvisor dr timothy blieden honor award bachelor science degree cum laude  
distinction immunology research awarded predoctoral fellowship immunology usphs  
training grant present awarded postdoctoral fellowship periodontology usphs  
training grant publication w borrello liou x r yao g warner b cell tolerance  
life death advance molecular cellular immunology b borrello e r fedyk brown r p  
phipps strategy studying regulation b lymphocyte prostaglandin fedyk e r  
borrello brown r p phipps regulation b cell tolerance complex chemical  
immunology borrello r p phipps fibroblast support outgrowth splenocytes  
simultaneously b lymphocyte macrophage characteristic journal immunology page ff  
has\_stopwords: ['m']

Índice da linha: 1303

Texto original: curriculum vitae mark brodie personal data birthdate august  
birthplace chicago marital status married social security number education ba  
university evanston biological science degree granted spring phd university  
chicago health science center chicago research professional experience  
undergraduate independent study anatomy department university medical school  
chicago il graduate student ph graduate trainee department pharmacology  
university illinois chicago health science center chicago il july july  
department pharmacology university colorado health science center denver co



postdoctoral research performed dr thomas v july july research scientist neuroscience research abbott laboratory abbott park il july september senior research scientist neuroscience research abbott laboratory abbott park il september present research assistant professor department physiology biophysics college medicine university illinois chicago chicago il membership neuroscience american society pharmacology experi mental therapeutic publication peerreviewed article book chapter abstract article selected brodie m proudfit hk analgesia induced local injection carbachol nucleus raphe magnus brain research brodie m proudfit hk antinociception induced local injection carbachol nucleus raphe magnus rat alteration intrathecal injection monoaminergic antagonist brain research brodie m lee fredholm l central versus peripheral mediation response adenosine receptor agonist evidence central mode action brain research brodie m dunwiddie cholecystokinin potentiates dopamine inhibition mesence phalic dopamine neuron vitro brain research mueller brodie intracellular recording putative dopaminecontaining neuron ventral tegmental area tsai brain slice preparation j neurosci

has\_stopwords: ['m']

Índice da linha: 1305

Texto original: fe biographical sketch key foreach parson research associate gretchen temeles bagin pa philadelphia pa b anthro yale university new ct phd biology merck sharp dohme west point mol biol u pa philadelphia pa re cell biol pa philadelphia pa re pres dev biol professional experience key ofthe atthe present three exceed two page award honor university fellowship nih training grant cell biology nih training grant biology yale comprehensive cancer center summer fellowship leukemia society america postdoctoral declined fellowship publication temeles gl defeojones tatchell k ellinger m scolnick em expression characterization ra saccharomyces cerevesiae mol cell biol papageorge ag defeojones robinson p temeles gl scolnick saccharomyces cerevesiae protein related gene product ra gene found mammal mol cell biol temeles gl gibbs jb j sigal l scolnick e yeast mammalian ra protein conserved biochemical sigal l gibbs jb j temeles gl wolanski bd socher sh scolnick mutant rasencoded protein altered nucleotide binding exert dominant biological effect proc natl acad sci usa kennett meyer b e temeles gl ikegaki n polakova immortalization stimulated transfection oncogenic dna first intemational symposium biotechnology monoclonal antibody treatment human irl press mclean va n temeles kennett protein expression associated chemically induced differentiation cell progress clinical biological research evan ae dangis g knudsen seeger r ed pp wileyliss new york conover temeles zimmermann burke b schultz rm stagespecific expression family prowins major product gene activation mouse embryo dev biol temeles schultz characterization expressed stagespecific manner cell mouse embryo preparation

has\_stopwords: ['m', 're']

Índice da linha: 1310

Texto original: curriculum vitae sundaesan cold spring harbor laboratory telephone po box cold spring harbor new york education phd university m university institute technology kanpur india physies college poona india teaching experience teaching assistant dept physic university teaching fellow

dept biochemistry harvard university present faculty program suny research  
experience senior staff investigator cold spring harbor laboratory cold spring  
harbor ny visiting assistant research geneticist university california berkeley  
pioneer international fellow laboratory dr michael freeling postdoctoral  
research fellow dept biology massachusetts general hospital boston dept genetics  
harvard medical school boston laboratory dr frederick ausubel graduate harvard  
university laboratory dr frederick ausubel graduate student rotation harvard  
university laboratory dr walter service federal advisory panel member nsf  
eukaryotic panel  
has\_stopwords: ['m']

