### YZ组的信息检索项目总结报告

### 71118313 张晓铮 71118415 叶宏庭

- 一、项目实现过程中遇到的问题及解决方案
- 1. 基于布尔模型的检索实验与研讨

第一次实验的问题主要集中在如何存储倒排索引,不使用数据库的情况下,采用文件存储倒排索引,如果存储为 txt 文件,需要我们自定义存储格式等,并自己写解析的代码,我们希望利用现成的模块,并且非常方便读取阅读等,最终确定了 json 格式的文件。

同时,对于如何处理输入表达式,我们采用了中缀转后缀,并用栈模拟运算的方式,实现了可以任意加括号的功能。

### 2. 基于向量模型的检索实验与研讨

向量模型中,我们采用了书中提到的 ddd.qqq 表示法,让查询与文档的参数全部可调整,为了实现这一目的,并且尽量提高代码的可复用性和可扩展性,我们将各种归一化方法,实现统一接口,利用统一的函数来获取,达到传递类型即可动态的调用相应函数的功能。

3. 基于概率模型的检索实验与研讨

本次实验中,主要问题集中在如何计算 Pi, Ri 值,以及计算 RSV 值中;在计算 RSV 值时会出现负数分数的情况,影响了最后结果的评判。最终采用了 e 指数的方法,将结果映射到正数中,解决了这一问题。

二、项目检索结果对比分析

查询一: 春 燕 柳

查询结果:

布尔模型:

权重方式选择: Utf-idf owf-idf

#### 长恨歌

白居易 (唐代)

#### 老将行

王维 (西代)

少年十五二十时,步行夺得胡马骑。射杀山中白鹤虎。肯敖邺下黄须儿! 一身转战三千里,一剑曾当百万师,汉兵奋击如露 赏,虏骑崩剥畏蒺藜。卫青不观由天幸,李广无功缘数奇。自从弃置使衰朽。世事遂的成白首。昔时飞箭无全目,今日垂杨 生左射,路旁时发战极风。 了朋学种外生粮,在花古木连写著,寄惠亳山对唐卿。 管令谕勒出飞泉,不似颍川空徙酒。 赞兰 山下降如云,郑钦独自夕周,节能之司祭年少,这书五通出将军,试排铁太如雪色,朝持室如刘星文,愿得燕弓射大将,耻令越甲鸣吾君,莫谢旧日云中守,犹堪一战取功勋。

本篇文档符分: 2.41701

#### 春思

皇甫冉 (唐代)

莺鸣燕语报新年,马邑龙堆路几干。家住层城临汉苑,心薤明月到胡天。(层城一作:秦城)机中锦字论长恨,楼上花枝笑独 眠。为问元戎窦车骑,何时返除勒燕然。

**本篇文档得分: 2.39246** 

#### 春思

(金色)

燕草如碧丝,秦桑低绿枝。当君怀归日,是妾断肠时。春风不相识,何事入罗帏。

#### 长安调冯著

客从东方来,衣上灞陵雨。问客何为来,采山因买斧。冥冥花正开,飏飏燕新乳。昨别今已春,鬓丝生几缕。

#### 长相思·其二

李白 (唐代)

日色欲尽花含烟,月明欲素愁不眠。(欲素一作:如素)赵瑟初停凤凰柱、蜀琴欲奏鸳鸯弦。此曲有ê无入传,愿随春风寄蒸然。忆君函迢隔青天。昔日横波目,今作流泪泉。不信妾断肠,归来看取明输前。(断肠一作: 肠断)

### 奉和圣制从蓬莱向兴庆阁道中留春雨中春望之作应制

王维 (唐代)

渭水自紫秦寒曲,黄山旧绕汉宫斜。銮舆迥出干门柳,阁道回看上苑花。云里帝城汉凤阙,雨中春树万人家。为乘阳气行时 令,不是宸游玩物华。(玩一作: 重)

#### 和晋陵陆丞早春游望

杜审点 [唐代]

独有宦游人,偏惊物候新。云霞出海曙,梅柳渡江春。淑气催黄鸟,晴光转绿蘋。忽闻歌古调,归思欲沾巾。

本施文档得分: 1.8702723125990217

#### 奉和中书舍人贾至早朝大明宫

鸡鸣紫貂暖光寒,莺啭凰州春色阑。金阙晓钟开万户,玉阶仙仗拥干官。花迎剑佩雕切落,棕烧旌旗露未干。独有凤凰池上客,阳春一曲和谐难。

#### 洛阳女儿行

洛阳女儿对门居,才可顺容十五余。良人王勒宗總马,侍女金盘脸鲤鱼。画阁朱楼尽相望,红柳绿柳垂幢向。罗帷送上七香 车,重扇迎归九华帐,狂子富贵在青春,息气场看剧学伦。自怜趣王宗教拜,不惜珊瑚诗与人。春霞喟灭九微火,九烟片片 飞花现,戏罢音无理曲时,妆成祇是黑香坐。城中相识尽繁华,日夜经过赵李家。谁怜越女献如玉,贫贱江头自浣纱。





