

# 第一步做什么

1. 按要求的日期去公众号和网址合集中搜索综述论文



图1 公众号中提及综述的文章

如图1所示，发现该公众号有一篇《注意力机制》综述的论文，点击进入该文章找到论文名字。

2. 在**新收集.xlsx**中填入除最后一列之外的其他列信息

先在**dblp** 中根据论文名字找到具体的论文信息填入 **新收集.xlsx** excel表格中，dblp没有再在 **arxiv** 中找；表格中汇总这列需要添入的是论文的分类，我们在 **新收集.xlsx** 中类别sheet里可以看到所有的类别

在 **第一步-示例.xlsx** 中可以看到详细的注意事项说明



图2 dblp上找到综述文章信息(有期刊会议信息，其中CoRR是指arxiv)

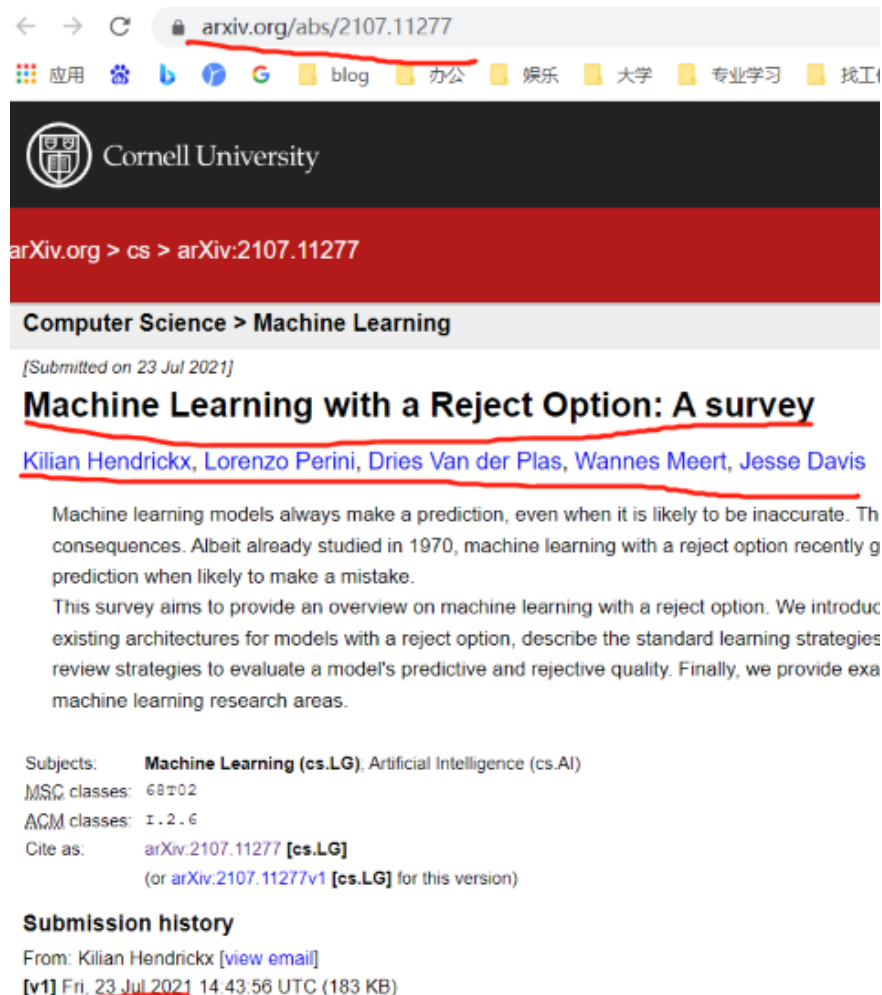


图3 arxiv上找到综述文章信息(缺少期刊会议信息)

#### 注意事项:

- 期刊会议填入缩写即可
- 汇总这列无法判断就空着，之后学长会来填。
- 该论文要是新收集的，已经收集过的就不用收集了(收集过的论文信息都在 **survey.xlsx** 中)。
- 论文链接如果有多个填入最新的链接。
- 论文如果是期刊会议，则用对应的时间；否则有多个时间用最早的时间。
- 一个paper可以有多个类别，如 **第一步-示例.xlsx** 所示的例子一样将该论文信息填写两次。
- 没有找到论文来源，标记为来源不明并标红

