

## 一、实验目的

- ## 二、实验任务

完成一个支持通配查询处理的检索系统

### 三、使用环境

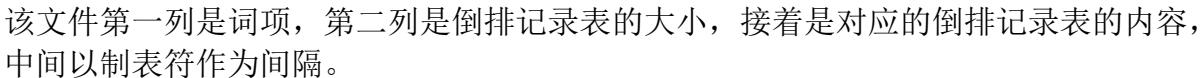
pycharm

python 3.7

#### 四、实验过程

## 1.构建轮排 索引

基于实验 2，我们已经构建了数据集的倒排索引表，如下图所示：



在此基础上，我们在字符集中引入一个新的符号\$，用于表示词项结束，得到扩展词项。然后构建一个轮排索引，对扩展词项的每一个旋转结果都构造一个指针来指向原始词项，得到轮排索引文件 `PermutermIndex.txt`。

构建轮排索引实现在 PermutermIndex.py 中。

首先创建字典 tokens 和 count，tokens 存储词项和倒排索引的列表[docID1、docID2...], count 存储词项和去重后对应的倒排索引表大小。

```
import os

tokens = {}
count = {}
path = r"D:\大四上\信息检索\pro02\data\数据"
path2 = "D:/大四上/信息检索/pro02/data/数据/"

for root, dirs, files in os.walk(path):
    for name in files:
        docID = name.split('.')[0] # 取出.txt前面的部分，记录为docID
        with open(path2 + name, "r") as f: # 打开文件
            text = f.read().lower() # 按要求把单词都转成小写
            words = text.split() # 直接用空格作为分隔符
            f.close()
            for word in words: # 用字典tokens存{word,docID}键值对
                if word in tokens:
                    tokens[word].append(docID)
                else:
                    tokens[word] = [docID]

for temp in tokens:
    tokens[temp] = list(set(tokens[temp]))
    count[temp] = len(tokens[temp])
    tokens[temp].sort()
```

table0 是 count 字典按照值的大小排序的结果，将出现次数排名前一百的词项从 tokens 字典和 count 字典去除。

```
table0 = sorted(count.items(), key=lambda obj:obj[1])
table0.reverse()

#把出现次数前100的排名去掉
k = 0
for c in table0:
    if k >= 100:
        break
    else:
        #print(c[0]+"\n")
        tokens.pop(c[0])
        count.pop(c[0])
    k += 1
```

定义 rotate 函数，返回字符串在第 n 位旋转的结果：

```
def rotate(str, n):
    return str[n:] + str[:n]
```

dkey 是扩展词项，对扩展词项每一位旋转得到 out，再将 out 和 key 以空格分隔，写入 PermutermIndex.txt，轮排索引构建完毕。

```
file = open("D:\大四上\信息检索\pro03\WildcardQuery\PermutermIndex.txt", "w")
keys = tokens.keys()
for key in sorted(keys):
    dkey = key + "$"
    for i in range(len(dkey)-1):
        out = rotate(dkey, i)
        file.write(out)
        file.write(" ")
        file.write(key)
        file.write("\n")
file.close()
```

## 2.实现通配符查询系统

对输入的查询分为 8 种情况

- 对于 X, 查询 X\$, 即实现精确查询，输入 hello, 查询 hello
- 对于 X\*, 查询 \$X\*, 输入 hel\*, 前缀匹配 \$hel
- 对于 \*X, 查询 X\$\*, 输入\*lo, 前缀匹配 lo\$
- 对于 \*X\*, 查询 X\*, 输入\*ll\*, 前缀匹配 ll
- 对于 X\*Y, 查询 Y\$X\*, 输入 h\*o, 前缀匹配 o\$h
- 对于\*, 不支持该查询
- 对于查询中多于一个符号\*出现，不支持该查询
- 对于 exit, 退出该查询系统

具体实现在 query.py 中：

首先在主函数中实现一个外循环，直到输入"exit"循环结束。

parts 是以符号\*作为分割符，得到的子字符串列表

- 1.判断输入的 query 中是否包含\*符号，如果不包含，将 1 作为标记传入 WildCardQuery 函数，表示查询符合情况 1，实现精准查询
- 2.如果 parts 包含三个子字符串且只有中间不是空字符串，符合情况 4，查询 \*X\*
- 3.如果不满足上述条件且 parts 的大小大于 2，说明出现了两个及以上的\*符号，不支持该查询，直接给出提示信息
- 4.如果上述情况皆不满足，说明 parts 刚好被分割为两个子字符串

- (1)如果第一个字符串为空，第二个字符串不为空，符合查询\*X，进入情况 3
- (2)如果第一个字符串不为空，第二个字符串为空，符合查询 X\*，进入情况 2
- (3)如果第一个字符串不为空，第二个字符串不为空，符合查询 X\*Y，进入情况 5
- (4)如果两个字符串都为空，查询仅包含符号\*，输出提示信息

```
if __name__ == '__main__':
    query = input("您可以开始通配符查询：")
    while query != "exit":
        parts = query.split("*")
        if '*' not in query:
            print("通配符查询需要包含符号*，为您精确查询...")
            WildCardQuery(parts, 1)
        elif len(parts) == 3 and parts[0] == '' and parts[2] == '':
            WildCardQuery(parts, 4)
        elif len(parts) > 2:
            print("只支持一个通配符查询，请确认输入有且仅包含一个符号*")
        else:
            if parts[0] == '' and parts[1] != '':
                WildCardQuery(parts, 3)
            elif parts[0] != '' and parts[1] == '':
                WildCardQuery(parts, 2)
            elif parts[0] != '' and parts[1] != '':
                WildCardQuery(parts, 5)
            else:
                print("不支持仅包含*的查询！")

        query = input("可以继续通配符查询：")
    print("退出成功！")
```