#### 2000

### 闺怨

王昌龄 〔唐代〕

闺中少妇不知愁,春日凝妆上翠楼。(不知一作:不曾)忽见陌头杨柳色,悔教夫婿觅封侯。

本篇文档得分: 0.03569412400057605

#### 征人怨 / 征怨

柳中庸 〔唐代〕

岁岁金河复玉关,朝朝马策与刀环。三春白雪归青冢,万里黄河绕黑山。

本篇文档得分: 0.03569412400057605

#### 寒食

韩翃 [唐代]

春城无处不飞花,寒食东风御柳斜。日暮汉宫传蜡烛,轻烟散入五侯家。

本篇文档得分: 0.03569412400057605

### 和晋陵陆丞早春游望

杜审言 [唐代]

独有宦游人,偏惊物候新。云霞出海曙,梅柳渡江春。淑气催黄鸟,晴光转绿蘋。忽闻歌古调,归思欲沾巾。

本篇文档得分: 0.03569412400057605

#### 新年作

刘长卿 [唐代]

乡心新岁切,天畔独潸然。老至居人下,春归在客先。岭猿同旦暮,江柳共风烟。已似长沙傅,从今又几年。

**本篇文档得分: 0.03569412400057605** 

#### 阙题

刘昚虚〔唐代〕

道由白云尽,春与青溪长。时有落花至,远随流水香。闲门向山路,深柳读书堂。幽映每白日,清辉照衣裳。

本篇文档得分: 0.03569412400057605

### 奉和圣制从蓬莱向兴庆阁道中留春雨中春望之作应制

王维〔唐代

渭水自禁秦塞曲,黄山旧绕汉宫斜。銮舆迥出干门柳,阁道回看上苑花。云里帝城双凤阙,雨中春树万人家。为乘阳气行时令,不是宸游玩物华。(玩一作: 重)

本篇文档得分: 0.03569412400057605

#### 奉和中书舍人贾至早朝大明宫

岑参 〔唐代

鸡鸣紫陌曙光寒,莺啭皇州春色阑。金阙晓钟开万户,玉阶仙仗拥千官。花迎剑佩星初落,柳拂旌旗露未干。独有凤凰池上客,阳春一曲和皆难。

本篇文档得分: 0.03569412400057605

### 赠阙下裴舍人

钱起 [唐代]

二月黄莺飞上林,春城紫禁绕阴阴。(黄莺一作:黄鹂)长乐钟声花外尽,龙池柳色雨中深。阳和不散穷途恨,霄汉长怀捧日心。 献赋十年犹未遇,羞将白发对华籍。

本篇文档得分: 0.03569412400057605

### 东郊

韦应物 [唐代]

吏舍赐终年,出郊門清曙。杨柳散和风,青山澹吾虑。依丛适自憩,缘涧还复去。微雨离芳原,春鸠鸣何处。乐幽心屡止,遵事迹犹遽。终罢斯结庐,慕陶直可庶。

本篇文档得分: 0.03569412400057605



首先,在布尔模型中,可以看到十分偏向长文档,查询结果排名靠前的和查询的春天,燕子和柳树等相关性并不高,查询结果不理想;对于 BIM 模型,可以看到分数都相同,区分度不高,前几条的得分都是一样的,结果没有区分度;向量模型和 MLE 效果都比较好,但相比较来说,MLE 的前十条中,相关度较高的更多也更靠前,效果略好。

查询二: 月 相思布尔模型



### My IR System



日色欲尽花含烟,月明欲素愁不眠。(欲素一作:如素)赵瑟初停凤凰柱、蜀琴欲奏鸳鸯弦。此曲有意无人传,愿随春风寄燕然。忆君迢迢隔青天。昔日横波目,今作流泪泉。不信荽断肠,归来看取明镜前。(断肠一作:肠断)

本義文档得分: 0.3038059595838587

#### BIM

#### My IR System



#### 望月怀远/望月怀古

海上生明月,天涯共此时。情人怨谣夜,竟夕起相思。灭浊怜光满,披衣觉露滋。不堪盈手赠,还瘦梦佳期。

#### 无题•重帏深下莫愁堂

本篇文档例分: 0.14626

#### 寄李儋元锡

去年花里達君別,今日花开又一年。(又一年一作:已一年)世事茫茫难自科,春愁結鏡独成眠。身多疾病思田里,曷有流亡 使俸钱。問題欲來相问訊,西楼望月几问园。

#### 长恨歌

白田県 「唐代]

#### 琵琶行并序 / 琵琶引

白居易 (唐代)

#### 详陈音甫

李颀 (唐代)

四月南风大麦黄,枣花未落柳阴长、青山煎别磨还乳,椰马出门则旧乡。陈梁立身何坦落,虬须虎眉仍大鹅。温中贮书一万卷,不肯低头庄蝉拜,筑门旅临饮我曹。 化羟万嗪四邻毛 新龄不知后日暮,有时空飘死底,长可淡头连天黑,津口停舟渡不得,则重风不以发源,淄阳行子空级。南临城林阳景,爱崖即千分城时。

#### 长相思•其一

长相思,在长安,络纬秋综金井阁,微部凄凄蕾色寒,孤灯不明思欲绝,卷帷望月空长叹,美人如花隔云端!上有青冥之长天,下有渌水之波澜。(长天一作:高天)天长路远难飞苦,梦魂不到关山难。长相思,排心肝!

