



Beijing University of Posts and Telecommunications

# 本科生课程设计

## 作业 2：信息检索实验系统

课 程 名 称： 计算机系统结构

姓 名（学号）： 唐义泓(2018211328)

姓 名（学号）： 曹世东(2018211330)

姓 名（学号）： 刘彦博(2018211327)

学 院： 计算机学院

专 业： 计算机科学与技术

指 导 教 师： 刘传昌副教授

二〇二一年 五月

# 信息检索实验系统

## 一、综述

我们按照实验要求，通过 python 编程语言，实现了**倒排索引**文档存储，和**向量空间模型算法**来完成检索。并对实验结果进行了多次检验，发现准确率满足实验要求，并对**实验结果进行人工测评，满足并集条件**。在扩展方面，一方面我们实现了**多媒体信息检索**，根据语音内容检索相应的语音文件，另一方面，我们积极探索**信息检索算法**，对向量空间模型算法进行扩展，扩展了包括 BM25 算法在内的四种检索算法。

## 二、模型的环境与可持续发展意义

我们设计的模型，在生态与现实意义上，我们具有一定的实际意义。如我们的数据来源是北邮信息门户，现在的信息门户上的搜索算法为字符串匹配算法，有时并不能满足用户的要求，我们给出了多种算法来方便大家更方便地搜到自己的想要的学校信息。另外我们的模型有了语音检索功能，能够更好地为弱势群体服务。我们的模型运行在网页端口上，相对于本地搜索使得计算集中，可以节能环保，搜索内容积极向上，弘扬社会主义核心价值观。

## 三、系统设计：

这部分包括整体架构说明，和爬虫，倒排索引实现，向量空间检索模型实现，扩展内容（扩展算法及多媒体检索），网页。

### 3.1 工具列表

本次课程实验使用 python 编写，使用了以下库进行开发：

Argparse jieba math operator sqlite3 configparser sys json

## 3.2 系统运行

### 爬虫:

在 PC 文件夹下,若要使用爬虫功能,需要先修改 `settings.json` 文件,在其中的 `cookie` 字段内填入自己登陆信息门户后获取的 `cookie`, `cookie` 可以在 `chrome` 浏览器“检查”功能中“`Network`”部分抓取的文件的包头得到,具体参考百度或教程 <https://blog.csdn.net/code4101/article/details/90208429>。设置好 `cookie` 后,确定环境没有问题之后,执行命令 `python main.py --total=[] --startId=[]` 便可以进行信息爬取,其中 `total` 是要爬取的文章总数;`startId` 是需要爬取的最新的新闻 ID,解释参考文档中爬虫部分,默认为 90297,这对应了 2021.5 的一篇文章。

### 文字信息检索:

在 `Search` 文件夹下,运行目录下的 `makeindex.py` 文件(注:需要提前将数据文本放入 `data` 文件夹下,命名格式为数字+.txt,数据集和索引文件已经提前放入了相应位置了。同时需要保证 `data` 文件夹下不存在 `ir.db` 文件,如之前生成过需要删除),这个程序会生成倒排索引及统计各个词与文档的相关信息存入数据库中,以便之后的检索只与数据库进行交互。

在执行 `makeindex.py` 之后,我们便可以进行搜索文本数据搜索,命名为 `python search.py --o=[] --input=[] --method=[]`,我们可以指定多个运行参数,比 `--o` 指定输出文件名 `--input` 指定输入字符串, `--method` 指定相关算法,有 `VSM`, `preVSM`, `Bool`, `BM25` 四种,推荐使用 `BM25`。输出文件为一个 `json`,是一个 `list`,包含了搜索结果的文件名(不含.txt)以及评价指标数据。在屏幕上会打印统计的召回率,作为检验准确度的指标,具体说明请见第七部分-人工评价。