在 WildCardQuery 函数中按照上面所述规则实现对查询的修改，即根据不同情况在合适的位置加上符号\$，修改后的查询内容只要进行前缀匹配即可，这一部分在 processQuery 函数中实现。

```
def WildCardQuery(paras, case):  
  
    if case == 1:  
        query = paras[0] + "$"  
        processQuery(query, False)  
    elif case == 2:  
        query = "$" + paras[0]  
        processQuery(query, True)  
    elif case == 3:  
        query = paras[1] + "$"  
        processQuery(query, True)  
    elif case == 4:  
        query = paras[1]  
        processQuery(query, True)  
    else:  
        query = paras[1] + "$" + paras[0]  
        processQuery(query, True)
```

在 processQuery 函数中，如果 flag 被置为 True，对传入的 query 在轮排索引中实现前缀匹配；如果 flag 是 False，对传入的 query 在轮排索引中实现完全匹配。其中，前缀匹配由 startswith 函数实现。所有的结果存储在 term\_list 列表中，将他们按格式打印出来。

注意这里 t1 是调用 from BTrees.OOBTree import OOBTree 后实例化的 OOBTree，用于引用 B 树加快查询速度。

```

def processQuery(query, flag):
    #temp_count = {}
    term_list = []
    for tk in t1.keys():
        if flag == True:
            #print(1)
            if tk.startswith(query):
                term_list.append(t1[tk])
                #temp_count[permuterm[tk]] = int(count[permuterm[tk]])
            else:
                if tk == query:
                    term_list.append(t1[tk])
                    #temp_count[permuterm[tk]] = int(count[permuterm[tk]])
    #table = sorted(temp_count.items(), key=lambda x: x[1], reverse=True)
    #temp_count_sorted = dict(table)

    #print(' '.join(str(i) for i in temp_count_sorted))
    #print(' '.join(str(temp_count_sorted[i]) for i in temp_count_sorted))
    print(' '.join(str(i) for i in term_list))

```

接下来得到所有输出词项中第一项和第二项的 OR 结果的输出文档名列表。

t2 是根据倒排索引表实例化的 OOBTree，在其中找到第一项和第二项词项对应的所有 docID 列表，并加入 docID 列表，如果列表为空，输出没有查询结果；否则得到二者的并集再按照格式输出

```

docID = []
i=0
for term in term_list:
    if i == 0 or i == 1:
        docID.append(t2[term])
    i += 1

if len(docID) == 0:
    print("没有匹配的结果...")
    return
if len(docID) >= 2:
    result = list(set(docID[0]).union(set(docID[1])))
else:
    result = list(set(docID[0]))
result.sort()
print(' '.join(str(i) for i in result))
# print(temp)

return

```

### 3.测试结果

自己尝试的测试用例：（包括考虑的上述 8 种情况）

```

Run: query
D:\大四上\信息检索\pro83\WildcardQuery\venv\Scripts\python.exe D:/大四上/信息检索/pro83/WildcardQuery/query.py
您可以开始通配符查询: hello
通配符查询需要包含符号*, 为您精确查询...
hello
6642 6756 6870 15021 16741 18910 23364 27678 35572 36332 37215 38035 47145 50567 52656 55797 66082 71649 84321 87779
可以继续通配符查询: ja
ja ja_rule jaa jaafari jaakko jaan jaan-e-mann jaap jaap_de_hoop_scheffer jaap_penraat jaap_stam jaar jab jabalia jabaliya jabar jabari jabbed jabber jabbo jaber jabez jablon jablon
15978 20091 24316 31734 39156 44921 59051 59635 68673 73671 75057 80942
可以继续通配符查询:
11-movement 20-installment 5,800-apartment ament abandonment abatement aberdeen_asset_management abolishment accompaniment accomplishment accouterment achievement acknowledgment ac
65109 66074
可以继续通配符查询:
aa aaa al-qusaa al-sharaa baa bambastaa bekaa diaa farouk_al-sharaa gamaa gauslaa ghurabaa gomaia hautamaa honokaa i-aa jaa kaa kazaa komunyakaa lataa masnaa nafaa odadaa paa rajaa
2423 2507 2556 2669 2718 6455 9080 18437 18853 23449 23634 25112 26041 27069 27770 30546 30628 30800 30913 31181 32897 36329 36486 36584 37206 37548 38814 38597 41581 42596 42934 4
可以继续通配符查询:
healthco heather_matarazzo heather_raffo heberto heino helen_shapiro helgemo helio hello henao hengelo henoeh-ryugo hero heriberto herman_badillo hermosillo hernando hernan_crespo
21810 49812 62320
可以继续通配符查询:
不支持仅包含*的查询!
可以继续通配符查询:
只支持一个通配符查询, 请确认输入有且仅包含一个符号*
可以继续通配符查询:
退出成功!

Process finished with exit code 0
|

```

您可以开始通配符查询：hello

通配符查询需要包含符号\*，为您精确查询...