### 长相思·其二

日色欲尽花念烟,月朔欲素愁不眠。(欲素一作:如素)赵慧初序凤凰柱,蹋琴欲奏鸳鸯弦。此曲有意无人传,愿随春风寄燕然。忆街迢迢阁青天,昔日楼波目,今忤流追唤,不信宴断扬,归来看取明崩崩。(明扬一件: 肠细)

本篇文物得分: 0.14626

#### 竹里馆

独坐幽篁里,弹琴舞长啸。深林人不知,明月来相隔。

本篇文档得分: 0.

#### 望洞庭湖赠张丞相 / 临洞庭湖赠张丞相

八月湖水平,通虚混太清。气蒸云梦泽,波撼岳阳城。欲齐无舟楫,端居耻圣明。坐观垂钓者,徒有羡鱼情。



在这次搜索中,布尔模型和 BIM 模型同样出现了 1 中出现的问题,布尔模型优先选择了长文档,BIM 区分度不高。而这次查询中向量模型要略好,前面的文档相思和月亮都有,后面的偏重相思,而 MLE 中,前几个中含有月和相思,后面的更偏重月,第十个中只有一个相字,相关性并不高,但仍被选中。

# 查询三:金陵布尔模型







## 西塞山怀古

刘禹组 (唐代)

王濬楼船下益州,金陵王气黯然牧。(王渚一作:西晋)干寻铁锁沉江底,一片跨幅出石头。人世几回伤往事,山形依旧枕寒流。今逢四海为家日,故垒萧萧芦荻秋。

#### 金陵洒肆留别

风吹柳花满店香,吴姬压酒劝客尝。(劝客一作:唤客)金陵子弟来相送,欲行不行各尽觞。请君试问东流水,别意与之谁短

本篇文档得分: 0.010327342625832233

### 石鼓歌

韩愈 (唐代)

整生手持石較文、对我试作石能歌。少陵无人源仙死、才薄将奈石鼓闷,周即凌迟四海沸,宣王他起挥天戈、大开明堂受朝 贺,诸侯曾飘鸣相磨。蒐于被死即弹缝役。万里禽兽皆湿罗,魄功勋成告万世。密石件被骤峻缐。从臣才艺咸革一,拣选撰刻 留山间。明州日交野火鸡,鬼物守护项调响。公从90处增旺木,毫发尽备无差化。辞严义密读前除,字体不变非与纲、丰深 是免有缺陋,处场所断生安顺。 郑即凤紫允归作,珊瑚脚吹女侍。 金绳铁茶顿坦吐土 古清张大沙海搜, 短嘴麻布水入、 江维编设无委妃。孔子西行不到秦,将理里宿遗袭境。 "继续好古生西晚,对此涂田以旁沱。忆诸却蒙博士正,其年给攻称元 机、故人从军在右捕,为按理超归44,福河北台密游,加址至室府岩多。毡成原覆可立致,十起只载数跨级,专潜太庙 比郑卿,光价也止百倍过,圣恩若许留太学,诸生讲解得印漆、观经多稀的填烟。坐见举国来奔波。则苗崩峭离节角,安置 安岭平不溉、大厦深爆与重量,经历久远明无佗。中朝大宫老于事,追尚感激往纳艰,牧童散火牛斫角,谁复著手为摩挲。 日销月轻姚毗发,六年西顺空吟雅。截之信书能变崩,数然间可调告颇。维阳从代命必要,未入极治理则那。万今太平日无 事,杨任儒术张任祠。安阳以此尚论列,感情阴口如思河,石鼓之歌止于此,鸣呼吾意具其能论。

### 哀王孙

杜甫 (唐代)

长安城头上白鸟,夜飞延秋门上呼。又向人家啄大屋。屋底达官走避胡。金鞭斯折九乌死,骨肉不得同感怒。腰下宝块青珊,可怜王孙边整隅。问之不肯道姓名。但直围苦乞为奴。已经百日取荆棘。身上无有完即肤、离帝子孙尽隆准,龙种自与常人来,财报任邑龙在野、王孙善保干金妪、不敢长语庙公寨,且为王孙立斯须、昨夜东风吹血腥,乐来秦论渊旧路、朔方使儿好身手。若何勇锐今何惠。窃闻天子已传位,圣德北顺南单于,花门脐面请雪耻,慎勿出口他人组。哀哉王孙慎勿疏,五陵佳气无时无。

#### 哀江头

少検野老吞声哭,春日潜行曲江曲。江头宫殿锁千门,细柳新浦为谁绿?忆昔霓旌下南苑,苑中万物生颜色,昭阳殿里第一人,同琴随着待君侧。辇前才人带弓箭,自马哪哒黄金勒。翻身向天仰射云,一英正坠双飞属。(一笑一作: 一箭明眸始结 今何在?血污游魂归不得。清渭东流剑阙深,去往彼此无消息。人生有情泪沾漉,江水江花岂终极!黄昏胡骑尘满城。欲往 城南蛩咙江。