Índice da linha: 1321

Texto original: investigator rory biographical sketch kent pinkerton phd  
nicotine analysis name role education begin location year brigham young  
univ provo utah b duke university nc pathology m duke university durham nc  
pathology research professional experience chronological professional including  
year citation al year le publication list author order published another name  
name exceed two medical research professor department pathology duke university  
assistant professor veterinary anatomy cell biology university california davis  
present professor residence department veterinary anatomy university california  
davis publication selected total ke jt gallen rr mercer vc wong cg bk  
acrosolized fluorescent microspheres detected lung using confocal scanning laser  
microscopy micro re technique ke de dodge j cederdahl demmler vj wong j peake ci  
haselton pw mellick g singh cg plopper differentiated bronchiolar epithelium  
duct rat exposed ozone month amer j ke j lewis mulder ikegami ah jobe surfactant  
effect alveolar type morphology rabbit j appl physiol ke ly chang barry  
morphometric assessment lung injury pattern following inhalation mineral fiber  
fiber toxicology db ed academic press pp ja tr gelzleichter ke pinkerton rm  
walker hp witschi new model progressive pulmonary fibrosis rat rev dis wv lg  
stewart ke pinkerton c ji ger g singh sl wm thurlbeck cg plopper product  
expression clara cell differentiation rabbit rat j amer physiol ke rr mercer cg  
plopper jd crapo distribution injury ozone ventilatory unit rat j appl physiol  
fm wc hubbard v wong k pinkerton cg ozone-induced acute epithelial relationship  
granulocyte emigration lung j respir cell mol biol c j w cardoso r wu k  
pinkerton development airway epithelium expression marker differentiation chest  
ke r vincent cg plopper sl young development growth aging pathobiology aging rat  
mohr u dl dungworth cc capen ed press washington dc pp  
has\_stopwords: ['m', 're']

Índice da linha: 1323

Texto original: r redacted material thomas j steichen vitae redacted education  
university kentucky kentucky m statistic st cloud state university st cloud  
minnesota mathematics professional experience reynolds tobacco company  
winstonsalem north senior staff rd statistician present redacted publication  
presentation steichen t sears sb ets metaanalysis u epa experience presented  
joint meeting smoke technology group budapest hungary september ogden mw sb  
steichen tj respiratory health effect passive smoking lung cancer disorder epaf  
critique metaanalysis background adjustment presentation osha may smith cj

steichen tj atherogenic carbon monoxide atherosclerosis steichen tj critical  
review passive smoking risk heart disease k jama provided osha rjr response rft  
exposure indoor air pollutant march ingebrethsen bj heavner dl ange al conner jm  
steichen tj green cr comparative study environmental tobacco smoke particulate  
mass measurement chamber jpnca  
has\_stopwords: ['m', 't']

Índice da linha: 1324

Texto original: positiontitle birthdate robert h mccusker jr re assistant  
professor education institution year location pennsylvania state univ b animal  
sci pennsylvania state univ m physiology university georgia phd nutr research  
professional experience senior research aide pennsylvania university research  
fellowship university north carolina chapel hill nrsa nih training grant  
lineberger comprehensive cancer research center research fellowship university  
north hill fellowship lineberger research center research associate university  
north chapel hill center thrombosis hemostasis research assistant professor  
university north carolina chapel hill department medicine division endocrinology  
july assistant professor university department animal science college  
agriculture professional american society animal science endocrine society  
american diabetes association representative publication published cw rh white  
gene expression secrétion insulinlike growth factor binding protein myoblast  
differentiation endocrinology rh cohick w busby wh clemmons dr evaluation  
developnental nutritional change porcine igfbp igfbp serum level immunoassay  
endocrinology camachohubner c mccusker rh clemmons dr secretion biological  
action insulinlike growth factor binding protein two human tumorderived cell  
line vitro j cell physiol rh busby wh dehoff mh camachohubner clemmons dr  
insulinlike growth factor binding cell monolayers directly modulated addition  
igf  
has\_stopwords: ['m', 're']