hello

6642 6756 6870 15021 16741 18910 23364 27678 35572 36332 37215 38035 47145  
50567 52656 55797 66082 71649 84321 87779

可以继续通配符查询：ja\*

ja ja\_rule jaa jaafari jaakko jaan jaan-e-mann jaap jaap\_de\_hoop\_scheffer jaap\_penraat  
jaap\_stam jaar jab jabalia jabaliya jabar jabari jabbed jabber jabbo jaber jabez jablon jablonski  
jablonsky jabr jabra jabran jabre jabrowski jabs jac jaca jacek jacey jacinda jacine jacinta  
jacinto jacir jack jack-in-the-box jack-o jack-o-lantern jack-of-all-trades jack\general  
jack\_aker jack\_benny jack\_brewer jack\_buck jack\_bush jack\_carter jack\_cassidy  
jack\_coleman jack\_daniel jack\_davenport jack\_davis jack\_del\_rio jack\_dempsey jack\_dunphy  
jack\_elliott jack\_giaraputo jack\_gould jack\_ham jack\_higgins jack\_ingram jack\_johnson  
jack\_jones jack\_kemp jack\_kennedy jack\_kerouac jack\_kirby jack\_knight jack\_kruschen  
jack\_larson jack\_lengyel jack\_lester jack\_ma jack\_mccarthy jack\_mccollough jack\_mcduff  
jack\_mckee jack\_moran jack\_morris jack\_newfield jack\_nicholson jack\_nicklaus  
jack\_nitsche jack\_noseworthy jack\_o'brien jack\_palance jack\_pardee jack\_plotnick  
jack\_ramsay jack\_reed jack\_roberts jack\_ruby jack\_ryan jack\_smight jack\_stephens  
jack\_straw jack\_teagarden jack\_tworkov jack\_vettriano jack\_williamson jack\_womack  
jackaay jackal jackass jacket jacketed jackets jacki jackie jackie\_collins jackie\_curtis  
jackie\_fields jackie\_gleason jackie\_hoffman jackie\_joyner-kersee jackie\_mason jackie\_mclean  
jackie\_robinson jackie\_robinson\_stadium jackie\_wilson jacking jackman jacknow jacko  
jackpot jackrabbit jacks jackson jackson-browns jackson-george jackson-lee jackson\_barracks  
jackson\_browne jackson\_heights jackson\_heights-elmhurst jackson\_hole jackson\_pollock  
jackson\_river jackson\_vroman jacksonian jacksonians jacksonville jacksonville\_jaguars  
jacksonville\_veterans\_memorial\_arena jackucewicz jacky jaclyn jaclyn\_smith jacmel jaco  
jacob jacob\_d.\_robida jacob\_druckman jacob\_e.\_smart jacob\_grimm jacob\_hacker  
jacob\_k.\_javits jacob\_landau jacob\_lawrence jacob\_mincer jacob\_riis jacob\_schiff  
jacob\_talmon jacob\_vargas jacob\_zuma jacobean jacobellis jacobhollefuneralhome.com jacob  
jacobovici jacobowitz jacobs jacobs-corria jacobs-haviland jacobs-inspired jacobsen jacobson  
jacobus jacobus jacque jacque\_vaughn jacqueline jacqueline\_batteast jacqueline\_bisset  
jacqueline\_kennedy\_onassis jacquelyn jacques jacques-henri jacques-imo jacques-louis\_david