### 多媒体信息检索:

在 `Search` 文件夹下,执行 `python VoiceSearch.py --source=[] --target=[] --method=[]` 可以执行语音检索,其中 `source` 是源语音地址,只能是 `wav` 格式, `target` 是检索输出,表现为文档 ID 和评价指标构成的 `json`, `method` 为检索方法,有四种方法分别为 `VSM`, `preVSM`, `Bool`, `BM25`,推荐使用 `BM25`。

网页：

在配置好的环境下，执行 `python backend.py` 即可打开服务器后端，网页输入 `localhost:8000` 即可进入网页版系统。目录下其中 Web 文件夹包含了网页系统的静态资源。

## 四、爬虫：

### 4.1 数据源选取

经过讨论，爬取的数据源为北京邮电大学信息门户校内新闻版块。每个新闻数据源网页如下图所示。



### 4.2 数据爬取方法

数据爬取采用 `urllib` 库中的 `request` 包。本组实现方法如爬虫文件中的 `getFileFromInternet` 函数。

经分析，信息门户信息版块的每一条信息的地址格式为 '`{前缀}+{新闻 ID}`'，其中 `{前缀}` 部分为：

`'http://my.bupt.edu.cn/xnxw_content.jsp?urltype=news.NewsContentUrl&wbtreeid=1269&wbnewsid='`

{新闻 ID} 部分为一个五位数，如上一部分中新闻的 URL 为：

`my.bupt.edu.cn/xnxw_content.jsp?urltype=News.NewsContentUrl&wbtreeid=1246&wbnewsid=90900`

可以看到其 {新闻 ID} 为 90900。因此，我们只需要对 {新闻 ID} 进行从大到小的遍历，便可以获取从当前到过去一段时间内的新闻 URL。在预先保存登录后的 cookie 后，通过遍历一定数量的 URL，每次通过 request 包抓取渲染好的网页 DOM，便可以获取到包含网页信息的 HTML，之后对网页数据进行清洗，提取需要的内容。

除此之外，经过进一步的探索，我们发现每一个有效的 {新闻 ID} 并不是连续的，如我们将 {新闻 ID} 从上面的 90900 改为 90899。网页如下：



可以看到这其实是一个空网页，那么在后续的处理过程中，我们的策略是对空网页进行识别和去除，不计入我们需要获取的网页数目中。

### 4.3 数据清洗方法

获取到的 HTML 数据通过 bs4 库中的 BeautifulSoup 包进行数据清洗。Beautiful Soup 提供一些简单的、python 式的函数用来处理导航、搜索、修改分析树等功能。它是一个工具箱，通过解析文档为用户提供需要抓取的数据。

该部分对应本组实现代码中的 washHtml 函数。

经过预先的分析，我们只提取新闻中的文本信息，一个新闻 HTML 数据中主要的新闻文本内容部分在标签 p 和 h 里面，新闻的大标题在标签 title 里面。明确了这一点后，我们通过 BeautifulSoup 解析网页的 DOM 结构，定向获取其中的 title, h, p 标签，组合到一起，实现新闻网页内容的清洗，清洗结果如下图所示：

标题:关于公布2021年北京邮电大学本科教材建设项目立项结果的通知 -欢迎访问信息服务门户  
关于公布2021年北京邮电大学本科教材建设项目立项结果的通知 北京邮电大学教务处  
教务[2021]82号

关于公布2021年北京邮电大学  
本科教材建设项目立项结果的通知  
各学院:

根据《关于开展北京邮电大学2021年教材建设项目立项工作的通知》(教务[2021]61号)的总体安排,教务处组织开展了相关评审工作。经各学院择优推荐,共有39本教材参与申报。经学校组织专家组评审,《智能信息网络》等30本教材获批2021年我校本科教材建设项目,具体名单如下:

各项目负责人请认真开展研究并按期完成项目,依据学科专业或课程教学标准编写教材,服务我校教育教学改革和人才培养。请各学院加强对立项项目的监督管理。

教务处  
2021年5月25日

相关新闻最新新闻

热门新闻

全部最新新闻

版权所有©北京邮电大学,联系我们: portal@bupt.edu.cn

## 4.4 数据保存

由数据检索和抽取部分的功能需求,本模块进行爬取和清洗的结果以{数字}.txt 的形式保存,其中{数字}部分从1开始,向上增长。保存在爬虫模块文件夹下的 output 文件夹中。

## 五、 倒排索引实现

在倒排索引的实现主要在 makeindex.py 这个模块中,我们设计了一个 Doc 类来存储某个词在某个文档中的文档 id, tf (频率), 文档长度以及 tf-idf, 以方便为们后面的检索信息计算。然后我们存入数据库的每一条数据都是一个词对应一个数 n (代表出现该词的文档个数) 和对应的 n 个 Doc 类。具体计算过程如下:

我们对文档进行分词,并删去停用词,我们进行两遍遍历,第一遍遍历我得到每个词的逆文档频率并暂时用变量存储。然后我们进行第二遍遍历,每当在某个文章中得到一个词,我们就计算这个词的在这个文档的频率,进而计算出这个词的 tf-idf。然后我们把(文档 id, 该词词频, 文档长度, tf-idf) 压入该词的 list 字典中暂时存储。

最终我们遍历词的字典,将每个词的 list 中的每条记录都写入数据库。

## 六、空间向量检索模型 VSM 的实现

在这里我们实现了两种空间向量检索模型 VSM。

老师的 ppt 里讲到，目前 VSM 中计算两个向量距离的方法有很多还不确定，于是我对两种主要的方法为向量内积距离和向量夹角距离进行了测试，效果相差不多，于是在此选择了向量内积距离进行测试：

我们实现的两种 VSM 算法如下。

### 6.1 简单 VSM:

简单 VSM 是指当向量的每一位都是 1 或者 0，就是只检索这个词有没有在这个文章中出现，而不会根据词的重要性进行加权。

这里我们就将输入的检索的字符串 str 进行分词，然后对每个词来说，我们去查询 db，然后给 db 中该词的每一个检索到的文档 id 得分加一。对每个检索词

### 6.2 tf-idf 修正的 VSM:

由于简单 VSM 并不能体现出每个检索词的重要程度，所以我采用了 tf-idf 修正的 VSM 检索模型。

为了根据每个词的重要程度进行加权，我们将向量中的每个词对应的那个数改为该词在该文档中的 tf-idf 值，同样为了标识检索词的重要程度，我们把检索向量的每个词的值由 1 改为 tf-idf 值，但在检索向量中逆文档频率为 1，故为 tf 值。

这样的 VSM 考虑到了每个词的重要程度，于是可以取得更好的效果。

**Tf-idf 简介:**

**词频 (TF)** 表示词条 (关键字) 在文本中出现的频率。

这个数字通常会被归一化 (一般是词频除以文章总词数)，以防止它偏向长的文件。

$$TF_w = \frac{\text{在某一类词条 } w \text{ 出现的次数}}{\text{该类中所有的词条数目}}$$

**IDF 是逆向文件频率 (Inverse Document Frequency)**

某一特定词语的 IDF，可以由总文件数目除以包含该词语的文件的数目，再将得到的商取对数得到。

如果包含词条 t 的文档越少，IDF 越大，则说明词条有很好的类别区分能力

$$IDF = \log \left( \frac{\text{语料库的文档总数}}{\text{包含词条 } w \text{ 的文档数} + 1} \right)$$

TF-IDF 实际上是:  $TF * IDF$

某一特定文件内的高词语频率, 以及该词语在整个文件集合中的低文件频率, 可以产生出高权重的 TF-IDF。因此, TF-IDF 倾向于过滤掉常见的词语, 保留重要的词语。

## 七、扩展内容——更多的检索方法

首先我们实现了更多的检索算法如 BM25 算法和布尔检索算法:

### 7.1 BM25 算法:

BM25 算法, 通常用来作搜索相关性平分。一句话概况其主要思想: 对 Query 进行语素解析, 生成语素  $q_i$ ; 然后, 对于每个搜索结果  $D$ , 计算每个语素  $q_i$  与  $D$  的相关性得分, 最后, 将  $q_i$  相对于  $D$  的相关性得分进行加权求和, 从而得到 Query 与  $D$  的相关性得分。

$$Score(Q,d) = \sum_1^n W_i \cdot R(q_i, d)$$

其中,  $Q$  表示 Query,  $q_i$  表示  $Q$  解析之后的一个语素 (对中文而言, 我们可以把对 Query 的分词作为语素分析, 每个词看成语素  $q_i$ 。);  $d$  表示一个搜索结果文档;  $W_i$  表示语素  $q_i$  的权重;  $R(q_i, d)$  表示语素  $q_i$  与文档  $d$  的相关性得分。

下面我们来看如何定义  $W_i$ 。判断一个词与一个文档的相关性的权重, 方法有多种, 较常用的是 IDF。

我们再来看语素  $q_i$  与文档  $d$  的相关性得分  $R(q_i, d)$ 。首先来看 BM25 中相关性得分的一般形式:



$$R(q_i, d) = \frac{f_i \cdot (k_1 + 1)}{f_i + K} \cdot \frac{qf_i \cdot (k_2 + 1)}{qf_i + k_2}$$

$$K = k_1 \cdot (1 - b + b \cdot \frac{dl}{avgdl})$$

其中， $k_1$ ,  $k_2$ ,  $b$  为调节因子，通常根据经验设置，一般  $k_1=2$ ,  $b=0.75$ ； $f_i$  为  $q_i$  在  $d$  中的出现频率， $qf_i$  为  $q_i$  在 Query 中的出现频率。 $dl$  为文档  $d$  的长度， $avgdl$  为所有文档的平均长度。由于绝大部分情况下， $q_i$  在 Query 中只会出现一次，即  $qf_i=1$ ，因此公式可以简化为：

$$R(q_i, d) = \frac{f_i \cdot (k_1 + 1)}{f_i + K}$$

从  $K$  的定义中可以看到，参数  $b$  的作用是调整文档长度对相关性影响的大小。 $b$  越大，文档长度的对相关性得分的影响越大，反之越小。而文档的相对长度越长， $K$  值将越大，则相关性得分会越小。这可以理解为，当文档较长时，包含  $q_i$  的机会越大，因此，同等  $f_i$  的情况下，长文档与  $q_i$  的相关性应该比短文档与  $q_i$  的相关性弱。

综上，BM25 算法的相关性得分公式可总结为：

$$Score(Q, d) = \sum_i^n IDF(q_i) \cdot \frac{f_i \cdot (k_1 + 1)}{f_i + k_i \cdot (1 - b + b \cdot \frac{dl}{avgdl})}$$

在本文中，我们采用的是如下的常数形式。

$$\frac{(ri + 0.5)/(R - ri + 0.5)}{(ni - ri + 0.5)/(N - ni - R + ri + 0.5)} \cdot \frac{(k_1 + 1)fi}{K + fi} \cdot \frac{(k_2 + 1)qfi}{k_2 + qfi}$$

## 7.2 Bool 算法：

布尔检索是最基础，也是使用最广泛的信息检索模型了。所谓布尔查询就是

通过 AND、OR、NOT 等逻辑操作符将检索词连接起来的查询。

当我进行布尔查询的时候，就是要求查询的词必须要在相应的文档中出现，然后对每个词查出来的文档集合去交集，或者根据某些关键词如 AND、OR 或者 NOT 进行集合运算操作。