Índice da linha: 1334

Texto original: director biographical sketch te fotowing key personnel principal  
ginette serrero senior scientist beain wih year institution location field study  
university nice nice france b biology biochemistry university marseille  
marseille france biochemistry university nice nice france doct e sei biology  
research professional experience employment honor key key wil ina way present  
membership federal goverment two assistant nice france scientist dept biology  
calif san diego ca leave trom centre national de recherche scientifique france  
centre de biochimie university nice france research biologist cancer center  
california san diego la jolla ca scientist w alton jones cell science center old  
barn road lake placid ny w alton jones cell science center old barn road lake  
placid ny assistant professor dept biology university vermont burlington vt  
present assistant professor biology clarkson university potsdam ny delegation  
generale de la recherche et technique france nato exchange program fellowship  
research award new york state health department publication since serrero g jc  
khoo vitro model study adipose differentiation serumfree anal gaillard negrel g  
serrero g ailhaud growth preadipocyte cell line cell strain rodent serumfree  
medium vitro cell devel biol serrero g study teratomaderived adipogenic cell

```

line isolation insulin independent variant serumfree medium hormonally defined
medium tool cell biology g fisher rj wieser ed springerverlag berlin heidelberg
new york tokyo pp serrero g differentiation preadipocyte cell line mammalian
cell culture use serumfree hormonesupplemented medium jp mather ed plenum press
ny pp serrero g associated loss differentiation response insulin adipogenic cell
line vitro cell devel biol yamada g serrero characterization insulinrelated
factor secreted teratoma cell line biochem biophys re commun serrero g endocrine
autocrine control growth differentiation teratomaderived cell line cellular
endocrinology hormonal control embryonic cellular differentiation g serrero j
hayashi ed alan r li inc ny pp serrero g mill assay adipose differentiation
using primary culture adipocyte precursor j tissue culture yamada purification
insulinrelated factor irf secreted teratoma derived mesodermal cell line j biol
chem serreto g mill page
has_stopwords: ['re']

```

```
[67]: df['has_stopwords'].notna()
```

```

[67]: 0      False
      1      False
      2      False
      3      False
      4      False
      ...
     1330    False
     1331    False
     1332    False
     1333    False
     1334     True
      Name: has_stopwords, Length: 1335, dtype: bool

```

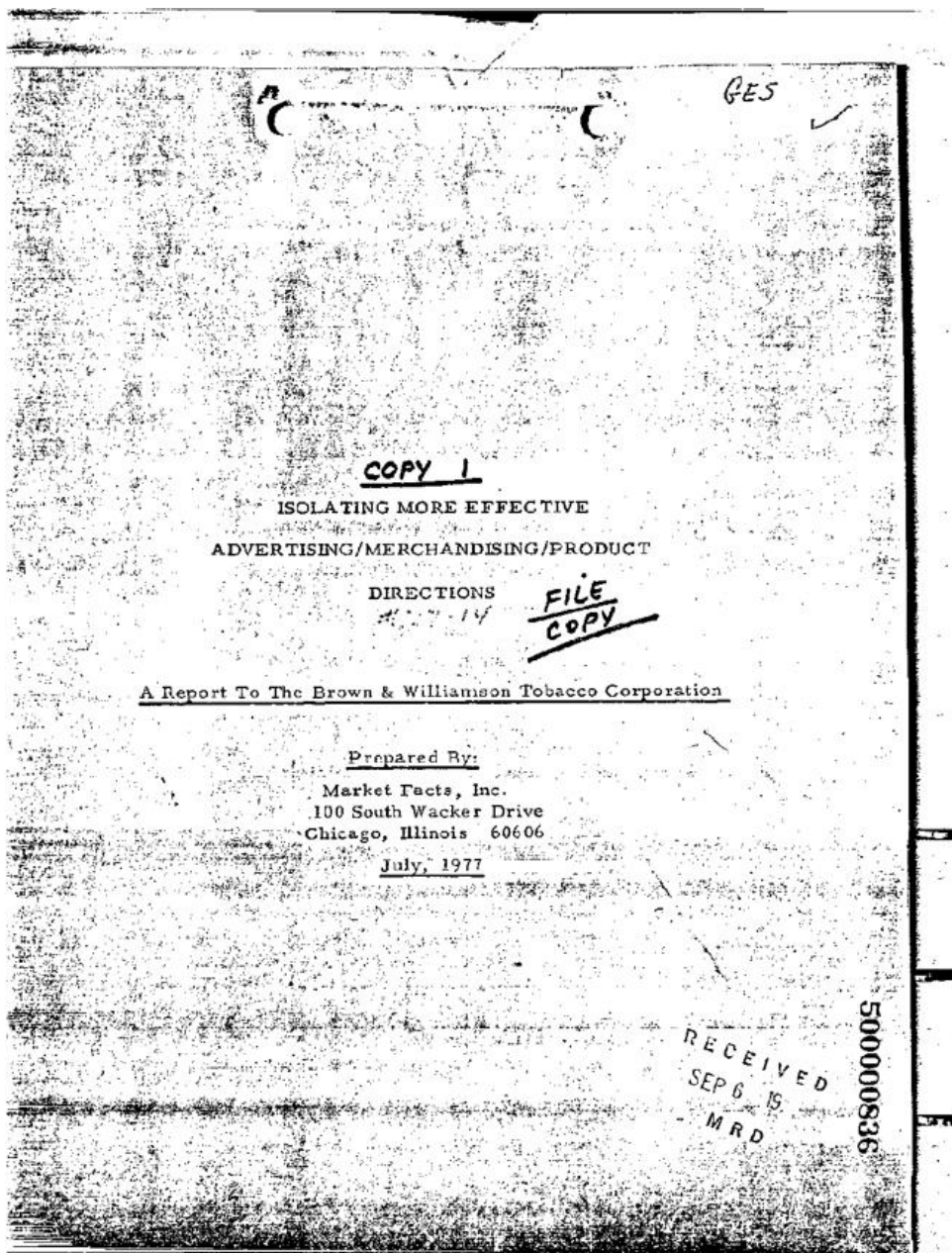
### 3.1 Analisando a imagem

```

[68]: from PIL import Image
      class_img = 'questionnaire'
      name_img = '0000002206.jpg'
      image = Image.open(f"{path_dataset}/{class_img}/{name_img}")
      image