jacques-marie jacques-pierre jacques\_attali jacques\_audiard jacques\_barzun jacques\_brel  
jacques\_callot jacques\_chaoulli jacques\_chirac jacques\_delors jacques\_demy  
jacques\_goudstikker jacques\_herzog jacques\_lipchitz jacques\_loeb jacques\_perrin jacques\_p  
茅 pin jacques\_tati jacques\_thibaud jacques\_villeneuve jacques\_villon jacquesimosnyc.com  
jacquet jacqui jacqui\_cooper jacquie jacquier jacquot jacuzzi jacyln jada jada\_pinkett\_smith  
jadakiss jadallah jade jade-hilted jaded jaden jadid jadow jadrolinija jadue-lillo jady jae  
jaeckie jaedene jaeger jaeger-lecoultre jafar jafar\_panahi jafari jafarzadeh jaff jaffa jaffe jaffe-  
braunstein jaffe-friede jaffee jaffery jaffna jaffray jaffree jaffrey jaffy jafr jagan jaganath jagdip  
jagdish jagdish\_bhagwati jagged jagger jagir jaglom jagodinski jagr jagrata jaguar jaguars  
jahanbegloo jahangir jaheim jahjah jahn jahnke jahore jai jai\_rodriguez jaideep jaidon  
jaidon\_codrington jail jailed jailer jailers jailhouse jailing jails jaime jaime\_moreno jaimee  
jaimes jain jaipur jairo jairzinho jaish jaish-e-mohammed jaish-e-muhammad jajuana jak  
jakarta jake jake\_cherry jake\_heggie jake\_kasdan jake\_lamotta jake\_plummer jake\_schifino  
jake\_siewert jake\_steinfeld jakes jakob jakob\_eklund jakobson jakub jakub\_klepis jal jala jalal  
jalal-abad jalal\_jalalizadeh jalalabad jalali jalaluddin\_haqqani jalen jaleo jalia jalil jalisco  
jalloh jalopies jam jam-band jam-packed jam-session jama jamaal jamaal\_magloire  
jamaar\_taylor jamaat jamaat-e-islami jamaica jamaica\_hills jamaica\_kincaid jamaica\_plain  
jamaican jamaican-born jamaicans jamal jamal\_crawford jamal\_jones jamal\_lewis  
jamal\_strong jamall jamar jamar\_martin jamar\_nesbit jamarcus jamba jambalaya jamboree  
jame jamea jamea\_jackson jameat jameel jamel jamel\_debbouze jamel\_white james james-  
lange james\_bartender james\_a.\_baker james\_a.\_michener james\_a.\_michener\_art\_museum  
james\_abdnor james\_agee james\_alan\_mcperson james\_algar james\_allchin james\_amann  
james\_arness james\_atlas james\_avery james\_b.\_jacobs james\_badge\_dale james\_baker  
james\_baldwin james\_barbour james\_beard james\_bennet james\_bentley james\_berry  
james\_blood\_ulmer james\_bond james\_boswell james\_boyle james\_braddock james\_brady  
james\_bridges james\_broderick james\_brooks james\_brown james\_bryant  
james\_buchanan\_duke james\_bulger james\_c.\_healey james\_c.\_russell james\_c.\_scott  
james\_callis james\_carlton james\_carpenter james\_carroll james\_castle james\_caviezel  
james\_chace james\_clark james\_cleveland james\_coburn james\_coleman james\_cooley  
james\_crumley james\_curran james\_d.\_stern james\_daly james\_david\_cain james\_davis  
james\_dawson james\_dean james\_dennis james\_denton james\_dickey james\_dobson  
james\_dolan james\_doohan james\_dyson james\_e.\_davis james\_e.\_tierney james\_elliott  
james\_ellroy james\_ensor james\_f.\_albaugh james\_f.\_gennaro james\_fisher james\_flanagan  
james\_fletcher james\_forten james\_fox james\_frain james\_franciscus james\_franco  
james\_franklin james\_frawley james\_g.\_watt james\_gandolini james\_gardner james\_garner  
james\_gleick james\_goldman james\_grant james\_greenwood james\_gustave\_speth  
james\_h.\_scheuer james\_hansen james\_heckman james\_hewitt james\_hill james\_hodgins  
james\_horner james\_howard james\_huang james\_j.\_braddock james\_johnson james\_keach  
james\_kent james\_kim james\_king james\_kugel james\_l.\_wattenbarger james\_lassiter  
james\_lee\_burke james\_lincoln\_collier james\_lomas james\_loney james\_long james\_m.\_cain  
james\_m.\_cox james\_m.\_harvey james\_m.\_kelly james\_m.\_kilts james\_m.\_lindsay  
james\_m.\_loy james\_m.\_mcperson james\_m.\_schneider james\_m.\_walton james\_macdonald  
james\_macmillan james\_madison james\_mangold james\_manning james\_martin james\_mason  
james\_mcmurtry james\_mcneill\_whistler james\_mcteigue james\_merrill  
james\_michael\_curley james\_monroe james\_murphy james\_murray james\_naismith  
james\_newton james\_newton\_howard james\_nolan james\_oddo james\_p.\_gorman  
james\_p.\_hoffa james\_park james\_parkinson james\_patterson james\_phelan james\_phillips  
james\_quinn james\_r.\_hedges\_iv james\_r.\_schlesinger james\_remar james\_risen james\_river  
james\_river\_coal\_company james\_robert\_brown james\_robertson james\_rosenquist  
james\_rubin james\_s.\_tisch james\_saito james\_sample james\_sanders james\_schamus