### 3. 在学长填好汇总这一列后，再填写引用

- 如图4所示，按照格式在引用这一栏添加路径：/bib/大分类/小分类/作者姓氏+年份+文章首个单词.md。

论文	汇总	作者	来源	时间	下载地址	引用
A survey of xxx	Computational Social Science and Social Media	Li Xiao, Mei Xiac	ACL	2021(虽然论文在2020年就挂在了arxiv上，但是是ACL21的论文，所以是2021年)	https://arxiv.org/pdf/2005.12584v2.pdf 错误的链接：v2要删去，因为作者以后可能会更新v3 v4版本，这样该链接就不是最新的了。而arxiv可以自动保持基础网址为最新论文链接。如果作者更新到了v4，基础网址会自动索引到https://arxiv.org/pdf/2005.12584.pdf 正确的链接：这个不加v2的是正确的paper链接	/bib/Natural Language Processing/Computational Social Science and Social Media/Xiao2021A.md
A survey of xxx	Machine Translation (一个paper可以有多个类别)	Li Xiao, Mei Xiac	ACL	2021	https://arxiv.org/pdf/2005.12584.pdf	/bib/Natural Language Processing/Machine Translation/Xiao2021A.md

图4 引用内容

- 如图5所示，在bib文件夹对应的路径下，新建刚刚命名的Xiao2021A.md文件。

« bib » Natural-Language-Processing » Computational-Social-Science-and-Social-Media				
	名称	修改日期	类型	大小
	Field2021A.md	2022/1/12 16:12	MD 文件	
	Joseph2020When.md	2022/1/12 16:13	MD 文件	
	Kiritchenko2021Confronting.md	2022/1/12 18:16	MD 文件	
	Nguyen2016Computational.md	2022/1/12 16:13	MD 文件	
	Sheng2021Societal.md	2022/1/12 15:35	MD 文件	
	Su2021A.md	2022/1/12 16:14	MD 文件	
	Zhou2018Fake.md	2021/4/19 13:30	MD 文件	
sona	Xiao2021A.md	2022/1/17 15:53	TXT 文件	

图5 文件夹新建md文件

如果发现文件名同名，如图6、7所示；我们按照规则(无、1、2、3、...)给文件名后面加数字，并将excel表格引用这列的内容也做修改。

Dabre2020A.md	2021/4
Dabre2020A1.md	2022/1
Dabre2020A2.md	2022/1

图6 文件名冲突解决方法

19069	/bib/Natural Language Processing/Machine Learning for NLP/Almeida2019Word.md
15395	/bib/Natural Language Processing/Machine Translation/Dabre2020A1.md
11115	/bib/Natural Language Processing/Machine Translation/Dabre2020A.md
12.07526	/bib/Natural Language Processing/Machine Translation/Yang2020A.md
10258	/bib/Natural Language Processing/Machine Translation/Chu2018A.md
10373	/bib/Natural Language Processing/Machine Translation/Gibadullin2019A.md
15395	/bib/Natural Language Processing/Machine Translation/Dabre2020A2.md
11391	/bib/Natural Language Processing/Machine Translation/Chakravarthi2021A.md