#### 燕歌行并序

高适 (唐代)

开元二十六年,客有从元戏出寒而还者。作《燕歌行》以示,适感征戌之事,因而和焉,汉家烟尘在东北,汉将游家破残贼。男儿本自重横行,天子非常赐颜色。批金伐鼓下榆关,旌旆逶迤竭石间,校尉羽节飞箭海,单于猎火照狼山,山川萧条极边土,结别疾除於风崩。战士军前半死生,美人除下邓远卿! 大娘穷秋寒草排,孤城落日斗兵梯,身当恩遇解狂敌,力尽关山未群围,犹太远戍车驰之,王蕃应绵则属于。少北场南邻城縣。征入制江空间首,边庭赐城宫即首。 地域花光无所有!杀气三射作阵云,寒声一夜传刁斗,相看白刃血纷纷,死节从来岂顾勋?君不见沙场征战苦,至今就忆李将军!

#### 听筝/鸣筝

鸣筝金栗柱,素手玉房前。欲得周郎顾,时时误拂弦。



本篇文档得分: 0.01629138689511461

这次查询中,前几条的查询结果各个模型表现都不错,但 BIM 和布尔模型还是出现了偏重长文档的问题;向量模型和 MLE 效果都不错,其中向量模型更偏重陵字,而 MLE 更偏重金字,甚至有一个结果只是作者名字中含有金。

### 查询四: 登鹳雀楼

### 布尔模型

My IR System
の関係性
開発: O boolean-model vector-model prob-model
1751C08件 ② 17-401
受鹳雀楼
五年
白日依山尽,黄河入海流。欲穷千里目,更上一层楼。
本載文物得分: 7.3769159625453655
登柳州城楼寄漳汀封连四州/登柳州城楼寄漳汀封连四州刺史 柳河元 (唐代) 此上高楼除大馬,海天愁思正茫茫,惊风乱起芙蓉水,密雨斜侵奇茄塘。岭树重遮干里目,江流曲松九回肠。共来百越文岳 地、犹自首爷那一乡。
<b>市職(2</b> 1999分) 5.170090086513516
长恨歌
汉皇重色思绘图,等字字本水不得,杨家有女切长成,养在深刻人未识,天生烈质难自弃,一朝选在君王剑。回眸一笑指 生,六宫阶层无颜色。春寒飘忽华渐危。温泉水香充湿。
登楼 杜川 [唐代] 花近高屋街房心,万方多度此景临。锦江青色未天地,玉金浮云变古今,北级朝廷终不改,西山琉盗莫相侵。可怜后主还称 唐,日晷朝为《梁卿吟》,《梁禺一作:梁父》 <sup>8</sup> 周军 <sup>9</sup> 888888888888888888888888888888888888
与诸子登岘山 / 与诸子登岘首
一角子立成は、一つ相子立成日
人事有代谢,往来成古今。江山留胜迹,我辈复登临。水落鱼梁浅,天寒梦泽深。羊公碑尚在,读罢泪沾襟。(尚在一作:
字在) ************************************
登岳阳楼
tim(唐代) 昔河洞庭水,今上岳阳楼。吴楚东南坼,乾坤日夜浮。亲朋无一字,老病有孤舟。戎马关山北,凭犴涕泗流。
語での時間をパープ、上し込むとはなって、そのとのではます。 そのようし ・フ・コンドラインをパープ・アインドのようには ・ 本質文格件: 3,055586702415735
乐游原 / 登乐游原
李商隐(唐代)
向晚意不适,驱车登古原。夕阳无限好,只是近黄昏。
專業文別提升: 2.286644284G4636
登高
杜甫 [唐代]
风色天鹅端壁。 港港沙白乌飞回。 无边路木荫萧下,不尽长江北浪来。 万里悲秋常作客,百年多响绘登台。 艰难苦恨繁霜 泉,诸姆研修迎进来。 #82758882:2886448886488
九日登望仙台呈刘明府容
能等 (新代) 江立白高玄喜春、東日縣物理色工 三葉二山蛇小石 一時日本白女女 然江入石北郷口 河上川於土工司 日郊に日本江
汉文皇帝有高台、此日谢临曜色开,三晋云山皆北向,二陵风雨自东来。关门今尹谁能识,河上仙寄去不回,且欲近寻丝洪 秦,陶然共静寄花杯。 <sup>集现了佛丽</sup> : 2.3884445664666
与高适薛堀同登慈恩寺浮園
等參 (副化)
塔勢如浦出,孤高耸天宫。發施出世界,侵道盘虚空。突兀压神州,峥嵘如鬼工。四角得白日,七层摩苍穹。下夜指高鸟。 俯听俯瞰风,连山岩波声,奔袭似脚东。(秦一作:走:似一作:如青槐突脸盖,宫馆的珍逸。(宿一作:观)秋色从西来, 苍然海美华,五郎北流上,万古清潔潭。净理了可信,起因刚乐宗。 攀林接冠宏,炎逝游光穷。