james\_schuyler james\_sherley james\_siegel james\_sims james\_spader james\_steinberg  
james\_stevenson james\_stewart james\_stirling james\_t.\_farrell james\_t.\_walsh james\_taylor  
james\_thornton james\_thurber james\_tissot james\_tobin james\_toney james\_upton  
james\_vacca james\_w.\_carey james\_walsh james\_wan james\_ware james\_weldon\_johnson  
james\_west james\_whale james\_whitaker james\_whitmore james\_wolfensohn  
james\_wong\_howe james\_wright james\_wyatt jamesbeck jamesburg jameses jamesian  
jameson jamesport jamestown jamey jamey\_carroll jamgotchian jami jamiat jamiat-qurania-  
arabia jamie jamie\_brown jamie\_carragher jamie\_davis jamie\_foxx jamie\_harris jamie\_hewlett  
jamie\_kaler jamie\_kennedy jamie\_king jamie\_lee\_curtis jamie\_lundmark jamie\_mcintyre  
jamie\_moyer jamie\_niven jamie\_o'neal jamie\_parker jamie\_sal 茅 jamie\_selkirk  
jamie\_stewart jamie\_storr jamie\_walker jamie\_watson jamie\_wyeth jamiel jamieson jamil  
jamil\_walker jamille jamin jamire jamiroquai jamison jamison\_brewer jamiya jamkaran  
jammed jammer jamming jammu jamnagar jamont jampacked jamrock jams jamuna jam 贸 n  
jan jan-ake jan-erik jan-marc jan. jan\_baalsrud jan\_eliasson jan\_gehl jan\_hernych jan\_koller  
jan\_kott jan\_maxwell jan\_sawka jan\_schakowsky jan\_ullrich jan\_van\_eyck jan\_veizer jana  
jana\_bennett janabi janacek janadas\_devan janaina janaki jananne janas janata janberg jance  
jandek jandl jane jane\_alexander jane\_anderson jane\_austen jane\_birkin jane\_campion  
jane\_cox jane\_elizabeth\_hodgson jane\_engelhard jane\_eyre jane\_fonda jane\_friedman  
jane\_froman jane\_gardam jane\_goldman jane\_goodall jane\_jacobs jane\_jacobs-inspired  
jane\_kaczmarek jane\_kenyon jane\_kim jane\_kober jane\_lawrence jane\_lubchenco jane\_mayer  
jane\_pauley jane\_perlez jane\_russell jane\_shepherdson jane\_siberry jane\_swift  
jane\_voorhees\_zimmerli\_art\_museum janeane janeane\_garofalo janeene janeiro janel janelle  
janes janet janet\_davis janet\_evanovich janet\_flanner janet\_jackson janet\_jagan  
janet\_l.\_robinson janet\_maslin janet\_napolitano janet\_reno janet\_street-porter janet\_suzman  
janette janeway janey janez jang jangel jangl jangled jangling jangly jani janica janice  
janice\_stein janicke janie janiey janikowski janina janine janis janis\_joplin janissary  
janiszewski janitor janitorial janitors janjaweed jankoski jankowski jann jann\_wenner janna  
janne\_ahonen jannero jannero\_pargo janney jannie jannings jannis\_kounellis jannus janofsky  
janollari janos janov janovic janowicz janowski jansa jansen jansma jansons janssen jansson  
jantho january january\february janus janusian janusz janvrin janx jaoura japan japan-born  
japan-centric japan-china japan-united japan-yoi japan\_tobacco japanese japanese-american  
japanese-americans japanese-born japanese-canadian japanese-canadians japanese-drafted  
japanese-made japanese-owned japanese-south\_korean japanese-style japanese-styled  
japantown jape japonaiserie jaquelin jaqueline jaques jar jara jaraba jaradat jaraisi jaramana  
jaramillo jaran jarblum jarboe jarcho jarden jardim jardin jardine jardines jardini 猫 re jared  
jared\_aulin jared\_bernstein jared\_diamond jared\_dudley jared\_fogle jared\_harris jared\_jeffries  
jared\_kushner jared\_lorenzen jarell jaren\_jackson jaret jaret\_wright jargon jarhead jari jarick  
jarkko jarkko\_nieminen jarkko\_ruutu jarman jarmusch jarndyce jarocin jarome jarome\_iginla  
jaromir jaromir\_jagr jaros jaroslav jarotschin jarrarsupariver.blogspot.com jarred jarrell jarret  
jarrett jarring jarrod jarrod\_washburn jarrow jarry jars jarule jarvi jarvis jarvis\_hayes jascha  
jaseion jasen jasenovac jasey jashni jashon jashon\_sykes jasia jasicki jasikevicius jasiulionis  
jasmin jasmin\_tabatabai jasmine jasna jasnoff jason jason\_bateman jason\_bay jason\_bentley  
jason\_biggs jason\_binn jason\_blake jason\_bloom jason\_bohn jason\_bostic jason\_bulger  
jason\_calacanis jason\_castro jason\_collier jason\_dunham jason\_elam jason\_fabini  
jason\_ferguson jason\_fowler jason\_hart jason\_hawes jason\_hirsh jason\_horton jason\_hunter  
jason\_labarbera jason\_lyons jason\_marquis jason\_maxiell jason\_mcelwain jason\_michaels  
jason\_middlebrook jason\_newsted jason\_patric jason\_priestley jason\_reed jason\_rhoades  
jason\_robards jason\_robert\_brown jason\_rogers jason\_schmidt jason\_schwartzman  
jason\_segel jason\_shaw jason\_smilovic jason\_sokol jason\_spezza jason\_statham jason\_taylor  
jason\_terry jason\_turner jason\_varitek jason\_webster jason\_white jason\_whittle jason\_willis

jason\_witten jason\_wright jasons jaspar jasper jasper\_county jasper\_francis\_cropsey  
 jasper\_johns jasper\_morrison jaspers jasperware jass jassim jatin-lalit jatis jauan jauchem  
 jaume jaume\_plensa jaundiced jaune jaunt jauntily jauntiness jaunts jaunty jauron jaur 猫 s  
 java javad javaid javan javanese javelin javelins javiar javier javier\_aguirresarobe  
 javier\_bardem javier\_c 谩 mara javier\_mar 柳 as javier\_mascherano javier\_rebollo  
 javier\_saviola javiers javish javits javon javon\_walker javor javy javy\_1 贸 pez jaw jaw-  
 dropping jaw-droppingly jawad jawad\_williams jawaharal jawaharlal jawaharlal\_nehru  
 jawaharlal\_nehru\_university jawara jawboning jawed jawed\_karim jawed\_ludin jaweed jawing  
 jawlensky jawline jawole\_willa\_jo\_zollar jaworski jaws jaxon jay jay-alexander jay-z  
 jay\_adelson jay\_berger jay\_bilas jay\_bouwmeester jay\_dardenne jay\_farrar jay\_fiedler  
 jay\_garner jay\_gibbons jay\_grinney jay\_haas jay\_harrington jay\_hernandez jay\_inslee  
 jay\_leonhart jay\_mccarroll jay\_o'brien jay\_o.\_sanders jay\_payton jay\_peak jay\_presson\_allen  
 jay\_r.\_ferguson jay\_rabinowitz jay\_thompson jay\_triano jay\_williams jay\_witasick jaya  
 jayant\_patel jayantha jayapura jayaraman jayawardena jayce jayceon jayde jayden jaye jaygee  
 jayhawk jayhawks jayme jaymes jaymie jayne jayne\_atkinson jayne\_houdyshell  
 jayne\_mansfield jays jayson jayson\_blair jayson\_williams jazeera jazira jazmin  
 jazmin\_grace\_rotolo jazz jazz-festival jazz-inspired jazz-themed jazz-tinged jazze jazzed  
 jazzes jazzfest jazzy