```

```
[68]:
```



```
[69]: # Usar query para filtrar e iloc para pegar a primeira linha
filtered_row = df.query("class_img == @class_img and name == @name_img").iloc[0]
filtered_row
```

```
[69]: text          effective brown tobacco corporation market inc...
      class_img          questionnaire
      name              0000002206.jpg
      dict_ocr          {'block_num': [0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
      has_stopwords          None
      Name: 829, dtype: object
```

```
[70]: # obtem a linha do DF onde essa imagem foi processada
ocr_df = pd.DataFrame(filtered_row['dict_ocr'])
ocr_df.columns
```

```
[70]: Index(['block_num', 'conf', 'height', 'left', 'level', 'line_num', 'page_num',
          'par_num', 'text', 'top', 'width', 'word_num'],
          dtype='object')
```

```
[71]: #
image = draw_bounding_boxes(image, ocr_df)
image
```

```
[71]:
```



GES ✓

COPY 1

ISOLATING MORE EFFECTIVE  
ADVERTISING/MERCHANDISING/PRODUCT

DIRECTIONS

FILE  
COPY

A Report To The Brown & Williamson Tobacco Corporation

Prepared By:

Market Facts, Inc.  
100 South Wacker Drive  
Chicago, Illinois 60606

July, 1977

RECEIVED  
SEP 6 1977  
MRD

500000836

## 4 Separando os dados em treino e teste

```
[72]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(df['text'],
↳ df['class_img'], test_size=0.2, random_state=678)
```

### 4.0.1 Vetorizando dos dados

```
[73]: from sklearn.feature_extraction.text import TfidfVectorizer
# Configurando o TfidfVectorizer
tfidf = TfidfVectorizer(ngram_range=(2,5), max_df=0.95, min_df=2
↳ ,max_features=10000)
```

```
[74]: # Ajustando e transformando os dados de treinamento
tfidf_train = tfidf.fit_transform(X_train)
# Convertendo a matriz esparsa em uma matriz densa
tfidf_test = tfidf.transform(X_test)
```

```
[75]: from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.svm import SVC
from sklearn.naive_bayes import MultinomialNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import StratifiedKFold, cross_val_score
from sklearn.metrics import accuracy_score, precision_score, recall_score,
↳ f1_score
import numpy as np

# Lista de modelos para comparar
models = {
    'Logistic Regression': LogisticRegression(),
    'Random Forest': RandomForestClassifier(),
    'SVM': SVC(),
    'Naive Bayes': MultinomialNB(),
    'KNN': KNeighborsClassifier(),
    'Gradient Boosting': GradientBoostingClassifier()
}
```

## 5 Ajuste de hiperparâmetros usando GridSearchCV

busca pela configuração ideal de parâmetros para cada modelo.