受鹳雀楼					
財素模型: boolean-model	o vector-model  prob-	model			
的提供选项: log	查阅document选项:	idf			
档价选项: log	· 文档document选项:		文档norm选项:	L2 Norm	
Hammer log	Xejdocumentajaja.	None	XHIROTHIZAGE:	L2 Norm	
分前10的文档为					
登鹳雀楼					
王之涣 [唐代]					
白日依山尽。 黄河入海	流。欲穷干里目,更上一层楼	s.			
本篇文档得分: 1.601211526768					
登岳阳楼					
杜甫〔唐代〕					
	楼。吴楚东南坼,乾坤日夜浮	全明于一字 老痒有	孤舟 戎马关山	と、佳軒逆源流	
本篇文档得分: 0.429273088111		- 201120 3, E/1313.	NACO NO DAME	or your promise	
中十井					
乌衣巷					
刘禹锡 (唐代)					
	巷口夕阳斜。旧时王谢堂前燕	<sup>1</sup> , 飞入寻常白姓家。			
本篇文档得分: 0.403793846785	9867				
赤壁					
杜牧 [唐代]					
折戟沉沙铁未销,自将	磨洗认前朝。东风不与周郎便	,铜雀春深锁二乔。			
本義文档得分: 0.403003551752	250013				
柳宗元 [唐代]	汀封连四州/登柳州 悉思正茫茫。惊风乱飐美蓉水			工流曲似九回肠。共来已	百越文具
	9间。长风几万里,吹度玉门关 作:望边色)高楼当此夜,叹 77587		海湾。由来征战	也,不见有人还。戌客望	型边邑,
乐游原/登乐游	盾				
亦 <b>が</b> 塚/豆亦附 李商篇 (唐代)	nas.				
	i原。夕阳无限好,只是近黄昏	ĝ.			
中成总个旦, <u>8</u> 4年豆C 本篇文档得分: 0.317023193425					
登幽州台歌					
陈子昂 (唐代)					
	者。念天地之悠悠,独怆然而	藩下!			
<b>削小火山へ</b> , <b>加小火水</b> 本蓋文档得分: 0.267760431043		ne (*)			
九月九日忆山东	兄弟				
王维 (唐代)	70/13				
	/ / / / / / / / / / / / / /	、 海坯花芯小 (			
<b>独住开乡//开各,母进</b> 本篇文档序分: 0.240830876321	佳节倍思亲。遥知兄弟登高处 ************************************	- / 西田本央シーへ。			
	Control of the Contro				



#### 长恨歌

白居易〔唐代〕

#### 登岳阳楼

昔闻洞庭水,今上岳阳楼。吴楚东南坼,乾坤日夜浮。亲朋无一字,老病有孤舟。戎马关山北,凭轩涕泗流。

#### 登楼

花近高楼伤客心,万方多难此蹬临。锦江春色来天地,玉垒浮云变古今。北极朝廷终不改,西山寇盗奠相侵。可怜后主还祠 庙、日暮聊为《梁甫吟》。《梁甫一作:梁父》

### 登柳州城楼寄漳汀封连四州/登柳州城楼寄漳汀封连四州刺史

城上高楼接大荒,海天愁思正茫茫。惊风乱趾芙蓉水,密雨斜侵薜荔墙。岭树重遮千里目,江流曲似九回肠。共来百越文身 地。犹自音书滞一乡。

#### 庐山谣寄卢侍御虚舟

李白 [唐代]

我本楚狂人,凤歌笑孔丘,手持绿玉杖,朝别黄鹤楼。五岳寻仙不辞远,一生好入名山游。庐山秀出南斗傍,屏风九叠云锦 张、影落明湖青端光。金阙前开二峰长,银河倒挂三石梁,香炉瀑布遮相望,回崖沓峋凌苍苍。翠影红霞映朝日,鸟飞不到 吴天长、鹭融壮观天地间,大江茫茫去不达。黄云万里动风色,白波九追流雪山,好为庐山湖,兴因庐山龙,周咏石镜清我 心,谢公行女苍苔没。早服还丹无世情,琴心三叠道初成。逼见仙人彩云里,手把芙蓉朝玉京,先期干浸九垓上,原接卢敖 游太清。

### 关山月

李白 (唐代)

明月出天山,苍茫云海间,长风几万里,吹度玉门关。汉下白登道,胡窥青海湾。由来征战地,不见有人还,戍客望边邑, 思归多苦颜。(望边邑一作:望边色)高楼当此夜,叹息未应闲。

### 芙蓉楼送辛渐

王昌龄 [唐代]

寒雨连江夜入吴,平明送客楚山孤。洛阳亲友如相问,一片冰心在玉壶。

#### 闺怨

王昌龄 [唐代]

闺中少妇不知愁,春日凝妆上翠楼。(不知一作:不曾)忽见陌头杨柳色,悔教夫婿觅封侯。

#### 黄鹤楼送孟浩然之广陵

故人西辞黄鹤楼,烟花三月下扬州。孤帆远影碧空尽,唯见长江天际流。(唯通:惟)



本次查询有意设计为一个精确查询,在四个查询中,排名第一的都正确搜索出来了我们要的正确的结果,但其中的 BIM 和布尔仍然存在偏向长文档的现象;向量和 MLE 则表现更好。