图7 解决冲突时，excel也要修改

- 在dblp上面查询该文章的bib，点击下图中的BibTex链接，将内容复制粘贴到刚刚创建的md文件中。

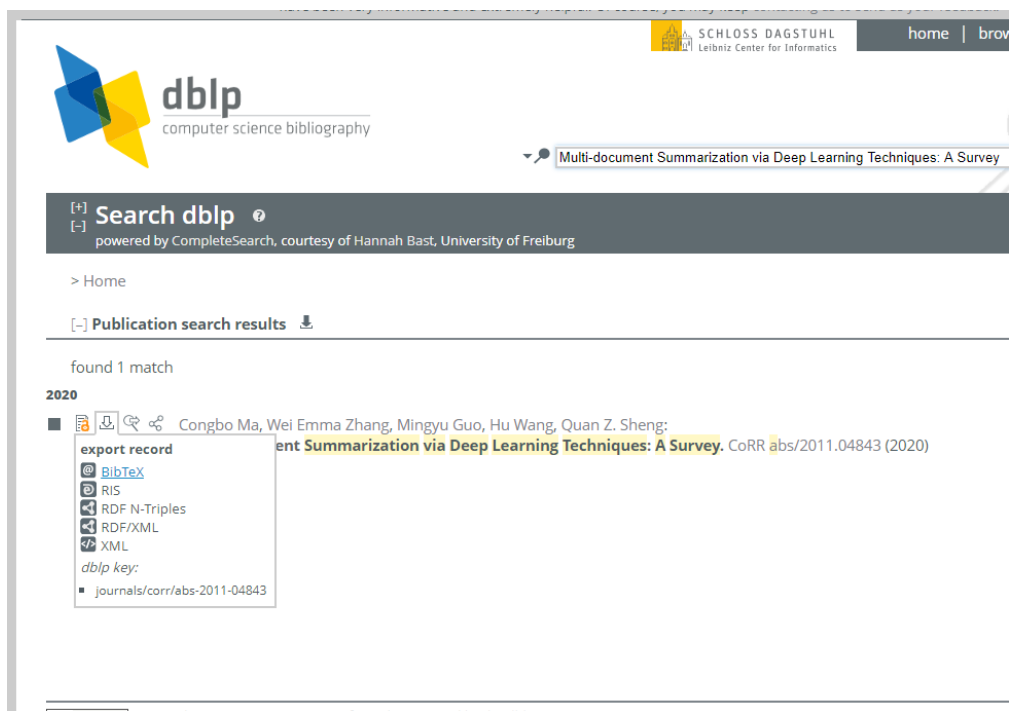


图8 BibTeX链接点击之后出现的内容

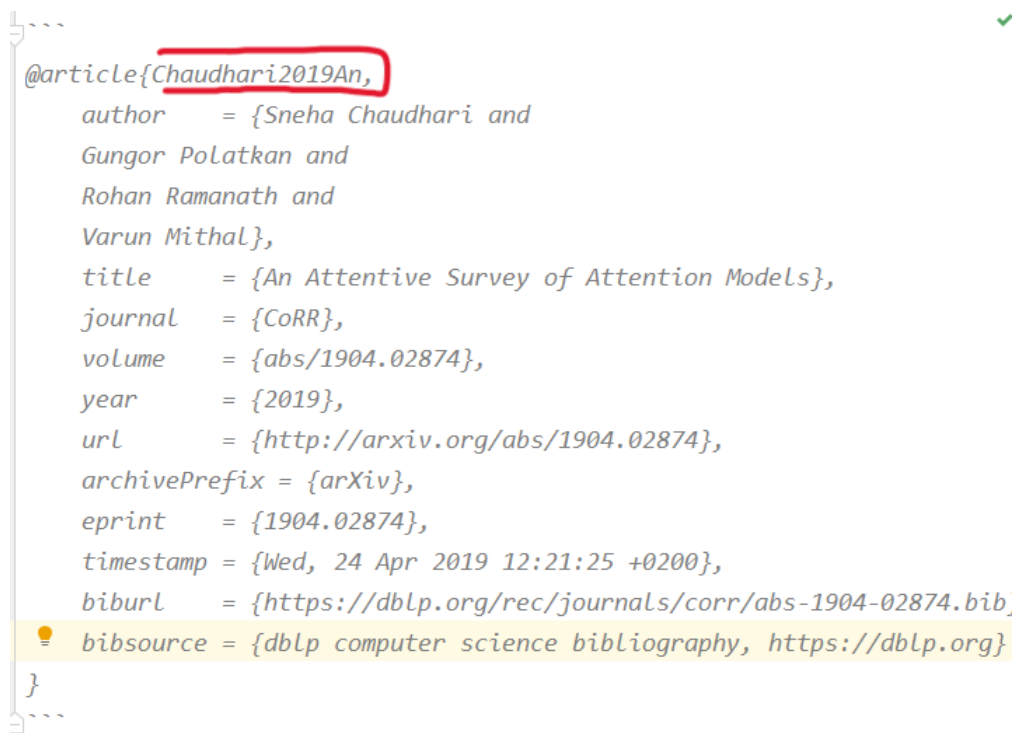


图9 md文件中的内容

#### 注意事项:

- 注意图9中画红线的地方，我们需要将 @article{ 后面的内容手动修改成作者姓氏+年份+文章首个单词（即和md文件名字保持一致）。
- 如果dblp中没有找到文章的bib，去arxiv、google学术、百度学术等其他渠道中找；如果依然没有找到，则将excel表格中该行标红。
- 如图9所示，不要忘记开头和结尾的 ````

## 公众号和网址合集

- AINLP
- 深度学习这件小事
- AI科技评论
- PaperWeekly
- AI科技大本营
- 专知
- 机器之心
- 深度学习与NLP
- 深度学习自然语言处理
- 机器学习算法与Python学习
- 夕小瑶的卖萌屋
- 香依科技
- MT论文列表: <http://paperreading.club/category?cate=NMT>
- 新智元
- arXivDaily
- 机器学习算法与自然语言处理
- <https://github.com/thunlp/SOS4NLP> (根据清华搜集的论文, 找一找我们没有的论文)
- 炼丹笔记
- 李rumor
- NLP工作站
- 机器学习与推荐算法
- AI有道
- 爱可可爱生活