NOT 操作就是先将 NOT 之后的内容取反，再进行 AND 操作。

我们可以发现使用词汇文档矩阵的话，进行布尔检索十分简单。

## 八、多媒体扩展内容——语音检索模型：

多媒体信息检索基于实现的基本文本信息检索系统，通过对语言信息进行文字转换，之后以转换后的文本信息，与语言数据库中对应的文本数据应用文本信息检索系统，将匹配的语音相关信息返回，在命令行模式下返回对应的语音信息 ID，在网页的模式下显示对应的文本信息和语音数据链接。使用方法见报告前部分与 README.md。语音文件仅支持 WAV 格式。

（注：语音信息的文字转换采用 google 开发的在线 api，其他的免费工具效果很不理想。因此需要在可以连接境外网络的环境下执行程序，否则会出现网络请求错误。本工具由谷歌团队进行维护和优化）。

该系统将语音信息的匹配转化到文字的匹配上，相比于纯粹的语言匹配，准确率有所提高。同时由文字抽取系统识别出的内容具有一定的语义信息和纠错机制，对结果有一定的优化效果。

选取文件

testby.wav

VSM

preVSM

Bool

BM25

搜索一下

## 八、人工评价方法：

经过一次匹配之后，我们将相关的文档输出，用户可以点击查看相关文档的全部信息，可以感受到我们的检索算法的优劣之处。

在逻辑上讲，我们对自己的检索系统进行了多次尝试，发现我们的系统检索结果（除 bool 这种强匹配算法外）都是满足并集条件的。

在这里我们将并集条件作为我们的人工评价方法之一，即如果我们对 AB 进行检索，我们的输出 A 的检索结果和 B 的检索结果的并集。在这个指标上，除了 Bool 算法的原则要求是必须强匹配外，我们其他的算法指标都是 100%

在满足并集条件，和分词算法很低出错的概率上，我们便做到了对召回率的 100% 的成功，对于准确率，可以通过观察相关的文章内容来进一步又用户人工评定。

为了检验该准确率。我们自定义了一种方法：我们假设我们的四种实现方法所查到的文档的并集接近真实情况下所有相关的文档集合，在这个条件下，我们将本次搜索到的文档集合与四种搜索方法得到的文档并集（下称：总集）进行交运算，然后将交集的元素个数与总集的元素个数进行比值，该值作为本次搜索结果的准确率。这反应了我们的搜索结果是否接近实际要搜索的文档集合。

另外，在这种情况下，由于 bool 算法需要完全匹配的特殊性，bool 算法在本组方法中的检验准确率没有太大意义。

运行效果如下：（关键词 世界一流大学）

```
In [19]: runfile('D:/BUPT/信息与知识获取/ALL/提交/实验二/程序/Search/search.py',
wdir='D:/BUPT/信息与知识获取/ALL/提交/实验二/程序/Search')
检索中：
['世界', '一流大学']
['世界', '一流大学']
['世界', '一流大学']

统计中：
Recall : 100.0 %

In [20]: |
```

# 九、运行实例

倒排索引数据库截图：

SQLiteSpy - C:\Users\914200081qq.com\Desktop\info\data\ir.db

File Edit View Execute Options Help

Name

Type

main

Tables (1)

postings

Collations (7)

postings

term	df	docs
四位	4	112230.00821035560334728
苏联	4	112230.00821035560334728
接受	21	112230.004980941686728639
特训	3	112230.00877061899620961
特工	3	112230.00877061899620961
小队	3	112230.00877061899620961
叛徒	2	112230.009560265897804595
出卖	2	112230.009560265897804595
面对	9	122230.013262123600314627
处境	1	112230.01091017619226191
非人	1	112230.01091017619226191
磨难	1	112230.01091017619226191
考验	3	122230.01754123799241922
始终	11	112230.00624025383729229
坚定	12	112230.006070798407294982
黎明	2	122230.01912053179560919
向往	1	112230.01091017619226191
共产主义	3	112230.00877061899620961
信仰	7	112230.007120498882780936
严峻	1	112230.01091017619226191
敌人	2	112230.009560265897804595
苦苦	1	112230.01091017619226191
周旋	1	112230.01091017619226191
舍生忘死	2	112230.009560265897804595
最终	18	112230.005281151505699998
体会	9	112230.006631061800157313
现今	4	112230.00821035560334728
幸福	10	112230.006425871259526483
安定	1	112230.01091017619226191
生活	35	112230.003986104244502825
来之不易	2	112230.009560265897804595
无赖	3	122230.01754123799241922