### 查询五: 渚青沙白鸟飞回

### 布尔模型

My IK System
清朝沙白鸟 (3)
服制限型 © boolean-model vector-model prob-model 認定的文字形 tf-idt © wf-idt
共西南图 50 条结果
登高 世郎 (唐代) 宋色天高的建筑,准遵约白乌飞湿。无边落木港萧下,不尽长江和北京。万里想秋常作客,百年多典绘璧台。观境苦憬繁霜 泉、游师陈等尘西怀。
獨這種  國子級說 器道之常,用于上青天!要从及自免,开握例定然 尔来四万八千岁,不与秦寨海人根。西当太白有乌 原子概。 信乎函說 器道之常,用于上青天!要从及自免,开握例定然 尔来四万八千岁,不与秦寨海人根。西当太白有乌 道。可以根据德國縣。她他从操任士外,然后又据在时期际底,上有大龙国门之梯后,下有冷型形分三则儿,舞道之"如怀 得没,我就这就这些事情,是何但他。但少少万余举色,仍是外中他的。以于其中是代众,但对此即时分了 股金城市 可谓,何见与乌马士术。官心战从操林局。汉周子四城时,然之山、秦龙之州,秦于上青天。任人年此郑丰后,这十余天 不是氏,林后就是他是一处是是一个战争。 特定的战人 使为这些人从另外于来说 的神经不是一 一大当天,万夫夷开,所守西寨来,化分除与时,都避延虎,少诸长蛇,能写死血,杀人如弃,修城是云圻,不如可还来, 据自之后,男子不见海寨来,化分除与时,都避延虎,少诸长蛇,能写死血,杀人如弃,修城是云圻,不如可还来, 那直之后,男子无以海军,以身无驾长,治
长信歌  ESSE [87]  以爱惠的经现。 由于多年来不得。杨家有女切长成。养在深观人太彩。天生用思用自身,一些热在甚至处。 图除一类四维 是一个大小的大型的一个大小型,一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个一个
庐山福寄产侍御虚舟  9日 (18代)  我本部任人,风吹荒礼丘、手持极王杖、朝别唐朝梯。五岳寻仙不恭远,一生好入名山游。庐山秀出海斗侍,屏风九盘云馀 宪、郑明明高献先,命昭对于"其代、明初胜王石泉、晋守师后居报是,回回帝师在岳龙、郑北田晚明日,与下至明 宋天代、曹尚林江河南,大江至华太东、黄汉为皇后的。 由发出成功能以上为为中国。兴园的人员,阳及田南政 人,此次为党者因。用最近共为无世情,举心三金命形成。温汉仙人多云生,手把英容领王弟,先期下者乃终上,即场冲放 游太海。
长干行-貸一  9日 (銀代)  20日 (銀行)  20日 (銀行)
参游天姥吟留别/参游天姥山别东鲁诸公 5日 (唐代) 5日
高歌行井序  AII (唐代)  开元十六年,各有从元连出商压去者,作《商歌行》以示,适感征注之事。因而机理,以家宪立在东北,汉特异家破败 居、男儿本自国研究,天子市常愿险。民产战功率的来,但该这些有关。因而机理,以家宪立在东北,汉特异家破败 战处土,胡伊克德州山、龙江南市平原之,成本下北极为,大学则从秦阳湖、南北省北十年,是,是秦阳城的山、山川南条 成处土,胡伊龙德州山、北大省北大学。成本下北极为一大学的大型场,由于张大州南加、山川南条 成立土,胡州市省、大学、大学、大学、大学、大学、大学、大学、大学、大学、大学、大学、大学、大学、
夜上受降城间笛  \$1 (8代)  原环境部分似意,受难纳外月知識。(厄乐峰一作: 厄乐峰) 不知何处次声管,一夜征人尽望乡。  *国尔姆斯》 24448416705444
白雪歌迷武判官归京  78 (唐代)  184 (唐代)  北风相他白母听,胡天八月即飞雪。忽如一改春风泉,干祝万树梨花开。撒入珠帘遮罗幕,绿菜不喝晚余篷。将军角号不得 龙。都呼收衣切难息,便看一十,张澄婉胸等于百丈夫。赵元传次万里战,中中置近代马虎,即将战处为来血,纷纷得宣 下榻了,风景似现在不知。轮台街了这巷去,去对雪滩天山路。山园窗桥不见老,里上空留马行处。
古柏行 村用(銀代) 利用組成化之。例如海狗學於方、高於國和四十世,當命参末二千尺、君配已時期会。明末坊为人聖信、云來气吸成核 於、丹北鄉海出出。《公司教授學學师、先主郑原而郎、崔朝於千成郡方、即四月海宁鄉之、海洛南國衛衛地。冥祝福 多利风、扶持是最等的力、正直即迎接化工、大國如孫要從称。万年回台丘山德,不寬文學世已地,未非劉化維修定?並ら 即公翰後後,西州安在衛星、北上諸人及即徙。古末林大淮为用。 ·周丁四衛士、2.64005/08114154