Definição de param\_grids: Para cada modelo, definimos um dicionário de param\_grid que contém os parâmetros que desejamos ajustar utilizando GridSearchCV. Os parâmetros específicos foram escolhidos com base na prática comum e podem ser ajustados conforme necessário.

Loop de Ajuste de Parâmetros: Iteramos sobre o dicionário models, que contém os modelos a serem



testados. Para cada modelo, realizamos o ajuste de hiperparâmetros usando GridSearchCV.

GridSearchCV: Criamos um objeto GridSearchCV para cada modelo, onde especificamos o modelo, o param\_grid correspondente, o número de folds para validação cruzada (cv=5 neste exemplo) e a métrica de avaliação (scoring='accuracy').

Execução da Busca em Grade: Chamamos o método fit() para executar a busca em grade no conjunto de treinamento (tfidf\_train, y\_train).

Armazenamento do Melhor Modelo: Após a busca em grade, armazenamos o melhor modelo encontrado (best\_estimator\_) no dicionário best\_models.

Avaliação no Conjunto de Teste: Finalmente, avaliamos os melhores modelos encontrados no conjunto de teste (tfidf\_test, y\_test) e exibimos a acurácia para cada modelo.

[76]: *# baseado nos melhores parâmetros encontrados*

```
param_grid_lr = {
    'C': [0.1, 1.0, 10.0],
    'penalty': ['l2'],
    'solver': ['liblinear']
}

param_grid_rf = {
    'n_estimators': [100],
    'max_depth': [None],
    'min_samples_split': [2]
}

param_grid_svc = {
    'C': [1.0],
    'kernel': ['linear'],
    'gamma': ['scale']
}

param_grid_nb = {
    'alpha': [0.1],
    'fit_prior': [True]
}

param_grid_knn = {
    'n_neighbors': [10],
    'weights': ['uniform'],
    'metric': ['euclidean']
}

param_grid_gb = {
    'n_estimators': [200],
    'learning_rate': [0.5],
    'max_depth': [10]
```

```
}
```

## 6 Ajuste de Hiperparâmetros e Avaliação de Modelos

```
[77]: from sklearn.model_selection import GridSearchCV

# Definir os primeiros parâmetros para começar a testar em cada modelo
param_grid_lr = {
    'C': [0.1, 1.0, 10.0],
    'penalty': ['l1', 'l2'],
    'solver': ['liblinear', 'saga']
}

param_grid_rf = {
    'n_estimators': [100, 200, 300],
    'max_depth': [None, 10, 20],
    'min_samples_split': [2, 5, 10]
}

param_grid_svc = {
    'C': [0.1, 1.0, 10.0],
    'kernel': ['linear', 'rbf', 'poly'],
    'gamma': ['scale', 'auto']
}

param_grid_nb = {
    'alpha': [0.1, 0.5, 1.0],
    'fit_prior': [True, False]
}

param_grid_knn = {
    'n_neighbors': [3, 5, 10],
    'weights': ['uniform', 'distance'],
    'metric': ['euclidean', 'manhattan']
}

param_grid_gb = {
    'n_estimators': [50, 100, 200],
    'learning_rate': [0.05, 0.1, 0.5],
    'max_depth': [3, 5, 10]
}

# Dicionário de param_grids para cada modelo
param_grids = {
    'Logistic Regression': param_grid_lr,
    'Random Forest': param_grid_rf,
```

```

'SVM': param_grid_svc,
'Naive Bayes': param_grid_nb,
'KNN': param_grid_knn,
'Gradient Boosting': param_grid_gb
}

# Lista para armazenar os melhores modelos ajustados
best_models = {}
def hyperparameter_tuning_and_model_evaluation(best_models, models,
param_grids, tfidf_train, y_train):
    # Executar GridSearchCV para cada modelo
    for name, model in models.items():
        print(f"Ajustando hiperparâmetros para {name} ...")
        # Definir o param_grid específico para o modelo atual
        param_grid: dict = param_grids[name]
        # Criar o objeto GridSearchCV
        grid_search = GridSearchCV(model, param_grid, cv=5, scoring='accuracy')
        # Executar a busca em grade no conjunto de treinamento
        grid_search.fit(tfidf_train, y_train)
        # Armazenar o melhor modelo ajustado
        best_models[name] = grid_search.best_estimator_
        # Mostrar os melhores parâmetros encontrados
        print(f"Melhores parâmetros encontrados para {name}:")
        print(grid_search.best_params_)
        print()

    # Avaliar os melhores modelos no conjunto de teste e mostrar métricas
    print("\nAvaliação dos melhores modelos no conjunto de teste:")

    for name, model in best_models.items():
        # Avaliar o modelo no conjunto de teste
        accuracy = model.score(tfidf_test, y_test)
        print(f'{name}: Acurácia no teste = {accuracy:.4f}')

hyperparameter_tuning_and_model_evaluation(best_models, models, param_grids,
tfidf_train, y_train)