15978 20091 24316 31734 39156 44921 59051 59635 68673 73671 75057 80942

可以继续通配符查询: \*ment

11-movement 20-installment 5,800-apartment ament abandonment abatement  
 aberdeen\_asset\_management abolishment accompaniment accomplishment accouterment  
 achievement acknowledgment acknowledgement adjustment admonishment advancement  
 advertisement agreement agreement ailment alignment allotment amazement  
 amblin\_entertainment amendment amusement anagreement anchor\_bay\_entertainment anger-  
 management announcement anointment anti-environment anti-establishment anti-government  
 antigovernment apartment apollo\_management appeasement appointment argument  
 arraignment arrangement arrondissement arrondissement assessment asset-management  
 assignment assortment astonishment atom\_entertainment atonement attachment attainment  
 augment backbone\_entertainment bandelier\_national\_monument banishment bargain-basement  
 basement befuddlement betterment bewilderment bombardment  
 canyon\_de\_chelly\_national\_monument castle\_clinton\_national\_monument cement  
 chicago\_police\_department classical-instrument clement cl 茅 ment co-payment code-  
 enforcement comment commandment commencement commitment compartment complement  
 compliment concealment confinement consignment containment contentment curtailment  
 dazzlement denouement dental-equipment department deployment derailment dethronement  
 detriment development dick\_clement digital-rights-management dinosaur\_national\_monument  
 disagreement disappointment disarmament disbursement discouragement disenchantment  
 disengagement disfigurement disgruntlement dishevelment disillusionment disinvestment  
 dismantlement dismemberment disparagement displacement divestment document  
 documentary-cum-indictment drug-enforcement easement eastment edutainment element  
 embankment embarrassent embellishment embezzlement embodiment embroilment  
 emergency-management employment empowerment entombment enactment encampment  
 enchantment encirclement encouragement encroachment endangerment endearment  
 endorsement endowment enforcement engagement enhancement enjoyment enlargement  
 enlightenment enlistment enrichment enrollment enshrinement enslavement entanglement  
 entertainment enticement entitlement entombment environment equipment establishment  
 estrangement eton\_park\_capital\_management  
 european\_bank\_for\_reconstruction\_and\_development excitement excrement experiment

farodokument ferment figment firmament foment four-movement fox\_film entertainment  
 fragment fulfillment funimation\_entertainment game-management garment general-assignment  
 general\_instrument generation\_investment\_management good-government government  
 government-to-government grand\_staircase-escalante\_national\_monument harassment  
 hastings\_entertainment health-management health-supplement hemment home-improvement  
 housing-development imagine\_entertainment impeachment impediment implement  
 imprisonment improvement in-development incitement inclement indictment inducement  
 infringement installment instrument interment internment international-development  
 international\_creative\_management intombment inurnment investment involvement  
 jeff\_clement jewel\_cave\_national\_monument john\_f\_kennedy\_school\_of\_government  
 judgment lament labor-management laboratory-equipment law-enforcement lead-abatement  
 ligament limited-engagement liniment little\_bighorn\_battlefield\_national\_monument live-  
 entertainment maltreatment managment management matt\_clement measurement merriment  
 mga\_entertainment micromanagement mismanagement mistreatment moment momument  
 money-management monument movement multigovernment new\_world\_development no-  
 comment nonmonument of-the-moment office\_of\_personnel\_management  
 office\_of\_technology\_assessment ointment ornament overallotment overbrook\_entertainment  
 overdevelopment overstatement pacquement parment parliament pavement payment pediment  
 pergament period-instrument pigment pinnacle\_entertainment placement post-enlightenment  
 post-retirement postponement postretirement pre-retirement predicament pretournament pro-  
 environment pro-government pro-settlement procurement product-placement pronouncement  
 punishment radiological-assessment rapprochement ravishment re-enactment re-enlistment re-  
 establishment readjustment realignment reassessment reassignment recruitment redeployment  
 redevelopment refinement refreshment refurbishment regiment reimbursement reinforcement  
 reinstatement reinvestment relignment repayment replacement replenishment requirement  
 research-and-development resentment resettlement restatement retail-office-entertainment  
 retirement r 茅 giment sacrament sediment segment self-assessment self-enlightenment self-  
 government self-improvement semiretirement sentiment settlement seven-movement sewage-  
 treatment shipment shopping-and-entertainment small-government smaller-government  
 sony\_bmg\_music\_entertainment sony\_computer\_entertainment sony\_pictures\_entertainment  
 sports-management spyglass\_entertainment statement sub-basement supplement sustainment  
 syriacomment temperament temporary-employment tenement testament time-management  
 torment tournament trans-shipment transshipment treatment understatement unemployment  
 united\_states\_agency\_for\_international\_development uranium-enrichment vehement  
 washington\_monument water-treatment weight-management western\_asset\_management  
 wildlife-management wolfgang\_clement wonderment yale\_school\_of\_management