青沙白乌飞网
题:   boolean-model   ② vector-model   ○ prob-model
態項: log v 煮illydocument透頂: idf v
(態項: log V 文档document透頭: None V 文档norm透頭: L2 Norm V
前10的文档为
登高
間 (唐代)
Q急天高猿啸哀,渚清沙白鸟飞回。无边落木萧萧下,不尽长江滚滚来。万里悲秋常作客,百年多病独登台。艰难苦恨繁 ************************************
篇文献明分: 1.0742210908374978
<b>友泊牛渚怀古</b>
自唐代
TREEDICK   ドラング (1 Ac 32.7 生 1/2 7)   上 にのがり ++  かか Roberで   オレマド・リッツ ** *** *****************************
寄左省杜拾遗
徐 (唐代)
关步趋丹陛,分曹限紫微。晓随天仗入,暮惹御香归。白发悲花落,青云羡鸟飞。圣朝无阙事,自觉谏书稀。
萬文樹得分: 0.5698522083155136
<b>划王十一南游</b>
  长脚  (唐代)
思君烟水阔,挥手泪沾巾。飞鸟没何处,青山空向人。长江一帆远,落日五湖春。谁见汀洲上,相思愁白蘋。
· 真文梅博分: 0.5665865887719286
瓜雁二首•其二
(唐代)
1行归塞尽,念尔独何之。暮雨相呼失,寒塘欲下迟。渚云低暗度,关月冷相随。未必逢矰缴,孤飞自可疑。
篇文档博分: 0.48782997798903854
· =
<b>工雪</b> 除元 (唐代)
F山岛飞绝,万径人踪灭。孤舟蓑笠翁,独钓寒江雪。 第28時分: 0.4818135377717353
開入(明度27: V-0-10133377 / 17333
5上受降城闻笛
益(唐代)
即乐烽前沙似雪,受降城外月如霜。(回乐烽一作:回乐峰)不知何处吹芦管,一夜征人尽望乡。
黨文樹門分: 0.48130597076642667
京州词
(唐代)
前荀美酒夜光杯,欲饮琵琶马上催。醉卧沙场君莫笑,古来征战几人回?
識文態時分: 0.46684587423663576
×
<b>六干行•其一</b>
的 (唐代) 24分719865、15757分别。6月254日75、6345平底45、同居区工用、东小工6458。 上四头更为 2545大学工 (区) 5月28月
发发初覆额,折花门前剧。郎骑竹马来,绕床弄青梅。同居长干里,两小无嫌猜。十四为君妇,羞颜未尝开。低头向暗留 F唤不一回。十五始展眉,愿同尘与灰。常存抱柱信,岂上望夫台。十六君远行,瞿塘滟滪堆。五月不可触,猿声天上穹
康声一作:猿鸣)门前迟行迹,——生绿苔。苔深不能扫,落叶秋风早。八月蝴蝶来,双飞西园草。(蝴蝶来一作:蝴蝶黄 似比伤妾心,坐愁红颜老。早晚下三巴,预将书报家。相迎不道远,直至长风沙。
第2点時分: 0.44586790005596
<b>音建德江</b>
浩然(唐代)
8舟泊烟渚,日暮客愁新。野旷天低树,江清月近人。
第文相限分: 0.4418139516599108

#### BIM

#### My IR System

odel vector-model prob-mod 授家方式选择: O BIM

#### 长恨歌

#### 庐山谣寄卢侍御虚舟

我本桩主人,风吹笑礼丘,手持续玉杖,朝奶黄鹤楼。五岳寻仙不辞远,一生好人名山鸽,庐山秀出海斗得。屏风九最云鸲 殊。 郑端明绮横梁九 金婉朗广上峰长。银穷相往三石梁,看好准布造相望。四盛宫峻龙苍苍。 瑟秋且明晓时,今下不到 吴大长,登翰杜汉为园。大江庄老木水,黄云刀里边风。 比这九金流场出,好为山庙。 兴起中战,积级石场清极 心,被公乃处击谷汉,早报公丹光世境,等心三般诸初成。遥见仙人彩云里,手尼芙蓉新玉泉,先称千夜九块上,原接卢故 游太高。

#### 蜀道难

· 哪好戲,是乎歲稅!獨直之推,是于上青天!蚕丛及鱼兔,开留何茫然!尔来四万八千岁,不与秦寒遇人危。四当太白有鸟 道。可以操他细菌癌。 地面山脉壮士死。然后天林花日桂阳原。 古书大龙四日之称宗。 不肯冲改进开之回川,黄鹤之飞和不 何况, 孫就從改發稅處,青河的但是。 目步山苏美密店,到防牙种印息。 以手并赐坐长死。 问官都是附归讨记? 则是她就在 可事。 组双路与号古木,旗:如此从除林园,又闻子郑明农月,愁空山,蜀湖之淮,淮于上青天, 侵入所迁成来原〕 连母朱承 不是尺、 任然后期待他是、 飞瀑率投入回路。 英雄华日之富莽,到世也如此,《安心宣道之人为万子来说》:如明帝祖朱来 一夫当朱、万夫畏开,所守庭田宗,化为指导对。 明朝任成,