```

Ajustando hiperparâmetros para Logistic Regression ...

```

/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:
ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:
ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
  warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:

```

```

ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:
ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:
ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/svm/_base.py:1237:
ConvergenceWarning: Liblinear failed to converge, increase the number of
iterations.
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/svm/_base.py:1237:
ConvergenceWarning: Liblinear failed to converge, increase the number of
iterations.
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/svm/_base.py:1237:
ConvergenceWarning: Liblinear failed to converge, increase the number of
iterations.
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:
ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:
ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:
ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:
ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
    warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/linear_model/_sag.py:350:
ConvergenceWarning: The max_iter was reached which means the coef_ did not
converge
    warnings.warn(

```

```

Melhores parâmetros encontrados para Logistic Regression:
{'C': 10.0, 'penalty': 'l2', 'solver': 'liblinear'}

```

Ajustando hiperparâmetros para Random Forest ...

Melhores parâmetros encontrados para Random Forest:

```
{'max_depth': None, 'min_samples_split': 2, 'n_estimators': 100}
```

Ajustando hiperparâmetros para SVM ...

Melhores parâmetros encontrados para SVM:

```
{'C': 10.0, 'gamma': 'scale', 'kernel': 'rbf'}
```

Ajustando hiperparâmetros para Naive Bayes ...

Melhores parâmetros encontrados para Naive Bayes:

```
{'alpha': 0.1, 'fit_prior': True}
```

Ajustando hiperparâmetros para KNN ...

```
/opt/conda/lib/python3.11/site-  
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring  
failed. The score on this train-test partition for these parameters will be set  
to nan. Details:
```

```
Traceback (most recent call last):
```

```
File "/opt/conda/lib/python3.11/site-  
packages/sklearn/model_selection/_validation.py", line 982, in _score  
    scores = scorer(estimator, X_test, y_test, **score_params)  
            ~~~~~
```

```
File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",  
line 253, in __call__  
    return self._score(partial(_cached_call, None), estimator, X, y_true,  
    **_kwargs)  
~~~~~
```

```
File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",  
line 345, in _score  
    y_pred = method_caller(  
            ~~~~~
```

```
File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",  
line 87, in _cached_call  
    result, _ = _get_response_values(  
            ~~~~~
```

```
File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",  
line 210, in _get_response_values  
    y_pred = prediction_method(X)  
            ~~~~~
```

```
File "/opt/conda/lib/python3.11/site-  
packages/sklearn/neighbors/_classification.py", line 259, in predict  
    probabilities = self.predict_proba(X)  
            ~~~~~
```

```
File "/opt/conda/lib/python3.11/site-  
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba  
    probabilities = ArgKminClassMode.compute(  
            ~~~~~
```

```
File "/opt/conda/lib/python3.11/site-
```

```

packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
    ~~~~~
ValueError: invalid literal for int() with base 10: 'form'

    warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
    ~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
    ~~~~~
ValueError: invalid literal for int() with base 10: 'form'

```

```

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
    ~~~~~
  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
    ~~~~~
  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
    ~~~~~
  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
    ~~~~~
  File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
    ~~~~~
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
    ~~~~~
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
    ~~~~~
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
    ~~~~~
ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):

```