65109 66074

可以继续通配符查询: \*aa\*

aa aaa al-qusaa al-sharaa baa bambaataa bekaa diaa farouk\_al-sharaa gamaa gauslaa  
 ghurabaa gomaa hautamaa honokaa i-aa jaa kaa kazaa komunyakaa lataa masnaa nafaa odadaa  
 paa rajaa sanaa shabaa shebaa usaa tiaa-cref www.aa.com www.flysaa.com alaa\_al-tamimi  
 alaa\_al-aswany aaa daab musaab raab saab shaab shahaab staab akwaaba baaba kaaba  
 baaba\_maal shaaban maabarot max\_raabe raabe al-kaabi paabo isaac maac oscar\_isaac  
 port\_isaac isaac\_bashevis\_singer isaac\_bruce isaac\_larian isaac\_newton isaac\_stern  
 isaac\_wilson aaccess isaach haacke aacm naacp isaacs isaacson al-saad ansyaad assaad elaad  
 gstaad raad rampersaad saad shaad shajaad saad-ghorayeb saad\_bin\_laden saad\_hariri  
 bou\_saada baadasssss saadat aaden al-saadi saadi saadoun\_al-dulaimi hamaady  
 saif\_saaeed\_shaheen aaf schaaf fuamatu-maafala al-jaafari ibrahim\_al-jaafari jaafari faafetai  
 burggraaff graaff taaffe naafi baifuloto praag haag\the haaga jochem\_uytdehaage karaage

haager praagh chunksaah daah jemaah maah pabaah.com maah\_daah\_hey maah\_daah\_hey  
 shaahed aahing www.alaskaair.com aaj al-haaj saajid haak paul-henri\_spaak raak yaak  
 mikheil\_saakashvili saakashvili haake jaakko laakkonen kraakman yaakov afzaal ariaal  
 baaba\_maal dewaal eric\_staal ferrostaal foner-gyllenhaal frans\_de\_waal gyllenhaal jamaal  
 kanaal kim\_staal maal schaal staal stephen\_gyllenhaal tsotsitaal waal jamaal\_magloire  
 hendrik\_ramaala iitaala ramaala shaalan mafaalani baalbeck baalbek daalder ivo\_h.\_daalder  
 waaldijk haale maale maali khaaliq taalman yaalon amin\_maalouf maalouf jan\_baalsrud aalto  
 alvar\_aalto abdul-salaam sadaam salaam www.paam.org salaam\_temple naama naaman  
 daamen aamer saaminki aamodt vlaams braamse aan al-shalaan antwaan ayaan bataan batelaan  
 caan canaan christiaan cole\_haan ghazi\_kanaan hasaan ishaan jaan jochanaan kanaan kiyaan  
 new\_canaan north\_canaan rahsaan suhaan tahaan tony\_kanaan jaan-e-mann shankar-ehsaan-  
 loy christiaan\_huygens antwaan\_randle\_el raanan raanana daane shaanxi aap jaap  
 jeremy\_schaap phil\_schaap schaap jaap\_de\_hoop\_scheffer jaap\_penraat jaap\_stam maaqal  
 alkmaar baar bazaar didaar dora\_maar emaar jaar kraar leefbaar maar paar saar schaar  
 straubhaar tichelaar jamaar\_taylor judy\_baar\_topinka daara malmivaara rautavaara aaranow  
 saarbr 眉 cken abildgaard boregaard edwaard hebsgaard jongewaard kierkegaard kirkegaard  
 klitgaard lisleggaard ordrupgaard peter\_sarsgaard rijkaard sarsgaard sverre\_lodgaard  
 thomas\_dausgaard thorsgaard westergaard tjaarda naarden aardman aardsma david\_aardsma  
 aardvark kaare haaretz baargeld ahtisaari martti\_ahtisaari saari saariaho eero\_saarinen  
 eliel\_saarinen saarinen aarion maariv jaariv paarlberg haarlem kirk\_saarloos saarlouis  
 paarmann aaron hank\_aaron stage/aaron aaron\_betsky aaron\_boone aaron\_burr aaron\_copland  
 aaron\_david\_miller aaron\_dismuke aaron\_francisco aaron\_gavey aaron\_glenn aaron\_guiel  
 aaron\_harang aaron\_harris aaron\_heilman aaron\_hill aaron\_holbert aaron\_lemonick  
 aaron\_lennon aaron\_mckie aaron\_merz aaron\_miles aaron\_murphy aaron\_neville  
 aaron\_peirsol aaron\_rome aaron\_rowand aaron\_russo aaron\_sele aaron\_spelling  
 aaron\_stanford aaron\_willard www.aarondavishall.org aaronovitch david\_aaronovitch aarons  
 slim\_aarons aaronson aarp bazaars vaart maarten baartman www.tacomaartmuseum.org  
 bazaarvoice aar 贸 n aar 贸 n\_galindo claas haas jay\_haas kaas koolhaas maas rem\_koolhaas  
 saravanaas schelhaas tommy\_haas aspaas-montoya haas\_school\_of\_business aase john\_haase  
 maaseik carl\_hiaasen paasewe aasif\_mandvi qaasim karaaslan maasri haass maass maastricht  
 taastrup advocaat anraat chaat cliostraat frans\_van\_anraat jaap\_penraat jamaat kaat  
 luijkenstraat penraat sadaat jamaat-e-islami bambaataa charles\_saatchi saatchi kaaterskill baath  
 shaath wangari\_maathai de-baathfication de-baathification baathist baathists el-maati ondaatje  
 staatliche chaats herenplaats mariaplaats staatsgalerie staatskapelle staatsoper maatta kaavya  
 kaavya\_viswanathan jackaay kraayeveld baayork faaz maazel