#### 长干行•其一

复发功重额,折说了前期,邮材归未,绕朱再青槐,同居长千里,两小无塘塘,十四为君妇,羞颜未尝开。低头向暗壁,干唤不一回,十五绘展。原阳尘与灰,常穷胜柱信。岂上望夫仓,十六君运行,理想能观难。五月不可绝,猿声天上哀。(唐声作:"瑜问·斯说行法。——生慈善善游不能思,避时秋风早,几时缓来,双飞西周草。(蝴蝶来一作:蝴蝶黄)绿此伤寒心,坐愁红颜老,旱晚下三巴,烧将书极家。相迎不通远,直至长风沙。

#### 听董大弹胡笳声兼寄语弄房给事

蔡女恰当郑范声,一排一十有八拍、胡人渐指出幼年,汉法斯部对刘格、古成苍苍峰火寒,大观况瓦飞驾白,先排南场后角 羽,四郊外村场城底。董夫子,通申明,深山即亦及林伟。当近原吏皆由于,将北复加政府体,空山自马郡还合,万原于云 阴国岛,秦帝被周末存取,新始别之帝与,川为海县发。《亦文景阳》。《孙宗朝》。《孙宗家圣》。爰之之宫故往。德令之之宫故。 唐、长风欢林相望远。进泉武汉飞木末,野德叫哈走堂下,长安城至东城归,风夷也对南湖门,高才即居名与利,日夕愿君 批等还。

#### 梦游天姥吟留别 / 梦游天姥山别东鲁诸公

等高铁温州,郑本观在他郑末,越人进天地、云南铜灭或引驰,天地连天向天果,势拔五压掩卧城。天台四万八千丈,对此 较明许南帆。我欲因之梦吴越,一夜飞度惊滅月,湖月阴极影。这我至彻溪。谢公帝处今尚在,源水落蒸南旋啼。同著帝公 原,是惟曾云城。年驱见即日,空中顺天地,干起为城路不远。这代自己即已观,顺地走沙接世郎,原深林兮彻虚赋。云齿 青兮似而。水高速与生地,则延霜形。正齿脂糕,源天石丽。临郊中井,南京孟东不见庙。日归南盛金岭。第六本兮如唐庙。云 与。云之君兮纷纷而来下。虎越悲兮喧回年,仙之人兮沙顺麻。宠魂野以境动,忧恨起而长城,惟觉即之北薄,失鸣龙之烟 唐、世间行所为如此。古来万事东流水,别君去兮何时还?且故白鹿青崖间,疾行前将访名山,安能据避肝腰事収贵,使我 不得开心颜?

#### 寄左省杜拾遗

联步趋丹财、分曹限紫微、磅礴天仗人、暮若御香归。白发悲花落、青云羡鸟飞、圣朝无烟事、自觉读书稀。

#### 饯别王十一南游

望君烟水阔,挥手泪沾巾。飞鸟没何处,青山空向人。长江一帆远,落日五湖春。谁见汀洲上,相思愁白蒴。

丽人行

三月三日天气新,长安水边多部人。态浓卷远混旦真,机理组联而均匀。携罗衣要圆幕者,爱金孔准银麒麟。头上何所有? 翠莺等叶植蚁居,曾后何所见了连压腰极脉停身,统叶云雕树原宗,随名大国称诗奉,繁饰之碑坛景参,水笔之独行素核。 摩灌历代父太下,鬼刀埃切空纷纶。黄门了格不动走,细断络绿送儿诊。萧鼓哀吟号鬼神,宾从必重实更津,后来鞍马何逡 逝。当叶下马入鸠园。杨花浩渐围里中,南与飞去拖江市、炙手可热劳悟危,值是近前途距离1

#### 官州谢朓楼饯别校书叔云

弃我去者,昨日之日不可留;乱脱心者,今日之日多烦忧,长风万里送秋雁,对此可以胜高楼。蓬莱文章建安骨,中间小密 又清发。俱怀逸兴壮思飞,欲上青天振明月,抽刀断水水更流,举杯酷愁愁更愁,人生在世不称意,明朝散发弄扁舟。



这次搜索的目的是模拟输错一个字符的情况,当我们把清输入为青的时候,大部分模型仍能搜索到正确结果,但只有 BIM 不可以,他更加偏向了长文档。其他结果同样,向量模型和 MLE 模型类似。

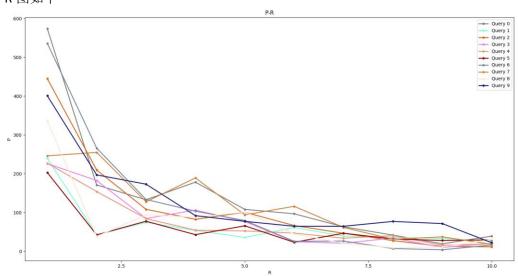
### 三、项目检索性能评价

选用概率模型中的 MLE 方法进行检索。

进行的十条查询为:

- 0- 春花风月夜
- 1- 江河湖海
- 2- 江南好风景
- 3- 月明 相思
- 4- 面白如玉
- 5- 东南西北
- 6- 田屋洒月夜
- 7- 春眠不觉晓
- 8- 镜花水月
- 9- 离人心秋愁

### P-R 图如下



将十次结果取平均后的 P-R 图如下