```

File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
    ~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
    ~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
    ~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
    ~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
    ~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
    ~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
    ~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
    ~~~~~

ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__

```



```

    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
~~~~~
File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
        ~~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
        ~~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
        ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
        ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
        ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
        ~~~~~~

ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
        ~~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
~~~~~
File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(

```

```

~~~~~
File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
~~~~~
ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",

```

```

line 210, in _get_response_values
    y_pred = prediction_method(X)
            ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
                  ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
                  ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
                        ~~~~~~

ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
            ~~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
            ~~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
              ~~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/neighbors/_classification.py",
line 259, in predict
    probabilities = self.predict_proba(X)
                  ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)

```

```

~~~~~
File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
    ~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
    ~~~~~

ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
    **_kwargs)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
    ~~~~~

File "/opt/conda/lib/python3.11/site-

```

```

packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
    ~~~~~
ValueError: invalid literal for int() with base 10: 'form'

    warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
    ~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
    ~~~~~
ValueError: invalid literal for int() with base 10: 'form'

```

```

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
    ~~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
    ~~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
    ~~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
    ~~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
    ~~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
    ~~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
    ~~~~~~

  File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
    ~~~~~~
ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):

```

```

File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
    ~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
    ~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
    ~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
    ~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
    ~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
    ~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
    ~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
    ~~~~~

ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
    ~~~~~

  File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__

```

```

    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
~~~~~
File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
        ~~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
        ~~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
        ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
        ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
        ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
        ~~~~~~

ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
        ~~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
~~~~~
File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(

```



```

~~~~~
File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",
line 210, in _get_response_values
    y_pred = prediction_method(X)
~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
~~~~~
ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py:993: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/lib/python3.11/site-
packages/sklearn/model_selection/_validation.py", line 982, in _score
    scores = scorer(estimator, X_test, y_test, **score_params)
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 253, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 345, in _score
    y_pred = method_caller(
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/metrics/_scorer.py",
line 87, in _cached_call
    result, _ = _get_response_values(
~~~~~

File "/opt/conda/lib/python3.11/site-packages/sklearn/utils/_response.py",

```

```

line 210, in _get_response_values
    y_pred = prediction_method(X)
            ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 259, in predict
    probabilities = self.predict_proba(X)
                  ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/neighbors/_classification.py", line 343, in predict_proba
    probabilities = ArgKminClassMode.compute(
                  ~~~~~~

File "/opt/conda/lib/python3.11/site-
packages/sklearn/metrics/_pairwise_distances_reduction/_dispatcher.py", line
590, in compute
    unique_Y_labels=np.array(unique_Y_labels, dtype=np.intp),
                        ~~~~~~

ValueError: invalid literal for int() with base 10: 'form'

warnings.warn(
/opt/conda/lib/python3.11/site-packages/sklearn/model_selection/_search.py:1051:
UserWarning: One or more of the test scores are non-finite: [0.26776798
0.29303673 0.24997587 0.26776359 0.2322408  0.24346014
          nan 0.25277961          nan 0.23218376          nan 0.22376377]
warnings.warn(

Melhores parâmetros encontrados para KNN:
{'metric': 'euclidean', 'n_neighbors': 3, 'weights': 'distance'}

Ajustando hiperparâmetros para Gradient Boosting ...
Melhores parâmetros encontrados para Gradient Boosting:
{'learning_rate': 0.5, 'max_depth': 10, 'n_estimators': 200}

```

Avaliação dos melhores modelos no conjunto de teste:

Logistic Regression: Acurácia no teste = 0.7453

Random Forest: Acurácia no teste = 0.7116

SVM: Acurácia no teste = 0.7341

Naive Bayes: Acurácia no teste = 0.7566

KNN: Acurácia no teste = 0.2846

Gradient Boosting: Acurácia no teste = 0.6966

- Seleção do Melhor Modelo: Escolha o modelo com o melhor desempenho nos dados de validação/teste.
- Retreinamento do Melhor Modelo: Retreine o modelo selecionado usando todos os dados disponíveis (conjunto de treinamento e teste combinados, se apropriado).
- Avaliação Final: Faça uma avaliação final do modelo em um conjunto de dados de teste ou validação separado, se disponível, para garantir que o modelo generaliza bem.
- Salvamento do Modelo: Salve o modelo treinado para uso futuro.
- Implantação: Implante o modelo em um ambiente de produção, onde ele pode ser usado para

fazer previsões sobre novos dados.

- Monitoramento e Manutenção: Monitore o desempenho do modelo em produção e faça ajustes conforme necessário.