2423 2507 2556 2669 2718 6455 9080 18437 18853 23449 23634 25112 26041 27069  
 27770 30546 30628 30800 30913 31181 32897 36329 36486 36584 37206 37548 38014 38597  
 41501 42596 42934 43288 43378 44577 45076 46356 46658 47055 48895 49388 50864 51831  
 54325 55672 56588 56735 57015 57701 58737 60185 60861 61703 62123 62667 63074 72555  
 73301 83557 86620 89345 89350 89352 89357 93379 95438 95439

可以继续通配符查询: he\*o

healthco heather\_matarazzo heather\_raffo heberto heino helen\_shapiro helgemo helio  
 hello henao hengelo henoch-ryugo hero heriberto herman\_badillo hermosillo hernando hern 谩  
 n\_crespo hetero

21810 49812 62320

可以继续通配符查询: \*

不支持仅包含\*的查询!

可以继续通配符查询: **he\*ll\*o**  
 只支持一个通配符查询, 请确认输入有且仅包含一个符号\*  
 可以继续通配符查询: **exit**  
 退出成功!

助教检查时的测试:

```
query
D:\大四上\信息检索\pro03\WildcardQuery\venv\Scripts\python.exe D:/大四上/信息检索/pro03/WildcardQuery/query.py
您可以开始通配符查询: assessment
assessment embarrassment harassment office_of_technology_assessment radiological-assessment reassessment self-assessment
1077 1141 1144 1166 1246 1294 1314 1477 1498 1593 1629 1637 1702 1767 1770 1792 1871 1920 1940 2102 2124 2218 2255 2263 2303 2359 4414 8477 10221 10869 13966 17498 17901 18425 18983
可以继续通配符查询: endea
endear endeared endearing endearingly endearment endeavor endeavors ended endemic ender enderle ends endesa
15885 24634 26089 35837 49260 56858 56871 70866 70879 88850 90895 90903
可以继续通配符查询: paarmann
paarmann pahlmann pat_quinn patrick_quinn penn peter_hermann peter_neumann peter_r._kann pierre_clostermann pohlmann
2432 2435 33479 39664 69260 75058
可以继续通配符查询: bnd
bnd
1324 1949 65420 68608 94117
可以继续通配符查询: |
```

您可以开始通配符查询: **\*sment**  
 assessment embarrassment harassment office\_of\_technology\_assessment radiological-assessment reassessment self-assessment  
 1077 1141 1144 1166 1246 1294 1314 1477 1498 1593 1629 1637 1702 1767 1770 1792 1871 1920 1940 2102 2124 2218 2255 2263 2303 2359 4414 8477 10221 10869 13966 17498 17901 18425 18983 19759 20193 20276 21062 21576 22351 22703 22850 23516 23538 23610 23922 24395 24837 24999 25003 25614 25671 25683 25705 26369 26519 27429 27430 27549 27648 28591 28931 31024 31031 32373 32513 32668 33161 34353 36063 36064 36126 36429 37897 37909 38251 39051 39703 40003 40077 40168 40275 40310 40517 41804 42304 42378 42604 43799 46769 47509 50354 52230 53403 53742 54079 55261 55715 55847 56409 57319 57328 57333 58347 60504 61548 61587 61651 61661 61669 61672 61824 64341 64342 65233 65316 65468 65670 65718 65839 65842 67912 67942 68196 68203 68261 68262 68268 68359 68360 68451 68470 68513 68516 68648 68732 68869 71625 71700 74213 74249 74544 74723 74835 75628 76113 77715 79512 80272 81248 81292 81471 81700 82027 82049 82643 82775 82959 83049 83432 84164 86264 88372 88815 90889 91753 92406 92603 92610 92921 92927 93875 93987 93988 94008 95664 96143 96499 96993 97982  
 可以继续通配符查询: **ende\***  
 endear endeared endearing endearingly endearment endeavor endeavors ended endemic ender enderle ends endesa  
 15885 24634 26089 35837 49260 56858 56871 70866 70879 88850 90895 90903  
 可以继续通配符查询: **p\*nn**  
 paarmann pahlmann pat\_quinn patrick\_quinn penn peter\_hermann peter\_neumann peter\_r.\_kann pierre\_clostermann pohlmann  
 2432 2435 33479 39664 69260 75058  
 可以继续通配符查询: **bn\*d**  
 bnd  
 1324 1949 65420 68608 94117

## 五、总结

1. B 树和轮排索引之间存在密切的联系，虽然我这里采取了调包的方式使用 B 树，但自己实现一个树的类也是可行的。

2. 轮排索引的结构非常简单，由于要保存词项旋转的所有结果，词典会变得很大，引起空间的急剧增长。为此可以引入 k-gram 索引，它是一种倒排索引结构，词典是 k-gram 表，倒排记录表是保存的所有包含该 k-gram 的词项，同样可以支持通配查询。

3. 查询尽可能考虑了用户所有的输入可能，对于多个通配符的查询按照项目要求没有做处理，要处理的话可能会牵涉到穷举法和过滤