1

2

223

0.01912053179560919

25

2

384

0.011103850495887628

我们点击某条记录：

光辉	5	112230.007775781553983798	
1	1	223	0.007775781553983798
132	1	443	0.003914219608438796
138	1	155	0.011187092171215399
20	2	241	0.01439003557293267
30	1	371	0.004673852524362229

我们可以看到光辉对应 5 代表这个词一共在 5 个文档中出现过，然后后面是五条记录，如下面所展示，代表每个出现的文档的 id，tf，文档长度，tf-idf。

其他的词也是如此：

党员		17	132230.016177404073938993
1	3	223	0.016177404073938993
114	1	573	0.0020986393883004047
115	2	190	0.01265810915259086
130	2	209	0.011507371956900781
132	14	443	0.03800290106759784
136	4	265	0.01815125086031897
138	6	155	0.04654917559339865
16	6	244	0.029570173020396683
17	4	439	0.010956905416821245
25	1	384	0.0031315634622295094
26	2	435	0.005528829285039686
33	2	412	0.005837477521825882
40	1	250	0.004810081477984527
50	8	493	0.01951351512366948
53	8	620	0.015516391864466217
64	3	301	0.011985252852120914
75	1	73	0.01647288177391961

搜索运行：（以网页版为例）

关键词：世界一流大学 方法：BM25

世界一流大学

VSMpreVSMBoolBM25

搜索一下

展开 / 收起

标题:歌声献给党!《唱支山歌给党听》快闪活动成功举办-欢迎访问信息服务门户  
歌声献给党!《唱支山歌给.....

评价指标:5.093451717221701

展开 / 收起

标题:校党委书记吴建伟与新一届校学生会、校研究生会主席团成员集体谈话-欢迎访问信息服务门户  
校党委书记.....

评价指标:4.945736982861305

关键词：数学 方法：VSM

数学

VSM

preVSM

Bool

BM25

搜索一下

标题:网络空间安全学院讲座-欢迎访问信息服务门户  
网络空间安全学院讲座 应网络空间安全学院王励成教授邀请,中国科学院信息工程所许军副研究员和史丹萍副研究员将于 5 月 27 日(周四)来访北京邮电大学,并做报告。欢迎感兴趣的师生踊跃参加!

报告一:求解多项式方程小整根的格方法

主讲人:许军 副研究员

主持人:王励成 教授

报告时间:2021-5-27 13:00-13:45

地点:教三楼 233教室

主讲人介绍:

许军,中科院信息工程所副研究员,美国数学学会的《数学评论》(Mathematical Reviews)的评论员。许老师长期从事公钥密码分析,特别是利用格基约化的思想对RSA密码体制、BNP问题等做出了多项重要工作,发表在 CRYPTO、Eurocrypt、Designs, Codes and Cryptography等密码学著名会议和期刊上。担任密码学旗舰会议Asiacrypt 2020、PKC 2019及重要期刊Theoretical Computer Science的审稿人。近期的的代表性工作:打破了著名密码学家美国工程院院士Dan Boneh等提出的近20年的关于模逆隐藏数困难问题的公开猜想;此外,还顺带攻击了一类伪随机数生成器-逆同余生成器,解决了种子密钥恢复问题。该结果发表于密码学国际顶级会议-CRYPTO 2019,并被评为中国科学院信息工程研究所2019年度所级最佳论文。

内容摘要:

格方法是一类求解多项式方程小整根问题的方法。本报告将介绍多项式方程小整根问题在公钥密码学中的重要作用,给出格方法求解该问题的主要策略,并呈现在RSA密码和相关数学困难问题中的主要分析结果。

报告二: Automatic Search of Meet-in-the-Middle Preimage Attacks on AES-like Hashing

主讲人:史丹萍副研究员

主持人:毕经国 副教授

报告时间:2021-5-27 13:50-14:35

地点:教三楼 233教室

主讲人介绍:

史丹萍,中国科学院信息工程研究所副研究员,入选中国密码学会第六届“青年托举”项目,密码学会第三届青年工作委员会委员。长期从事对称密码的设计与自动化分析,研究成果发表在 CRYPTO、ASIACRYPT, FSE 等著名密码学会议上。

内容摘要: The Meet-in-the-Middle (MITM) preimage attack is highly effective in breaking the preimage resistance of many hash functions, including but not limited to the full MD5,

展开 / 收起

如下所示为主要匹配内容:

102

标题:赵纪宁副书记会见原卫生部外事司司长宋允孚先生-欢迎访问信息服务门户

赵纪宁副书记会见原卫生部外事司司长宋允孚先生 5月21日下午,我校党委副书记赵纪宁于行政楼101室会见原卫生部外事司司长、世界卫生组织前伙伴关系与联合国改革协调员宋允孚先生。双方就推动学校高层次国际化人才培养,鼓励和支持北邮学子赴国际组织实习任职进行深入交流。

赵纪宁副书记对宋允孚先生的到来表示欢迎,并感谢宋先生受邀为我校师生做“培养家国情怀,逐梦国际组织”专题讲座。赵书记表示,学校高度重视国际组织人才培养与输送,已于今年3月获批教育部中外人文交流中心高层次国际化人才培养创新实践基地,并于近日与国际电信联盟等专业性国际组织达成合作意向。未来,学校将进一步鼓励和支持学生赴国际组织实习任职,参与全球治理,在国际舞台上展现北邮风采。

宋允孚先生感谢赵书记的亲切接见。他表示今年是中国恢复联合国合法席位50周年,但目前中国籍职员在联合国等国际组织中占比与中国的国际地位严重不匹配。要提高中国在全球治理体系的话语权,需要更多青年人才加入国际组织,发出中国声音。北邮作为享誉海内外的“信息黄埔”,应充分发挥学科优势,培养有工科专业背景和经管、法律、语言等复合能力,同时具备家国情怀、国际视野和时代担当的高层次国际化人才。宋允孚先生向赵纪宁副书记赠送自己所著的《国际公务员与国际组织任职》和《国际公务员素质建设与求职指南》,期待更多北邮有志青年在学校支持下奋发努力,将个人理想抱负与{\*\*\*世界\*\*\*}和平安发展相结合,为全球治理贡献中国力量。

教育部中外人文交流中心代表孙中雯女士、侯钢先生陪同来访,我校国际合作与交流处陈岩处长、学生处就业与创业指导中心王晋主任陪同会见。

国际合作与交流处

2021年5月22日

《唱支山歌给党听》的快闪视频中不仅展示了北邮师生的精神风貌，还体现了学校的发展成就，视频发布后在全校师生和校友中引发了热烈反响，受到了新华社和v思想（北京市教工委）等多家权威媒体的广泛关注和转发报道，截至目前，视频全网播放量已突破60万，并在发布时登上哔哩哔哩网站热门视频第一。

红心永向党，永远跟党走，北邮师生将不忘初心，勇担使命，用青春和智慧奏响全面建设社会主义现代化国家的时代最强音，用责任和担当谱写全面建设信息科技特色{\*\*\*世界\*\*\*}{\*\*\*一流大学\*\*\*}新篇章，永葆初心，永不止步，让青春在不懈奋斗中绽放绚丽之花，以优异成绩迎接建党百年！

校团委

2021.5.22

附视频链接：