```
[78]: from sklearn.naive_bayes import GaussianNB
import joblib

def transform_text_data(tfidf, X_train, X_test):
    tfidf_train = tfidf.fit_transform(X_train).toarray()
    tfidf_test = tfidf.transform(X_test).toarray()
    return tfidf_train, tfidf_test

def initialize_models(best_params):
    models = {
        'Logistic Regression': LogisticRegression(**best_params['Logistic_
↪Regression']),
        'Random Forest': RandomForestClassifier(**best_params['Random Forest']),
        'SVM': SVC(**best_params['SVM']),
        'Naive Bayes': GaussianNB(var_smoothing=best_params['Naive_
↪Bayes']['alpha']),
        'KNN': KNeighborsClassifier(**best_params['KNN']),
        'Gradient Boosting': GradientBoostingClassifier(**best_params['Gradient_
↪Boosting'])
    }
    return models

def train_and_evaluate_models(models, tfidf_train, y_train, tfidf_test, y_test):
    best_model_name = None
    best_accuracy = 0
    for name, model in models.items():
        print(f"Treinando {name} ...")
        model.fit(tfidf_train, y_train)
        y_pred = model.predict(tfidf_test)
        accuracy = accuracy_score(y_test, y_pred)
        print(f'{name}: Acurácia no teste = {accuracy:.4f}')
        if accuracy > best_accuracy:
            best_accuracy = accuracy
            best_model_name = name
    return best_model_name, best_accuracy

def retrain_best_model(best_model_name, models, tfidf, df):
    best_model = models[best_model_name]
    X_combined = tfidf.fit_transform(df['text']).toarray()
    best_model.fit(X_combined, df['class_img'])
    return best_model
```

```

def save_model_and_vectorizer(model, tfidf, model_filename,
    ↪vectorizer_filename):
    joblib.dump(model, model_filename)
    joblib.dump(tfidf, vectorizer_filename)
    print(f'\nModelo salvo como {model_filename}')

# Definir os melhores parâmetros encontrados
best_params = {
    'Logistic Regression': {'C': 10.0, 'penalty': 'l2', 'solver': 'saga'},
    'Random Forest': {'max_depth': None, 'min_samples_split': 2, 'n_estimators':
    ↪200},
    'SVM': {'C': 10.0, 'gamma': 'scale', 'kernel': 'rbf'},
    'Naive Bayes': {'alpha': 0.1, 'fit_prior': False},
    'KNN': {'metric': 'euclidean', 'n_neighbors': 3, 'weights': 'distance'},
    'Gradient Boosting': {'learning_rate': 0.5, 'max_depth': 5, 'n_estimators':
    ↪200}
}

# Supondo que `tfidf`, `X_train`, `X_test`, `y_train`, `y_test`, `df` já
    ↪estejam definidos

tfidf_train, tfidf_test = transform_text_data(tfidf, X_train, X_test)
models = initialize_models(best_params)

best_model_name, best_accuracy = train_and_evaluate_models(models, tfidf_train,
    ↪y_train, tfidf_test, y_test)
print(f'\nMelhor modelo: {best_model_name} com acurácia de {best_accuracy:.4f}')
best_model = retrain_best_model(best_model_name, models, tfidf, df)
save_model_and_vectorizer(best_model, tfidf, 'ML_best_model_shaz13.pkl',
    ↪'tfidf_vectorizer_shaz13.pkl')

```

```

Treinando Logistic Regression ...
Logistic Regression: Acurácia no teste = 0.7416
Treinando Random Forest ...
Random Forest: Acurácia no teste = 0.7004
Treinando SVM ...
SVM: Acurácia no teste = 0.7341
Treinando Naive Bayes ...
Naive Bayes: Acurácia no teste = 0.7266
Treinando KNN ...
KNN: Acurácia no teste = 0.2509
Treinando Gradient Boosting ...
Gradient Boosting: Acurácia no teste = 0.6891

Melhor modelo: Logistic Regression com acurácia de 0.7416

Modelo salvo como ML_best_model_shaz13.pkl

```